

Research Article

Multworking Conditions Anomaly Detection of Mechanical System Based on Conditional Variational Auto-Encoder

Wenping Lei , Chenyang Li , Xinmin Dong , Junhui Wang , and Huajie Liu 

School of Mechanical and Power Engineering, Zhengzhou University, Zhengzhou 450001, China

Correspondence should be addressed to Wenping Lei; lwp@zzu.edu.cn

Received 21 October 2022; Revised 2 January 2023; Accepted 2 January 2023; Published 1 February 2023

Academic Editor: Yi Qin

Copyright © 2023 Wenping Lei et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Existing anomaly detection models of mechanical systems often face challenges for the equipment under multiple working conditions: the learning model under a single working condition is challenging to adapt to new working conditions, and centralized learning of multicondition samples leads to too low detection accuracy. A multworking condition variational auto-encoder (MW-CVAE) is proposed to solve the problem. Based on the variational auto-encoder model, the working conditions of the equipment are regarded as the input. The anomaly evaluation threshold of each independent working condition is established by the centralized learning of normal multworking condition samples. At the same time, it is found that the representation of each working condition sample in the space of the hidden layer forms a distribution close to the prior probability, providing a theoretical basis for the separation and evaluation of working conditions. By comparing and verifying the CWRU, JNU, and PU datasets, the new method significantly improves anomaly detection (the *F1*-score value is increased by 18-19%) and can be widely used in anomaly detection mechanical systems with various discrete working conditions.

1. Introduction

Anomaly detection is a method for identifying abnormalities and illogical data mining, and it is an essential branch of machine learning. Especially in the era of big data, the speed of manual data processing has been far behind that of computers, and therefore, faster detection of abnormal data is a valuable task nowadays [1, 2]. In industries, the anomaly detection of mechanical devices is crucial. More immediate and more accurate anomaly detection help prevent accidents and improves reliability and production efficiency [3, 4].

Anomaly detection algorithm-based machine learning mainly includes One-Class Support Vector Machines (OC-SVMs), Principal Component Analysis (PCA), and Local Outlier Factor (LOF). For OC-SVM, the normal data is used to train the model to obtain a hyperplane, which is utilized to circle the positive data. OC-SVM takes the hyperplane as a criterion and considers the samples inside the circle are positive. Since the computation of kernel function is time-consuming, OC-SVM is not widely used under massive data [5-7]. PCA is a statistical algorithm to convert a set of

potentially correlated variables into a set of linearly uncorrelated variables, and the transformed set of variables is called principal components [8]. LOF measures the density deviation of a given sample concerning its neighbours and determines whether a point is an outlier by comparing the density of the sample with its neighbour. This algorithm is suitable for data with noticeable density differences, and its complexity is high; therefore, it does not apply to big data [9].

In recent years, the anomaly detection algorithm based on deep learning has been an effective method and academic focus gradually [10-22]. Auto-Encoder (AE) is widely used for anomaly detection due to its excellent deep representation. The AE-based anomaly detection algorithm minimizes the reconstruction error to establish a representation model of normal samples and identifies the samples whose reconstruction error exceeds the threshold as anomalies [19]. Chen et al. proposed a novel quadratic function-based deep convolutional auto-encoder (DCAE) in predicting the remaining useful life (RUL) of bearing [23]. The bearing vibration signals are first preprocessed by

low-pass filtering and then fed into the quadratic function-based DCAE neural networks. It can generate a bearing Health Indicator (HI) from raw vibration signals and can be better applied to RUL prediction than other existing HI. However, AE has a disadvantage in that it only gives a low-dimensional hidden space representation and cannot learn the characteristics of a sample's probability distribution. However, by employing a variational auto-encoder (VAE), a specific sample can be expressed as the distribution of possible samples, and the latent space becomes a continuous distribution space. VAE, therefore, is competitive in the field of anomaly detection as a generation model. In 2015, An and Cho [17] proved the feasibility of VAE in unsupervised anomaly detection and applied VAE in network intrusion detection. Literature [11] proposed an unsupervised anomaly detection model donut based on VAE. The encoder extracts representative features from Key Performance Indicator (KPI) sequences. The decoder reconstructs the sequence according to the features and calculates the anomaly in the deviation sequence, which is between the reconstructive sequence and the origin sequence. This model fully uses the deep representation ability to model KPI series and takes advantage of the representation abilities of KPI sequence in the deep generation models. References [12, 13] discuss how the condition K affects VAE. According to the different conditions, the potential distribution can be generated and be used for detection. Due to the different thresholds of anomalies, the model can detect local and global anomalies. References [24, 25] discussed the unsupervised construction of HI. In [24], it presented a new unsupervised HI construction approach. The method innovatively constructs the HI of a distribution contact ratio metric health indicator (DCRHI) to represent the degradation process well and obtain a uniform failure threshold. Qin et al. proposed a novel degradation-trend-constrained VAE (DTC-VAE) to construct the HI vector with the distinct degradation trend [22]. Compared with other typical unsupervised HI construction methods, this method can more easily determine the uniform failure threshold.

Mechanical devices work under different conditions, such as the change of load, rotating speed, and input power, leading to two challenges in the anomaly detection of a mechanical system. (i) The model learned from a single working condition is not appropriate for a new working condition and even identifies the normal samples under another condition as anomalies. (ii) The concentration learning of samples under multiconditions may lead to low detection accuracy. The research on anomaly detection under multiple working conditions has been relatively inadequate [4, 26].

The Multiworking Conditions Variational Auto-Encoder (MW-CAVE) is presented in the paper to solve the low detection accuracy under multiworking conditions. MW-CAVE takes the working conditions as the conditional input of VAE, obtains the distribution of samples by concentration learning on normal data, and determines the threshold of anomaly detection, increasing the detection accuracy under multiworking conditions.

2. Anomaly Detection of Multiworking Conditions

2.1. Problem Analysis. Figure 1 demonstrates the vibration responses of normal and abnormal cases, of which data are quoted from the JNU dataset (more details on the JNU dataset will be introduced in Section 5.1). Figures 1(a)–1(c) show the vibration of normal cases, and others show the abnormal.

Figure 2 illustrates the results using VAE anomaly detection (introduced in Section 3.1). A typical VAE anomaly detection utilizes the reconstruction error (L_{MSE}) as the anomaly score to evaluate the abnormal condition. Threshold 1 in Figure 2 represents the best threshold according to the maximum principle of the $F1$ -score (introduced in Section 5.2.3) and the normal and abnormal are expressed as 0 (False) and 1 (True), respectively. Anomaly scores that exceed the threshold are considered abnormal. Table 1 compares the detection performance of the JNU dataset and PU dataset by VAE anomaly detection and shows both low Area Under Curve (AUC) and $F1$ -score and poor performance of anomaly detection. Scores of the three normal cases divide into three layers in Figure 2, representing three typical working conditions, respectively. If threshold 2 is used as the standard, some abnormal samples (in region 1) may be misclassified into normal. If threshold 3 is used as the standard, all the samples in region 1 and region 2 are misclassified as normal. The phenomenon of the misclassification is called Misclassification Caused by Working Condition Interference, and this is the main reason leading to a low anomaly detection performance.

2.2. Idea of Paper. Figure 3 shows the idea of the paper. The ellipses 1 and 2 represent the distributions of the normal samples of two working conditions obtained by generating distributed model learning, and the ellipse edges represent the classification boundary. Points 1 and 2 stand for abnormal and normal samples of working condition 1, respectively. Point 3 stands for an abnormal sample of working condition 2. If the two working conditions were not distinguished, two kinds of errors would have happened: (1) point 1 is an abnormal sample of working condition 1 but would be considered normal according to the distribution of working condition 2. (2) Point 2 is a normal sample of working condition 1 but would be abnormal according to the distribution of working condition 2. An anomaly detection method is established to avoid these two kinds of errors, as shown in Figure 3(b). Two working conditions will be separated. In low-dimensional space, the normal samples of each independent working condition are learned to form a Gaussian distribution. After two working conditions are isolated, points 1 and 2 will be compared with the distribution of working condition 1. According to the probability density, point 1 is judged as an abnormal sample, and point 2 is deemed as a normal sample. Similarly, point 3 can also be correctly identified.

By utilizing the ability of a variational auto-encoder (VAE) to learn the data distribution, and a conditional

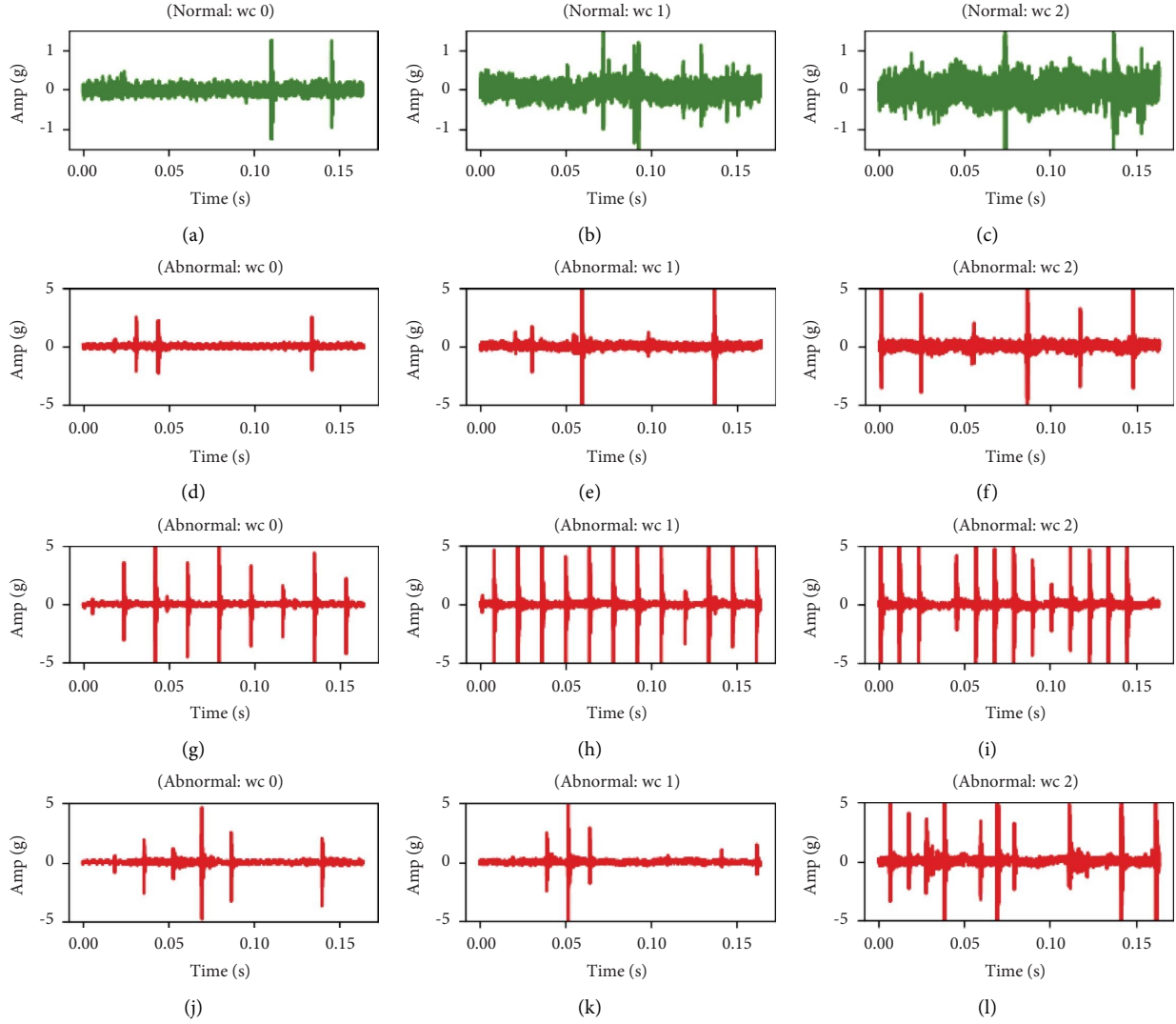


FIGURE 1: Examples of multiple working conditions (data from JNU dataset. “wc” is short for working condition). (a) Normal: wc 0. (b) Normal: wc 1. (c) Normal: wc 2. (d) Abnormal: wc 0. (e) Abnormal: wc 1. (f) Abnormal: wc 2. (g) Abnormal: wc 0. (h) Abnormal: wc 1. (i) Abnormal: wc 2. (j) Abnormal: wc 0. (k) Abnormal: wc 1. (l) Abnormal: wc 2.

variational auto-encoder (CVAE) to separate the working conditions, a Multiworking Conditions Anomaly Detection Method, MW-CVAE, is proposed in this paper. Based on the centralized learning of various working conditions, the proposed method establishes the respective distribution characteristics (latent space) of different working conditions samples and calculates the exclusive anomaly metric of different working conditions samples.

The new method can overcome the misclassification problem caused by working condition interference under multiple working conditions.

2.3. Definition of Multiworking Conditions Anomaly Detection. Multiworking conditions anomaly detection is defined as a mechanical system that has C kinds of working conditions, $C = \{0, \dots, c-1\}$, of which l data samples of each working condition are collected, respectively; for C kinds of working conditions, the total number of samples is

$n = C \cdot l$. The datasets of samples are time series $D = \{X_1, X_2, \dots, X_n\} \in R^{l \times n}$, $X_i = [x_{i1}, x_{i2}, \dots, x_{ij}]^T \in R^J$ ($i = 1, 2, \dots, n$), the subscript “ i ” represents the i^{th} sample, and the length of the data sequence is J . A model, described by similar function $P(x|c)$, will be established to determine the anomaly algorithm and corresponding threshold ϵ . If $P(x|c) < \epsilon$, the sample is normal; otherwise, it is abnormal.

3. Methodological Foundation

3.1. Variational Auto-Encoder (VAE). VAE, shown in Figure 4, is a deep Bayesian network that can establish a relationship between visible variable x and latent variable z , which is usually a multivariate unit Gaussian distribution. VAE simulates the data distribution $P_\theta(x)$ through a neural network $P_\theta(x|z)$ with parameters θ and takes samples from hidden layers z . The data that conforms to the distribution of $P_\theta(x)$ be generated by $\int P_\theta(x|z)p(z)dz$. Since the true posterior $P_\theta(z|x)$ is intractable by analytic methods, similar

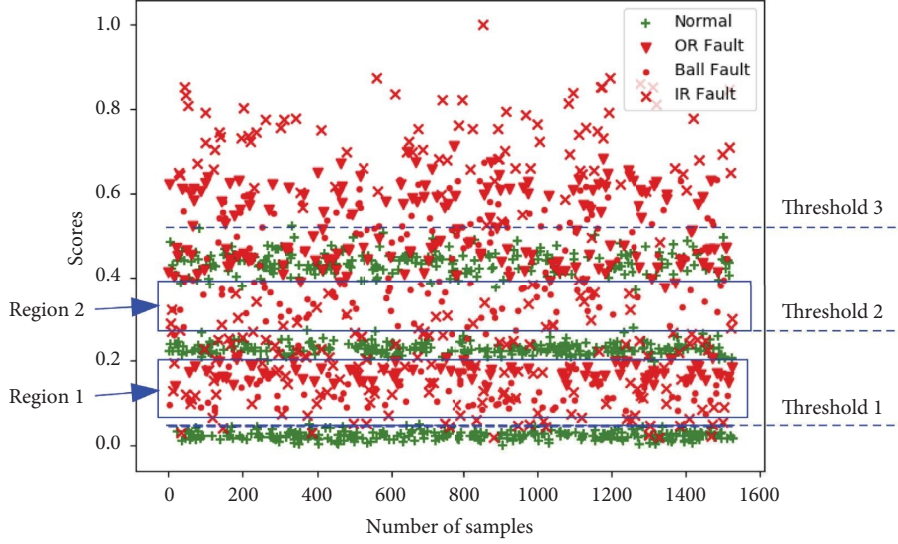


FIGURE 2: Anomaly detection of multiworking conditions based on VAE (JNU dataset with three working conditions).

TABLE 1: Results of VAE-based anomaly detection.

Cases	AUC	Accuracy	Best_three	F1-score
JNU dataset	0.6949	0.6616	0.0464	0.7448
PU dataset	0.6109	0.5400	0.1107	0.6604

to an auto-encoder, the parameter estimation approach of a VAE is approximating the distribution $P_\theta(z|x)$ through a

simple distribution $q_\phi(z|x)$. Refer to [10], the process of calculation is as follows:

$$\log P_\theta(x) - \mathbb{D}_{KL}(q_\phi(z|x) \| P_\theta(z|x)) = \mathbb{E}_{z \sim q}[\log P_\theta(x|z)] - \mathbb{D}_{KL}(q_\phi(z|x) \| P_\theta(z)). \quad (1)$$

The training loss of a VAE is as

$$\mathcal{L}_{\text{ELBO}} = \mathbb{E}_{z \sim q}[\log P_\theta(x|z)] - \mathbb{D}_{KL}(q_\phi(z|x) \| P_\theta(z)). \quad (2)$$

To calculate Kullback–Leibler (KL) divergence, the model considers $q_\phi(z|x)$ and $P(z)$ following normal distribution, $P(z) \sim N(0, I)$, $q_\phi(z|x) \sim N(\mu_x, \sigma_x^2 I)$. μ_x , and σ_x^2 represents the mean and variance generated by the encoder network $q_\phi(z|x)$, and the reparameterization is utilized to sample z .

3.2. Conditional Variational Auto-Encoder (CVAE). CVAE, shown in Figure 5, is a conditional-directed graphical model where input observations modulate the prior on latent variables that generate the outputs, in order to model the distribution of high-dimensional output space as a generative model conditioned on the input observation [13]. There are three types of variables. For random observable x , z (unknown, unobserved), and c (known, observed) are independent random latent variables. The conditional probability $P_\theta(x|z, c)$ is formed by a nonlinear transformation, with the parameter θ . ϕ is another nonlinear function that

approximates inference posterior $q_\phi(z|x, c) = N(\mu, \sigma I)$. The latent variable z allows for modeling multiple modes in conditional distribution of x given c making the model enough for modeling one-to-many mapping. To make an approximation of ϕ and θ , ELBO in (3) is given as

$$\log P_\theta(x) \geq \mathbb{E}_{q_\phi(z|x, c)}[\log P_\theta(x|z, c)] - \mathbb{D}_{KL}(q_\phi(z|x, c) \| P(z)). \quad (3)$$

Gaussian latent variable z samples hidden variable z . Refer to [7], assume $P(z) \sim N(0, I)$, $q_\phi(z|x, c)$ should be $q_\phi(z|x) \sim N(\mu_x, \sigma_x^2 I)$, the second part in (3) has an analytical solution.

$$\mathbb{D}_{KL}(q_\phi(z|x, c) \| P(z)) = \frac{1}{2} \sum_{j=1}^J (1 + \log \sigma_j^2 - \mu_j^2 - \sigma_j^2). \quad (4)$$

The first item in (4), called reconstruction error, represents the difference between the input x and reconstruction \hat{x} , then loss function $\mathcal{L}_{\text{CVAE}}$ is expressed as

$$\mathcal{L}_{\text{CVAE}}(x, c) = \mathcal{L}(x, \hat{x}) - \frac{1}{2} \sum_{j=1}^J (1 + \log \sigma_j^2 - \mu_j^2 - \sigma_j^2), \quad (5)$$

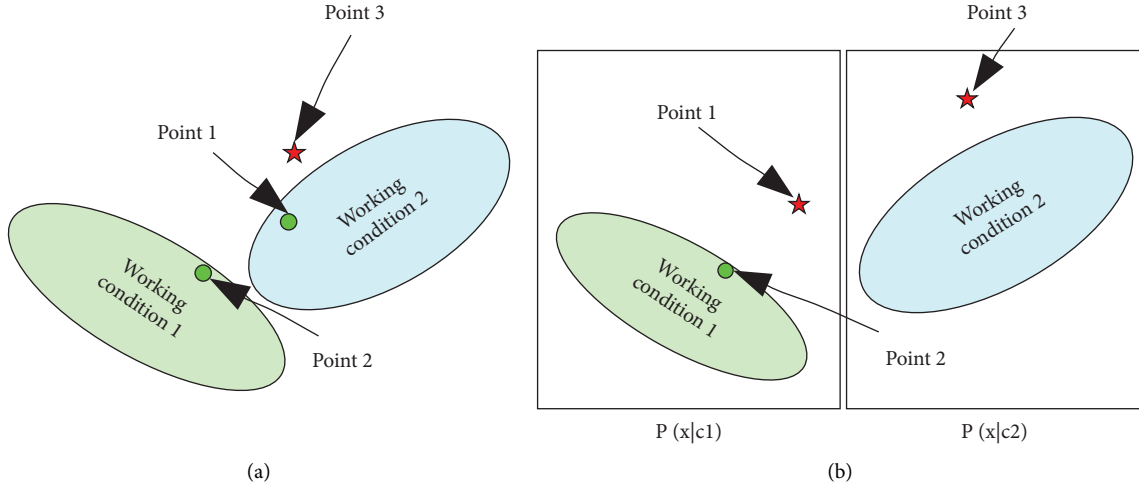


FIGURE 3: The idea of the paper.

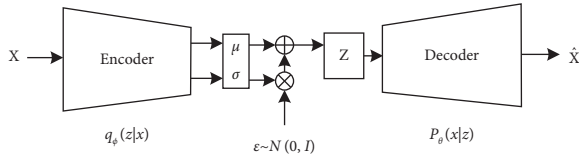


FIGURE 4: VAE model.

where J is the dimension of x and \hat{x} .

4. MW-CAVE Anomaly Detection

4.1. Process of Anomaly Detection. The MW-CVAE model is based on CVAE, and its process is shown in Figure 6. The details of detection are described as follows:

- (a) Data preprocessing. In the data preprocessing of anomaly detection, the most critical part is normalization. References [8, 16] adopt the normalization of mean and variance, namely, $x_{\text{scaled}} = (x - \mu)/\sigma$. μ and σ stand for the mean and variance of training data. However, the above-given method is only applicable to the same ranges of normal samples and anomalous samples. Usually, the range of anomalous samples is larger than normal. In the training step, only normal samples, which have a relatively small range will be trained. When we get to the testing phase, the mean and variance of testing data may change. To avoid the error caused by the normalization of testing data, the article suggests adopting linear global normalization:

$$x_{\text{scaled}} = \frac{(x - \min_x)}{(\max_x - \min_x)}, \quad (6)$$

where \min_x and \max_x represent the static minimum and maximum ranges of all samples, respectively, rather than the statistical functions, like $\min(x)$ and $\max(x)$. In practical application, we choose a proper value of \min_x and \max_x to

ensure all the training and testing samples are in the range (\min_x , \max_x).

- (b) Network structure. The network structure is shown in Figure 7. The input layer adopts the normalized vibration signal X , and the number of nodes is 8192. Working condition c adopts a 30-node one-hot coding mode, which is detailed in part (c). Encoders are composed of 400 and 200 full connection layers, respectively, and the activation function is ReLU. The hidden layer size is 2. Decoders, on the contrary, use 200,400 nodes of the fully connected layer, the number of nodes in the output layer is 8192.
- (c) Encoding of working conditions. The encoder and decoder add the same input representing a specific working condition c . The working condition c is a known variable in the training and testing process and maybe one or more of the following: (i) the devices are at different speeds. (ii) The devices are under different loads. (iii) The devices produce workpieces with different specifications. The working condition c is encoded by a one-hot form. Since the speed of devices is continuous, for example, the rotating speed is between 200 to 950 rpm, the number of types is infinite. In practice, for the sake of safety, energy-saving, and high efficiency, the working speed is only a few, such as low speed (0), medium speed (1), and high speed (2). The load and specification can be encoded similarly. The one-hot encoding of working conditions is shown in Figure 8.
- (d) Training process. In the training process, Stochastic Gradient Variational Bayesian (SGVB) is used [7]. Since the CVAE model is adopted, each sample and its working condition information are as input, and the normal samples containing all working conditions are the input to learn.
- (e) Anomaly score. In this paper, we adopt the reconstruction-based approach to evaluate the degree of abnormality [27]. When the samples are abnormal,

Input: Trained CVAE model, X_{test} , C_{test} , A_{test} ; Testing data set $X_{\text{test}} = \{x^{(1)}, x^{(2)}, \dots, x^{(N_{\text{test}})}\}$ working condition $C_{\text{test}} = \{c^{(1)}, c^{(2)}, \dots, c^{(N_{\text{test}})}\}$ abnormal label $A_{\text{test}} = \{a^{(1)}, a^{(2)}, \dots, a^{(N_{\text{test}})}\}$ $c^{(i)} \in W$, $a^{(i)} \in \{0, 1\}$; $W = \{w_1, w_2, \dots, w_C\}$

Output: best_threshold: ϵ (W)

```

scores  $\leftarrow$  score( $X_{\text{test}}$ ) # get anomaly score by equation (7)
for  $i = 1$  to  $C$ 
  wi_ind  $\leftarrow$   $C_{\text{test}} = w_i$  #get index by  $w_i$ 
  Si  $\leftarrow$  scores [wi_ind]
  Ai  $\leftarrow$   $A_{\text{test}}$  [wi_ind]
   $t = 0.01$ 
  while  $t < 1$  do
    Ai_pred  $\leftarrow$  Si  $> t$ 
    best_t  $\leftarrow$  F1-score (Ai, Ai_pred, Si)
     $t = t + 0.01$ 
  end while
   $\epsilon(w_i) \leftarrow$  best_t #best threshold for  $w_i$ 
end for

```

ALGORITHM 1: Threshold learning algorithm.

Input: samples and corresponding working conditions ($x^{(i)}, c^{(i)}$)

Output: anomaly or not

```

score  $\leftarrow$  score( $x^{(i)}$ ) # get anomaly score by equation (7)
If score  $<$  best_threshold ( $c^{(i)}$ )
   $x^{(i)}$  is normally
else
   $x^{(i)}$  is anomaly

```

ALGORITHM 2: Anomaly detection algorithm.

the L_{MSE} error is high, namely, high abnormality. The anomaly score of sample x is expressed as

$$\text{score}(x) = L_{\text{MSE}}(x, \hat{x}) = \frac{1}{J} \sum_{j=1}^J (x_j - \hat{x}_j)^2, \quad (7)$$

where J is the dimension of input samples.

(f) Determination of anomaly threshold. The paper's determination method of abnormal threshold is significantly different from a traditional CVAE method. An anomaly detection method based on an anomaly score must determine an anomaly threshold to distinguish whether the sample is normal or abnormal. The selection of abnormal threshold is the key to the detection performance [27]. Referring to the paper [28], we first obtain a set of test samples with known abnormal labels and obtain all abnormal scores. By drawing the relationship curve between F1-score (including precision rate and recall rate) and the threshold, we find that the threshold corresponding to the maximum value of the F1-score is the best threshold ϵ . However, in the multicondition scenario, due to the 'Misclassification Caused by Working Condition Interference', the optimal threshold obtained by the above method still cannot guarantee good anomaly detection performance (problem analyzed in Section

2.1). In order to prevent this problem, the proposed method uses the idea of working condition separation to learn the best threshold.

The method given in Algorithm 1 can greatly improve the accuracy of multiworking condition anomaly detection. The core idea of the algorithm is to classify the test samples with abnormal labels according to the working conditions, that is, to learn the optimal threshold for each working condition.

(g) MW-CAVE anomaly detection. After the detection threshold is determined by step (f), Algorithm 2 can be used to determine the abnormal state under a certain working condition.

4.2. Testing Algorithm. In Step 1, the testing data set X_{test} , working condition C_{test} , abnormal label A_{test} , and trained CVAE model for testing are collected. The number of samples is N .

In Step 2, the optimal threshold is identified for each working condition. We obtain the set of test samples with known abnormal labels and the anomaly scores by equation (7). The index of the data under different working conditions is recorded as wi_ind . Corresponding to the index, the sample anomaly score and the true anomaly label are recorded as Si and Ai , respectively. The normalized threshold t (range: 0, 1) starts at 0 and increases in steps of

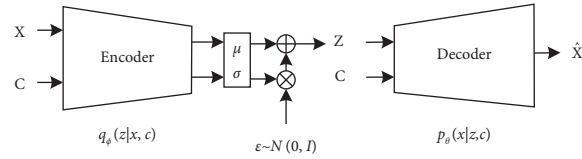


FIGURE 5: CVAE model.

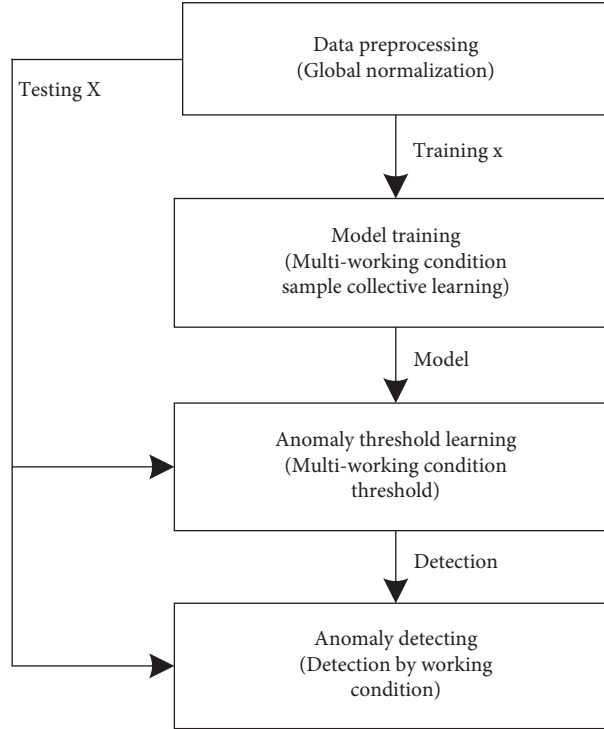


FIGURE 6: The detection process of MW-CAVE.

TABLE 2: Four working conditions.

Working condition	Rotate speed (r/min)	Load (HP)
0	1797	0
1	1772	1
2	1750	2
3	1730	3

0.01. According to S_i , A_i , and t , the best threshold ($best_t$, the value of t which generate the maximum $F1$ -score) is calculated by the $F1$ -score. The above-given operation is repeated to obtain each optimal threshold under different working conditions.

5. Examples

5.1. Introduction of Dataset. The CWRU dataset provided by Case Western Reserve University Bearing Data Center [29] is one of the most famous open-source datasets in fault diagnosis research. Data of CWRU was collected by accelerometers attached to the housing with magnetic bases. This paper uses the driver end-bearing fault data whose sampling frequency is 12 kHz, and four working conditions are listed in Table 2.

JNU dataset is a dataset on the rolling bearings provided by Jiangnan University [23] and contains four health statuses: normal, inner ring fault, outer ring fault, and roller fault. Accelerometers collect the vibration signal under three rotating speeds, 600 rpm, 800 rpm, and 1000 rpm, at the sampling frequency of 50 kHz. Three typical working conditions are listed in Table 3.

PU dataset is a dataset on the rolling bearings provided by Universität Paderborn [24]. The type of rolling bearing is 6203. The faults are divided into practical faults and artificial faults. The latter is discussed in the paper. In the artificial faults, the crack, spalling, and pitting are machined by electrical discharge machining, drilling, and electrical engraving. The vibration responses are acceleration signals, of which the sampling frequency is 48 kHz. The working conditions are divided according to the torque, radial force,

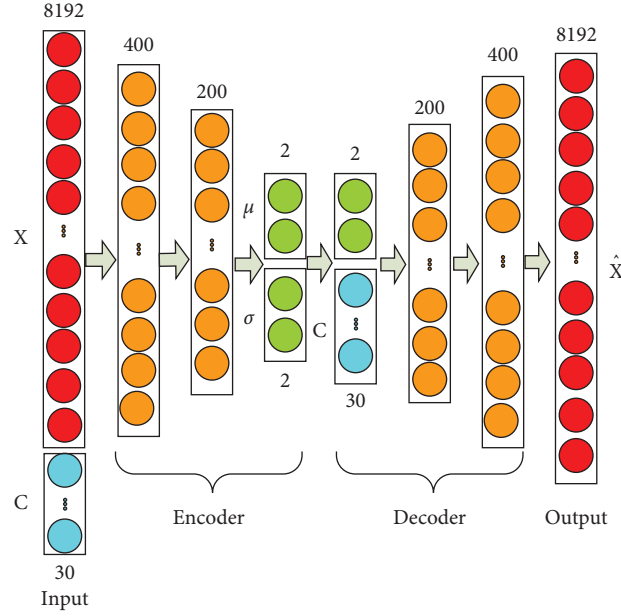


FIGURE 7: The network structure of MW-CAVE.

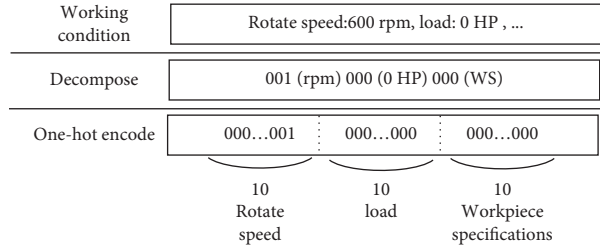


FIGURE 8: The way of encoding a working condition.

TABLE 3: Three working conditions.

Working condition	Rotate speed (r/min)
0	600
1	800
2	1000

and rotating speed. Four typical working conditions in PU dataset are listed in Table 4.

5.2. Experimental Method and Criterion

5.2.1. Baseline Method. In the experiment, five methods are utilized for comparison. The former two methods are machine learning, and the latter three are deep learning.

- (a) **Principal Component Analysis (PCA).** After the eigenvalues of the covariance matrices of samples are decomposed, the eigenvalues are the variances corresponding to the samples projected onto axes. A smaller eigenvalue indicates that the sample is concentrated. Meanwhile, the anomaly is easier to shift. It can be used as an indicator to distinguish anomalies [8].

- (b) **Local Outlier Factor (LOF).** LOF method compares the density of given data points to that of their neighbors. Since the outliers come from the areas with a lesser density, the ratio of abnormal data points is higher. LOF method detects whether the data are normal or not by comparing the densities of given data points with the data points near them [9].
- (c) **Variational Auto-Encoder (VAE).** For a VAE, the reconstruction error is considered as a score. According to the maximum of the $F1$ -score on the testing set, the best threshold is determined by linear normalization. The best threshold is regarded as a criterion for anomaly detection [25, 27].
- (d) **Conditional Variational Auto-Encoder (CVAE).** References [12, 13] introduce anomaly detection by CVAE. For the case of learning of multiple working conditions, $best_thr$ is the threshold corresponding to the maximum of the $F1$ -score.
- (e) **Multiworking Condition Anomaly detection (MW-CVAE).** When there are C kinds of working conditions, it is divided into C kinds of independent working conditions, and the threshold determination and anomaly detection are carried out for each independent working condition.

TABLE 4: Four working conditions.

Working condition	Torque	Radial force	Rotating speed (r/min)
0	0.7	1000	1500
1	0.7	1000	900
2	0.1	1000	1500
3	0.7	400	1500

TABLE 5: Confucian matrix.

Confucian matrix		Prediction	
		Positive	Negative
Real	Positive	TP	FN
	Negative	FP	TN

5.2.2. *Experimental Parameters.* The parameters of 3 deep learning models in the article are listed as follows:

- VAE. For the encoder, x size of the input layer is 8192. The nodes in the middle layers are 400 and 200, respectively. z size of the hidden layer is 2. For the decoder, the input is 2. The nodes in the middle layers are 200 and 400, respectively. The size of the output layer is the same as that of the input layer, and both are 8192.
- CVAE. For the encoder, x size of the input layer is 8192. The nodes in the middle layers are 400 and 200, respectively. z size of the hidden layer is 2. For the decoder, the input is 2. The nodes in the middle layers are 200 and 400, respectively. The size of the output layer is the same as that of the input layer, and both are 8192. The working conditions are encoded with one-hot of length 30 bits. The training adopts Adam optimization, and the learning rate is $10e^{-3}$.
- MW-CVAE. The parameters are the same as these of CVAE.

5.2.3. *Criteria.* Under the centralized learning of multiple working conditions, the detection accuracy of the samples covering all working conditions and health status is compared to verify the effectiveness of MW-CVAE. The area under curve (AUC), $F1$ -score, and accuracy are treated as indicators for comparing different methods.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}, \quad (8)$$

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (9)$$

$$\text{Recall} = \frac{TP}{TP + FN}, \quad (10)$$

$$F1 - \text{score} = 2 \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (11)$$

In order to introduce the performance indexes $F1$ -score and Accuracy, the confusion matrix in Table 5 is introduced.

TP is the number of positive (abnormally) samples predicted to be positive. FN is the number of positive samples predicted to be negative. FP is the number of negative (normal) samples predicted as positive and TN is the number of negative samples predicted as negative. The formulas of Accuracy, Precision, Recall, and $F1$ -score are shown in equations (8)–(11), respectively.

5.3. Results of Anomaly Detection

5.3.1. *Performance.* The anomaly detections by MW-CVAE and CVAE are compared on the JNU dataset, as shown in Figure 9. Figures 9(a)–9(c) demonstrate the visual display of anomaly detection by MW-CVAE under $wc = 0$, $wc = 1$ and $wc = 2$, and Figure 8(d) shows that by CVAE under $wc = [0, 1, 2]$. The blue dash line stands for the position of the best threshold. In Figure 9(d), there are stratifications between the three normal working conditions, indicating that the three working conditions are at different abnormal levels (green “+” points in Figure 9(d)). Figures 9(a)–9(c) show the distributions of the best threshold and samples. Due to the separation of working conditions, the interference of working conditions is avoided.

Table 6 compares the results by MW-CVAE and CVAE. The method proposed in this paper has a significant improvement. Accuracy is improved by 0.16–0.1645, AUC is improved by 0.1139–0.1154, and $F1$ -score is improved by 0.1815–0.1918. The determination method of abnormal threshold presented in this paper is different from the CVAE method. In MW-CVAE, it learns the best threshold according to the idea of working condition separation and avoids the interference of working conditions.

Figure 10 compares the receiver operating characteristic curve (ROC) using VAE and CVAE on the JNU dataset. The AUC of CVAE is higher than that of VAE. Also, the AUC of MW-CVAE is higher than that of MW-VAE. The reason is that there is an interference between different working conditions using VAE.

5.3.2. *Comparison of Different Datasets.* Table 7 compares the $F1$ -score and AUC of the methods on CWRU, JNU, and PU datasets. It can be seen from the comparison in Table 7

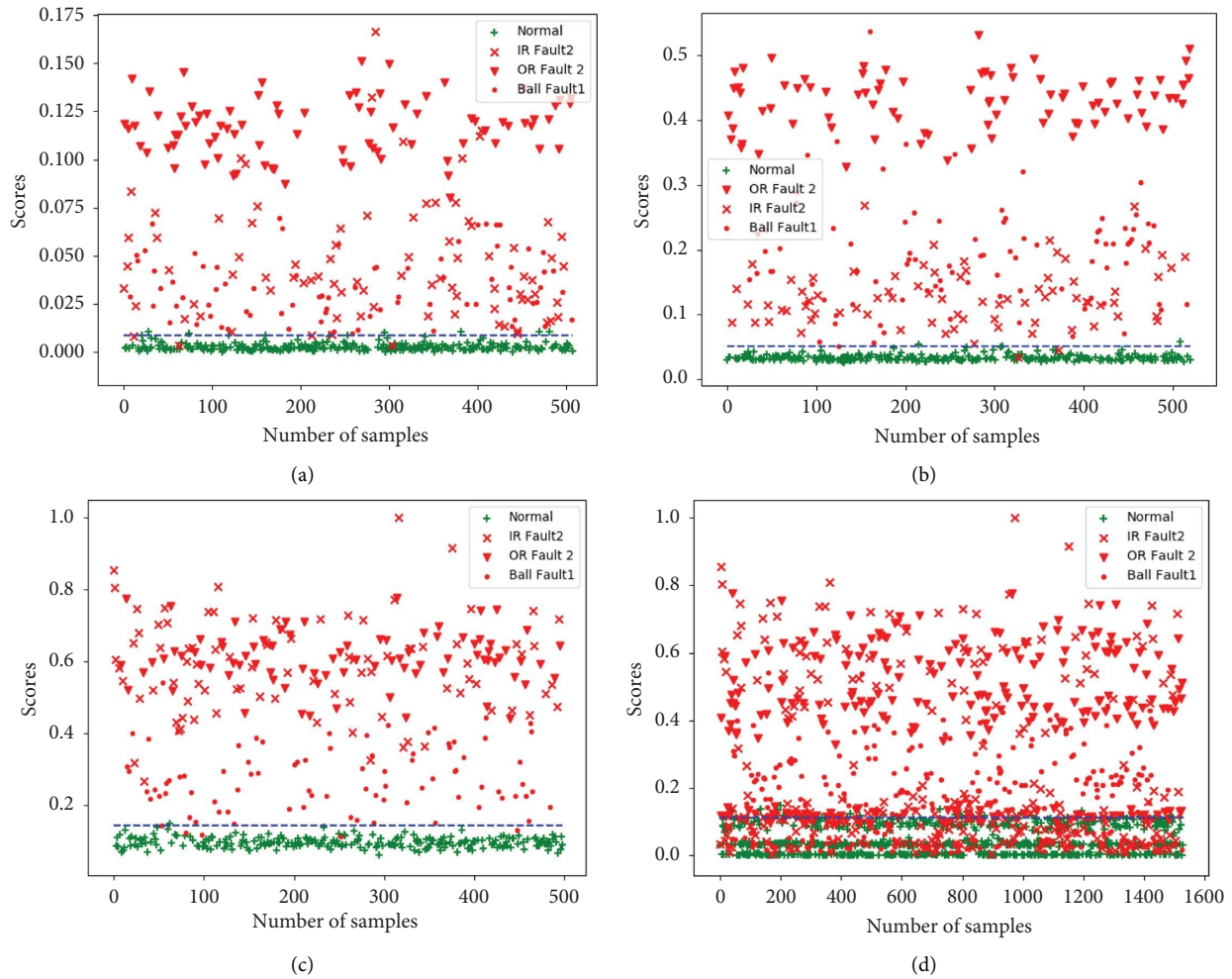


FIGURE 9: Anomaly detection by MW-CVAE and CVAE.

that in addition to the AUC performance of one PCA method on the PU data set, the MW-CVAE method has achieved the best $F1$ -score and AUC values on all data sets compared with other methods.

5.4. Effects of Parameters

5.4.1. Effects of Latent Space. Figure 11 shows the distribution of hidden learning samples under different epochs when the hidden layer size is 2. The red, blue, and green points represent three different working conditions of learning samples. As the epoch progresses, the network learning gradually converges. Meanwhile, the final

distribution of each working condition will form the normal distribution corresponding to the respective working condition. However, the mean and variance are different for each working condition, precisely with the assumption of CVAE.

5.4.2. Effects of the Size of Latent Layers. Figure 12 shows the $F1$ -score on JNU and PU datasets under the different sizes K of hidden layers. When K increases from 2 to 200, $F1$ -score on JNU dataset fluctuates slightly between 2 and 10. In other cases, $F1$ -score remains almost the same. It indicates that the size of hidden layers has little effect on MW-CVAE accuracy.

TABLE 6: Comparison of accuracy.

Method	Working condition	Accuracy	Auc	F1-score
CVAE	WC = [0, 1, 2]	0.8259	0.8834	0.8003
MW-CVAE	[WC = 0]	0.9803	0.9973	0.9818
	[WC = 1]	0.9904	0.9984	0.9921
	[WC = 2]	0.9859	0.9988	0.9884

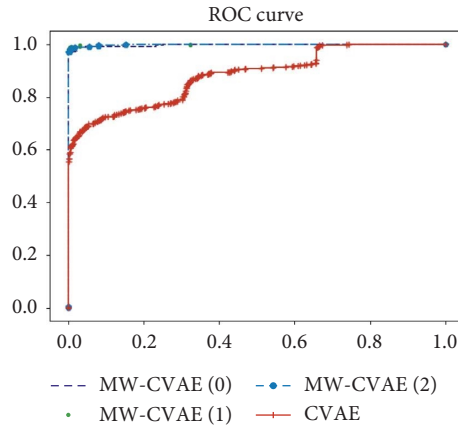
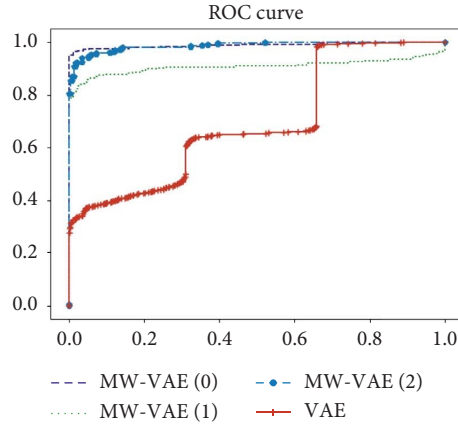


FIGURE 10: Comparison of ROC by MW-CVAE and MW-VAE for JNU dataset.

TABLE 7: Anomaly detection accuracy of different methods.

Dataset	PCA		LOF		VAE		CVAE		MW-CVAE	
	F1-score	AUC	F1-score	AUC	F1-score	AUC	F1-score	AUC	F1-score	AUC
CWRU	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.01.01.0	1.01.01.0
JNU	0.8464	0.8986	0.8484	0.8981	0.7448	0.6949	0.8155	0.8895	0.98190.99220.9829	0.99730.99790.9970
PU	0.6606	0.6723	0.6441	0.5091	0.6604	0.6109	0.6441	0.4972	0.76530.69980.7185	0.63880.39870.6075

The bold type indicates that the method obtains the best performance index in this database. For example, in database PU, 0.6723 (bold) indicates that PCA methods outperform other methods on AUC performance indicators. In addition, MW-CVAE wins every performance index (F1-score and AUC).

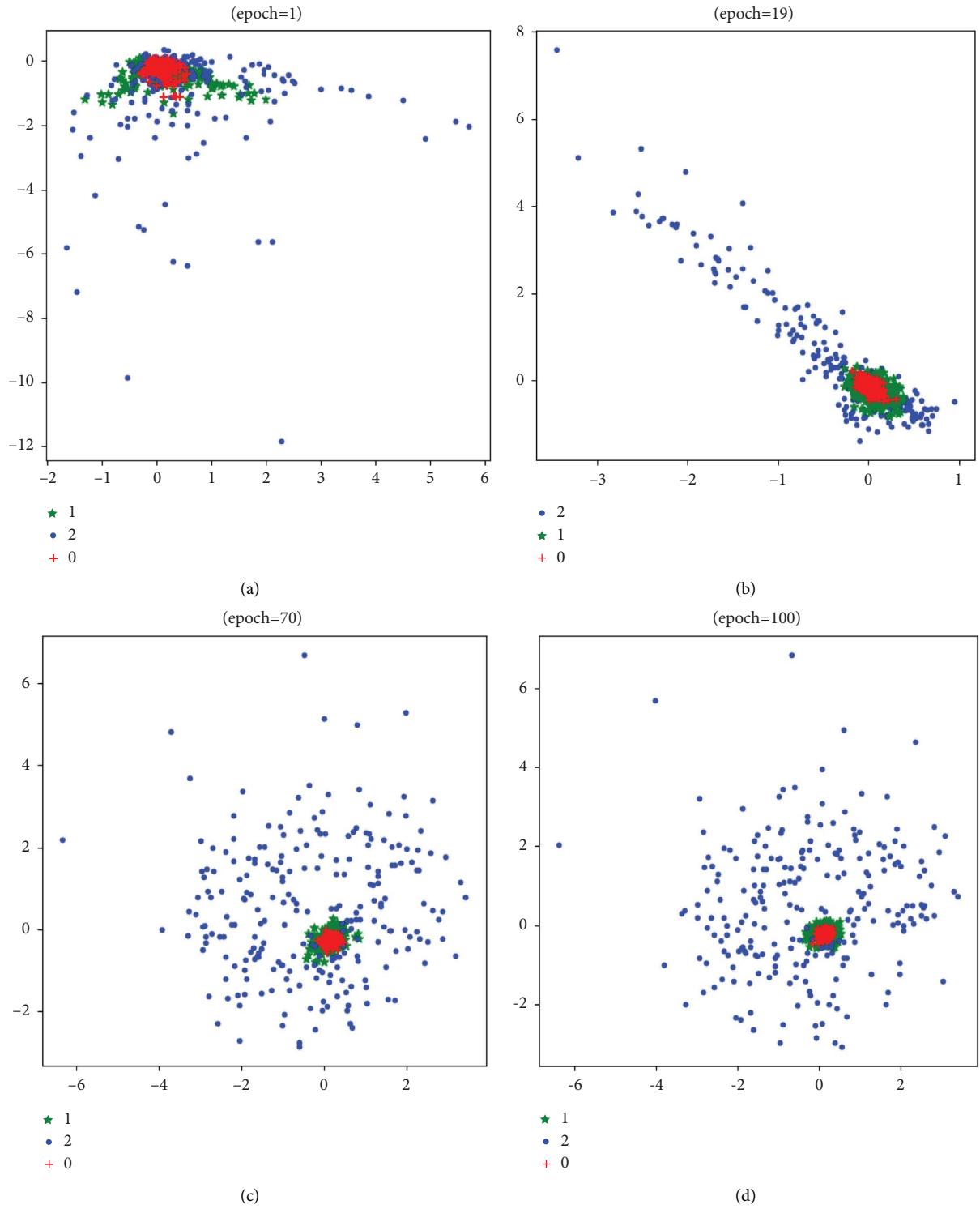


FIGURE 11: Distribution of MW-CVAE latent space. (a) Epoch = 1. (b) Epoch = 19. (c) Epoch = 70. (d) Epoch = 100.

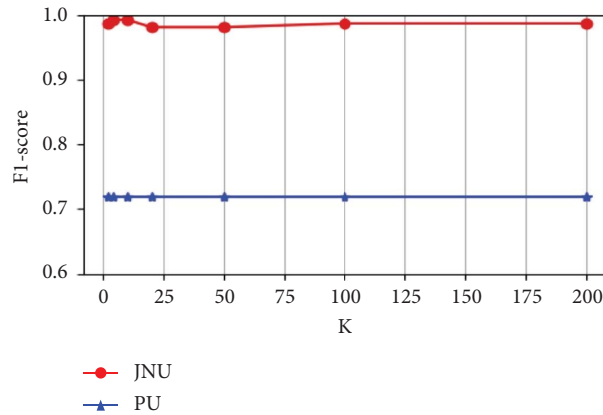


FIGURE 12: The relationship between $F1$ -score and K by MW-CAVE (average of 5 replicates).

6. Conclusion

Aiming at the problem that the accuracy of mechanical system anomaly detection is significantly reduced under multiple working conditions, an MW-CVAE is proposed in the paper. The working condition is encoded as conditional input to establish the anomaly detection model. Compared with the typical CVAE model, each working condition has a corresponding best threshold of anomaly detection in the MW-CVAE model, which improves the detection. For instance, the AUC increases by 11-12%, and the $F1$ -score increases by 18-19%. The method proposed in the paper can be applied to the anomaly detection of discrete working conditions in the industry. According to the distribution of multiple working conditions in the latent space of the MW-CVAE model, each working condition tends to the normal distribution $N(0, \text{diag}(x)I)$ after the convergence of learning, providing a basis for classification by working conditions and anomaly detection.

Data Availability

The data used in this study can be obtained from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by a Special Project of National Key Research and Development Program—Research of China on the technology of whole-life-cycle detection, monitoring and integrity evaluation of manned equipment in amusement parks and scenic spots (Project no.: 2016YFF0203100).

References

- [1] R. Chalapathy and S. Chawla, "Deep Learning for Anomaly Detection: A Survey," 2019, <https://arxiv.org/abs/1901.03407>.
- [2] M. Gupta, J. Gao, C. C. Aggarwal, and J. Han, "Outlier detection for temporal data: a survey," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 9, pp. 2250–2267, 2014.
- [3] G. Dewangan and S. Maurya, "fault diagnosis of machines using deep convolutional beta-variational autoencoder," *IEEE Trans. Artif. Intell.*, vol. 3, no. 2, pp. 287–296, 2022.
- [4] T. Li, Z. Zhao, C. Sun, R. Yan, and X. Chen, "Domain adversarial graph convolutional network for fault diagnosis under variable working conditions," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–10, 2021.
- [5] J.-C. Wu, S. Lu, C.-S. Fuh, and T.-L. Liu, "One-class anomaly detection via novelty normalization," *Computer Vision and Image Understanding*, vol. 210, Article ID 103226, 2021.
- [6] Y. Kim and H. K. Kim, "Anomaly Detection Using Clustered Deep One-Class Classification," in *Proceedings of the 2020 15th Asia Joint Conference on Information Security (AsiaJCS)*, pp. 151–157, Taipei, Taiwan, August 2020.
- [7] X. k. Wang, W. h. Hou, H. y. Zhang et al., "KDE-OCSVM model using Kullback-Leibler divergence to detect anomalies in medical claims," *Expert Systems with Applications*, vol. 200, Article ID 117056, 2022.
- [8] M. Ding and H. Tian, "PCA-based network traffic anomaly detection," *Tsinghua Science and Technology*, vol. 21, no. 5, pp. 500–509, 2016.
- [9] M. M. Breunig, H. P. Kriegel, R. T. Ng, and J. Sander, "LOF: Identifying Density-Based Local Outliers," *Procacm Sigmod Intconfon Manag. Data*, vol. 29, 2000.
- [10] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," 2014, <http://arxiv.org/abs/1312.6114>.
- [11] H. Xu, Y. Feng, J. Chen et al., "Unsupervised anomaly detection via variational auto-encoder for seasonal KPIs in web applications," in *Proceedings of the 2018 World Wide Web Conference*, pp. 187–196, ACM Press, Lyon, France, April 2018.
- [12] Z. Li, W. Chen, and D. Pei, "Robust and unsupervised KPI anomaly detection based on conditional variational autoencoder," in *Proceedings of the IEEE 37th International Performance Computing and Communications Conference (IPCCC)*, pp. 1–9, Orlando, FL, USA, June 2018.
- [13] A. A. Pol, V. Berger, C. Germain, G. Cerminara, and M. Pierini, "Anomaly detection with conditional variational autoencoders," in *Proceedings of the 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, pp. 1651–1657, IEEE, Boca Raton, FL, USA, November 2019.
- [14] Zhangqiaokeyan, "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery," 2022, https://www.zhangqiaokeyan.com/open-access_resources_thesis/0100091318667.html.

- [15] B. Zong, Q. Song, M. R. Min et al., "Deep autoencoding Gaussian mixture model for unsupervised anomaly detection," in *Proceedings of the International conference on learning representations*, p. 19, Vancouver, Canada, May 2018.
- [16] W. Chen, H. Xu, Z. Li et al., "Unsupervised anomaly detection for intricate KPIs via adversarial training of VAE," in *Proceedings of the IEEE INFOCOM 2019 - IEEE Conference on Computer Communications Workshops*, pp. 1891–1899, IEEE, Paris, France, May 2019.
- [17] J. An and S. Cho, "variational autoencoder based anomaly detection using reconstruction probability," *Special Lecture on IE*, vol. 2, no. 1, 2015.
- [18] A. Hannan, C. Gruhl, and B. Sick, "Anomaly based resilient network intrusion detection using inferential autoencoders," in *Proceedings of the 2021 IEEE International Conference on Cyber Security and Resilience*, pp. 1–7, Rhodes, Greece, July 2021.
- [19] Y. Zhou, X. Liang, W. Zhang, L. Zhang, and X. Song, "VAE-based Deep SVDD for anomaly detection," *Neurocomputing*, vol. 453, pp. 131–140, 2021.
- [20] D. Chen, Y. Qin, Y. Wang, and J. Zhou, "Health indicator construction by quadratic function-based deep convolutional auto-encoder and its application into bearing RUL prediction," *ISA Transactions*, vol. 114, pp. 44–56, 2021.
- [21] J. Zhou, Y. Qin, J. Luo, and T. Zhu, "Remaining useful life prediction by distribution contact ratio health indicator and consolidated memory GRU," *IEEE Transactions on Industrial Informatics*, vol. 34, pp. 1–11, 2022.
- [22] Y. Qin, J. Zhou, and D. Chen, "Unsupervised health indicator construction by a novel degradation-trend-constrained variational autoencoder and its applications," *IEEE*, vol. 27, no. 3, pp. 1447–1456, 2022.
- [23] Z. Zhao, Q. Zhang, X. Yu et al., "Unsupervised deep transfer learning for intelligent fault diagnosis: an open source and comparative study," 2019, <https://arxiv.org/abs/1912.12528>.
- [24] C. Lessmeier, J. K. Kimotho, D. Zimmer, and W. Sextro, "Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: a benchmark data set for data-driven classification," in *Proceedings of the PHM Society European Conference 2016*, p. 17, Bilbao, Spain, July 2016.
- [25] Y. Kawachi, Y. Koizumi, and N. Harada, "Complementary set variational autoencoder for supervised anomaly detection," in *Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2366–2370, Calgary, AB, Canada, April 2018.
- [26] H. Shao, M. Xia, G. Han, Y. Zhang, and J. Wan, "Intelligent Fault diagnosis of rotor-bearing system under varying working conditions with modified transfer convolutional neural network and thermal images," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 5, pp. 3488–3496, 2021.
- [27] H. Qin, X. Zhan, and Y. Zheng, "CSCAD: correlation structure-based collective anomaly detection in complex system," *IEEE Transactions on Knowledge and Data Engineering*, vol. 1, 2022.
- [28] J. Ramakrishnan and E. Shaabani, "Anomaly detection for an E-commerce pricing system," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Anchorage, AK, USA, August 2019.
- [29] Kaggle, "Case Western Reserve University (CWRU) Bearing Data Center," 2019, <https://www.kaggle.com/datasets/brjapon/cwru-bearing-datasets>.