IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 5, 2019, accepted September 25, 2019, date of publication September 30, 2019,
date of current version October 16, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2944689

A GAN-Based Anomaly Detection Approach for Imbalanced Industrial Time Series

WENQIAN JIANG[1], YANG HONG[2], BEITONG ZHOU[2], XIN HE[3], AND CHENG CHENG[2]
[1]China-EU Institute for Clean and Renewable Energy, Huazhong University of Science and Technology, Wuhan 430074, China
[2]School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, China
[3]School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China

Corresponding author: Cheng Cheng (c_cheng@hust.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 91748112, and in part by the Primary Research and Development Plan of Jiangsu Province under Grant BE2017002.

ABSTRACT Imbalanced time series are universally found in industrial applications, where the number of normal samples is far larger than that of abnormal cases. Traditional machine learning algorithms, such as support vector machine and convolutional neural networks, are struggling to attain high classification accuracies for class-imbalanced problems, because they tend to ensure the accuracy of the majority class. Hereby, this paper proposes a novel anomaly detection approach based on generative adversarial networks (GAN) to overcome this problem. In particular, an encoder-decoder-encoder three-sub-network generator is trained involving the elaborately extracted features from normal samples alone. Anomaly scores for anomaly detection are made up of apparent loss and latent loss. Without having any knowledge of the abnormal samples, our approach can diagnose faults by generating much higher anomaly scores when a fault sample is fed into the trained model. Experimental studies are conducted to verify the validity and feasibility of our approach, including a benchmark rolling bearing dataset acquired by Case Western Reserve University and another rolling bearing dataset which is acquired by our laboratory. Our approach can distinguish abnormal samples from normal samples with 100% accuracies on both datasets.

INDEX TERMS Anomaly detection, generative adversarial networks, imbalanced industrial time series, rolling bearings.

NOMENCLATURE

AUC	The area under curve of the receiver operating characteristic.
B	Faults at the rolling element.
CNN	Convolution neural networks.
CWRU	Case Western Reserve University.
DCGAN	Deep convolutional generative adversarial networks.
DBN	Deep belief network.
IR	Faults at the inner raceway.
GAN	Generative adversarial networks.
LSTM	Long short term memory networks.
OR	Faults at the outer raceway.
RNN	Recurrent neural networks.
ROC	Receiver operating characteristic.
SVM	Support vector machine.
SMOTE	Synthetic minority oversampling technique.
SNR	Signal to noise ratio.
R	The set of real numbers.
$\mathbf{D_{train}}$	Training set.
D_{test}	Testing set.
D^{v}_{test}	Testing set under normal conditions.
D^{u}_{test}	Testing set under abnormal conditions.
X_i	A time series sample recorded by sensor.
$\mathcal{M}$	The trained model.
F	Original feature matrix.
$\hat{F}$	Regenerated feature matrix.
f_i	An extracted feature.
Z	Original latent vector.
$\hat{Z}$	Regenerated latent vector.
G_e	The first encoder in generator.
G_d	The decoder in generator.
$\hat{G}_e$	The second encoder in generator.
L_f	Fraud loss.
L_a	Apparent loss.
L_l	Latent loss.

The associate editor coordinating the review of this manuscript and approving it for publication was Yungang Zhu.

σ Binary cross entropy loss function.
A Anomaly score.
λ Weighted parameter of the anomaly score.

I. INTRODUCTION

Anomaly detection is of utmost importance for the reliability and safety of modern industrial systems [1]. Timely and accurate anomaly detection can help to prevent fatal accidents and increase manufacturing efficiency [2]. However, class-imbalanced observed data, where samples under normal conditions are much more prevalent than those under abnormal conditions, set tremendous obstacles in diagnosing industrial faults precisely [3]. In addition, industrial systems always feature nonlinearities and uncertainties [4], which poses great challenges to model training.

Generally, industrial anomaly detection is based on physical signals recorded by diversified sensors over a certain duration, such as current and temperature, which is also known as time series. For industrial anomaly detection [5], time series usually acts as the input data with which to train models. Taking the time series as input, a common anomaly detection framework often consists of two stages: feature extraction and fault recognition [6], [7]. Through feature extraction algorithms [7], time series is preprocessed to low dimensional feature vectors, which are fed into fault detector for fault detection. As a powerful pattern recognition tool for anomaly detection, machine learning algorithms have become the focus of attention [1], including Bayesian classifier [8], support vector machine (SVM) [9], [10], neural networks [10], [11], and deep learning methods [12]. However, above methods are all struggling to attain high classification accuracies for imbalanced data because they are based on a class-balanced hypothesis [13].

With the exception of the class-balanced hypothesis on datasets, labeled data are essential for machine learning algorithms in training stage. However, in many practical industrial systems, samples from abnormal operating conditions are often of insufficient size. In addition, when a system runs under the normal condition for a long period followed by a sudden abnormality, it is extremely hard to precisely locate the starting time of the abnormality. Consequently, inaccurate abnormal labels will also have an adverse impact on the former imbalanced situation. When normal and abnormal labels are imbalanced, classifiers of machine learning methods will ensure the accuracy of the majority classes by sacrificing the minority classes [13], which means the diagnostic results will bias towards normality for all testing samples. However, for anomaly detection in industrial systems, our focus should be specifically on those minority classes.

Recently, the development of generative adversarial networks (GAN) proposed by Goodfellow *et al*. offers a new perspective for the class-imbalanced problem. It was initially proposed in [14] for image recognition. The basic idea of GAN is that it generates prototypical samples through a generator with random data points that satisfy a certain distribution (*e.g*. Gaussian distribution). In the field of image anomaly diagnosis, some competitive GAN-based network architectures have been designed accordingly, such as AnoGAN [15], BiGAN [16], and GANomaly [17]. These GAN-based methods train models only with normal images, distinguishing the abnormal images according to the distribution difference between the normal and abnormal. In this sense, these GAN-based models are effective for unbalanced datasets to prevent diagnostic results bias towards normality. However, regarding industrial applications, GAN-based methods for anomaly detection are rarely seen. [18], [19] propose GAN-based networks for mechanical faults. [20] proposes a GAN-based method for imbalanced fault diagnosis. [21] demonstrates the feasibility of using GAN-based networks to detect cyber-attacks for cyber-physical systems. Above researches inspire us to further investigate the effectiveness of GAN regarding the problem of industrial fault detection, especially for the cases without anomalous data. Therefore, based on GANomaly [17], we place emphasis on the characteristics of industrial time series, and improve the overall generator loss to achieve high accuracy for class-imbalanced scenarios.

In this work, specifically for the imbalanced industrial data, we propose a GAN-based approach to address the problem of intelligent anomaly detection. A generator and a discriminator are trained for diagnosis. The generator employs an encoder-decoder-encoder three-sub-network, based on deep convolutional generative adversarial networks. To reduce training time and increase diagnosis performance, a feature extractor is inserted between the original data and GAN. Experiments on rolling bearing datasets are conducted to verify the validity and feasibility of the approach. Meanwhile, comparison results with another three networks, and classic algorithms for solving class-imbalanced problems reveal that our approach has the superior anomaly detection performance.

The main contributions of this paper are as follows: 1) For the imbalanced time series in an industrial field, a novel GAN-based approach for anomaly detection is proposed. 2) To reduce data dimension and computing time, a well-designed feature extractor is designed to help train the whole network. 3) The proposed network only requires normal samples. This is a more realistic network than other existing ones, as abnormal samples are often of insufficient size in real industrial scenarios. 4) Our idea is the first to promise diagnosis performance on a rolling bearing benchmark dataset. Additionally, a new rolling bearing dataset acquired by our own laboratory is employed to further verify its effectiveness. Both diagnosis results show that our approach achieves the best performance compared to another three networks, and classic algorithms for solving class-imbalanced problems.

II. RELATED WORKS

Anomaly detection has long been a question of great interest in industrial systems. A considerable amount of works have been published to propose efficient theories and algorithms.

A. DEEP LEARNING FOR INTELLIGENT FAULT DIAGNOSIS

With the increasingly dynamic, complex and huge quantity of time series streams generated in various systems of industrial production, many anomaly detection techniques used in deep learning have been well developed [22]. They are purpose-free, which means that they are not clear about what the final output features will be. These feature black box modes are aimed to extracting specific patterns for specific datasets, such as long short term memory networks (LSTM) [23], recurrent neural networks (RNN) [24], convolution neural networks (CNN) [25], and autoencoders [26]. Although many state-of-the-art anomaly detection performances have been acquired by above deep learning models, their performances are still not satisfactory when faced with imbalanced datasets. In addition, due to differences in data representations (images and time series), many deep learning models are applicable to image fields while they are often difficult to apply in the industrial fields.

B. CLASS IMBALANCED PROBLEMS

Based on imbalanced time series for anomaly detection, two key methods are usually considered: data-level and algorithm-level methods [3]. Data-level methods [27], [28] utilize sampling policies to change imbalanced data distribution, where under-sampling and over-sampling [29] have been widely used. Algorithm-level methods [27] are used to change the classifier to fit imbalanced data, where bagging and boosting ensemble-based methods [27] are commonly applied. More specifically, EasyEnsemble [3] and BalanceCascade [30] algorithms are proposed to deal with the problem of class-imbalance problem. The synthetic minority oversampling technique (SMOTE) [31], [32] is a synthetic technique that can add new minority class examples. To overcome both binary and multi-class imbalance problems, [3] proposes the Easy-SMT ensemble algorithm based on synthesizing SMOTE-based data augmentation policy and EasyEnsemble algorithm.

C. GAN AND GANOMALY

Recently, adversarial training, especially GAN, occupies an increasingly pivotal position for anomaly detection of class-imbalanced images. GAN, initially introduced by Goodfellow *et al*. and viewed as an unsupervised machine learning algorithm, has achieved outstanding application effects in the field of image recognition. Based on GAN, various kinds of adversarial algorithm have emerged. For further details, we refer interested readers to [33], which gives a very comprehensive summary of GAN and its variants.

In the industrial field, Lim *et al*. proposes a novel GAN-based anomaly detection method combining GAN with LSTM-RNN to detect cyber-attacks for cyber-physical systems; a data augmentation technique based on GAN and focused on improving performances in anomaly detection is put forward [34]. Moreover, a generic anomaly detection architecture called GANomaly put forward by Akcay *et al*.

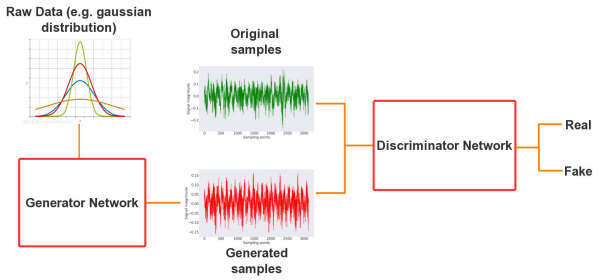


FIGURE 1. Overview of GAN.

in [17] shows superiority and efficacy compared with previous state-of-the-art approaches over several benchmark image datasets, which prompts us to adopt the technique for anomaly detection in the industrial field. The following is a brief introduction to GANomaly. Samet *et al*. employs an encoder-decoder-encoder sub-network in a generator network to train a semi-supervised network; this architecture uses deep convolutional generative adversarial networks (DCGAN) [35] and employs three loss functions in the generator to capture distinguishing features in both input images and latent space. One of the features of the algorithm is that it considers no negative samples, and it achieves state-of-the-art performances for anomaly detection in some benchmark image datasets.

The rest of the paper is organized as follows. Section III proposes our GAN-based anomaly detection framework. Experimental setup and results are described in Section IV and Section V respectively. Finally, conclusion and future work are drawn in Section VI.

III. METHODOLOGY

A. PROBLEM FORMULATION

Given a time series dataset $\mathcal{D} = [X_1, X_2, \ldots, X_n] \in \mathbb{R}^{t\times n}$, where $X_i = [x_{1i}, x_{2i}, \ldots, x_{ti}]^T \in \mathbb{R}^t$ (for $i = 1, \ldots, n$) represents a sample recorded by sensors over a period of time t. The proposed model is trained with a subset of the dataset $D_{train} \in \mathbb{R}^{t\times b}$ with only normal samples, where b is the number of training samples. Corresponding samples for testing are $D_{test} = [D^v_{test}, D^u_{test}] \in \mathbb{R}^{t\times(v+u)}$, where v and u are the number of normal and abnormal samples, respectively, and $n = b + v + u$. For imbalanced time series data, the size of normal samples is much greater than that of abnormal samples, i.e. $b + v \gg u$.

In the training stage, a GAN-based model $\mathcal{M}$ is trained with D_{train}. The objective of training process is to minimize the output of $\mathcal{M}$ for each $X_i \in D_{train}$. After training, D_{test} will be fed into the trained model $\mathcal{M}$. The trained generator will encode and decode fault and normal samples accordingly. Since the trained network only explores possible representation modes of normal data, with abnormal samples D^u_{test} as inputs, the outputs of $\mathcal{M}$ will deviate largely compared to the outputs with normal inputs D^v_{test}. This deviation ultimately help us to determine the existence of abnormal samples.

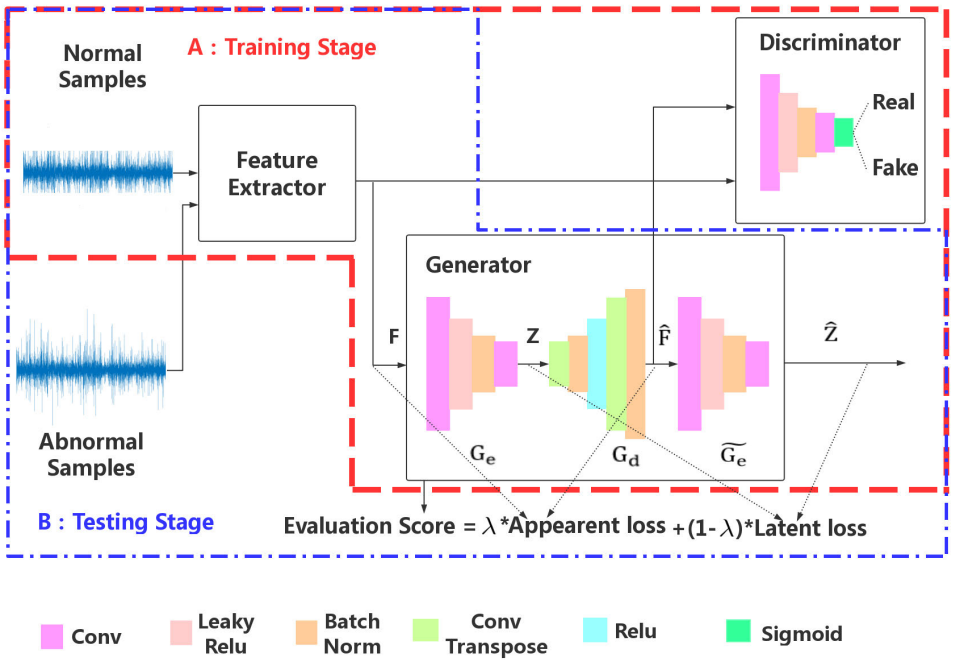


FIGURE 2. Overview of our proposed training procedure. In the training stage, only normal samples are taken into consideration. We use feature extractor to extract useful features which are carefully selected to avoid directly feed large dataset into the network. The generator which adopts encoder-decoder-encoder three-sub-network and the discriminator are based on DCGAN. In the testing stage, abnormal samples can be discriminated by higher anomaly scores compared to normal samples.

B. TRAINING PIPELINE

As shown in Figure 2, the network structure of our approach consists of three parts: a feature extractor, a generator, and a discriminator. The basic network architecture is denoted by DCGAN.

In the design of the generator, an encoder-decoder-encoder three-sub-network is developed. Prior to feeding data into the generator, a well-designed feature extractor is designed to extract distinguishing features and to help reduce the training time. All extracted features originate from [36]–[39]. Based on the principle of universality, we elaborately select 16 most representative features which any time series can be encapsulated into. During the training process, we first extract features of the time series just under normal conditions, and then obtain data distributions and potential representative modes of these features by our designed anomaly detector. In the testing phase, anomalies will be diagnosed by higher scores from the trained anomaly detector.

1) FEATURE EXTRACTOR

With the given training dataset D_{train}, which also can be written into the following matrix form:

$$D_{train} = [X_1, X_2, \ldots, X_b] = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1b} \\ x_{21} & x_{22} & \ldots & x_{2b} \\ \ldots & \ldots & \ldots & \ldots \\ x_{t1} & x_{t2} & \ldots & x_{tb} \end{bmatrix}. \quad (1)$$

To reduce the size of D_{train} and save computational time, feature extractor is employed to explore the most representative q features from D_{train}, feature matrix F is given by:

$$F = [f_1, f_2, \ldots, f_b] = \begin{bmatrix} f_{11} & f_{12} & \ldots & f_{1b} \\ f_{21} & f_{22} & \ldots & f_{2b} \\ \ldots & \ldots & \ldots & \ldots \\ f_{q1} & f_{q2} & \ldots & f_{qb} \end{bmatrix}, \quad (2)$$

where $F \in \mathbb{R}^{q \times b}$, and each element in f_i represents a extracted feature.

2) GENERATOR AND DISCRIMINATOR

For the generator, the two encoders G_e and $\hat{G}_e$ learn to acquire representations of original features F and regenerated features $\hat{F}$ respectively. The decoder G_d tries to reconstruct $\hat{F}$ at the same time. The whole process is as follows:

a) G_e consists of convolutional layers followed by batch-norm layers and leaky *ReLU* activation layers. G_e downscales F into latent representation Z (usually $h < q$)

$$Z = G_e(F), Z \in \mathbb{R}^{h \times b}, \quad (3)$$

where h is the dimension of latent vector Z.

b) G_d adopts convolutional transpose layers, *ReLU* activation layers and batch-norm layers. G_d uses Z to recreate $\hat{F}$

$$\hat{F} = G_d(Z), \quad (4)$$

where the dimension of $\hat{F}$ is the same as F.

c) Finally, the architecture of last encoder $\hat{G}_e$ is the same as G_e but with different parametrization, and the dimension of output $\hat{Z}$ is the same as Z

$$\hat{Z} = \hat{G}_e(\hat{F}). \tag{5}$$

The generator guarantees that both the characteristics of the original feature set F, and the pattern of the latent vector Z can be learned at the same time.

The discriminator C adopts the standard discriminator network introduced in DCGAN, which is used to distinguish whether input data is real or generated.

Having defined our overall network architecture, we now continue to discuss how we define loss functions for training.

3) OBJECTIVE FUNCTIONS

In the training phase, because only the normal set D_{train} is considered, the model $\mathcal{M}$ only obtains normal patterns. But in the testing phase, $\mathcal{M}$ needs to determine fault samples by outputting higher abnormal scores. That means G_d and $\hat{G}_e$ will decode Z and re-encode $\hat{F}$ similar to the patterns acquired in the training stage. $\hat{F}$ and $\hat{Z}$ will inevitably deviate from the original F and Z, so that help us identify faulty.

Since the encoder-decoder-encoder three-sub-network is adopted in the generator, the final loss function of the generator consists of three parts: fraud loss, apparent loss, and latent loss.

Fraud Loss : The fraud loss L_f is aimed to induce the discriminator to misjudge generated samples from generator as real industrial samples. We compute the fraud loss L_f on the discriminator output by feeding the generated samples into the discriminator, and the formula is as follow:

$$L_f(F) = \sum_{i=1}^{N} \sigma(C(\hat{F}), \alpha), \tag{6}$$

where σ is the binary cross-entropy loss function, and $C(\hat{F})$ is the probability that the sample i is predicted to be real. To fool discriminator and to promote the generator to generate fake samples as real samples as possible, we define $\alpha = 1$.

Apparent Loss : It is not enough for the generator to learn potential patterns under normal conditions and to reconstruct generated samples as realistic as possible, so we define the apparent loss L_a by measuring L_1 distance between the real and generated samples, and the formula is as follow:

$$L_a(F) = \sum_{i=1}^{N} \left\| F - \hat{F} \right\|. \tag{7}$$

Latent Loss : In addition to the fraud loss and the apparent loss, we also define the latent loss to minimize the distance between latent representations of real samples and encoded bottleneck features of generated samples. This loss can help to learn latent representations both in real and fake examples. Moreover, the latent loss will form final abnormal scores in the testing phase with the apparent loss

$$L_l(F) = \sum_{i=1}^{N} \|G_e(F) - G(F)\|_2 . \tag{8}$$

In summary, the loss function of the generator is:

$$L_g(F) = \omega_f * L_f(F) + \omega_a * L_a(F) + \omega_l * L_l(F), \tag{9}$$

where ω_f, ω_a and ω_l are used to adjust the importance of L_f, L_a and L_l in generator loss.

For the discriminator, a feature matching loss $\mathcal{L}$ is adopted for adversarial learning, which is proposed by Salimans et al. [40] to reduce the instability of GAN training

$$L_d(F) = \sum_{i=1}^{N} \|\mathcal{L}(F) - \mathcal{L}(G_d(G_e(F)))\| . \tag{10}$$

We use Adam [41] optimizer to update equations (9) and (10) using backward propagation.

C. TESTING STAGE

In the testing phase, our model uses the latent loss and the apparent loss for scoring the abnormality of a given sample. Anomaly score is defined as

$$A(F) = \lambda * L_a(F) + (1 - \lambda) * L_l(F). \tag{11}$$

In this part, we use the ratio of ω_a to ω_l which performs the best training result (i.e., with the smallest generator loss and discriminator loss) as the suitable weighted hyper-parameter λ. Because we only train our anomaly detector on normal data, our anomaly detector will only capture normal latent patterns and data distributions, $A(F)$ will be close to 0 for a normal sample, while it will be abnormally large for a fault sample. Then, we can easily find abnormalities by the value of $A(F)$.

D. GENERAL PROCEDURE OF THE PROPOSED APPROACH

The general outline of our approach for anomaly detection of imbalanced industrial datasets are summarized in Algorithm 1.

IV. EXPERIMENTAL SETUP

In order to evaluate the feasibility and effectiveness of our approach, we first test it on rolling bearing data from Case Western Reserve University (CWRU) [42], and then further validate it by using a rolling bearing dataset collected from our laboratory [43].

A. DATASET DESCRIPTION

Rolling bearing data from CWRU [42]. This is a benchmark bearing anomaly detection dataset measuring a vibration signal using an accelerometer on a 2 hp reliance electric motor. Motor bearings were seeded with faults using electro-discharge machining. Faults ranging from 0.007 to 0.028 inch in diameter were created respectively at the inner raceway (IR), the rolling element (*i.e.* ball) (B), and

Algorithm 1 Outline of Our Approach for Imbalanced Industrial Dataset

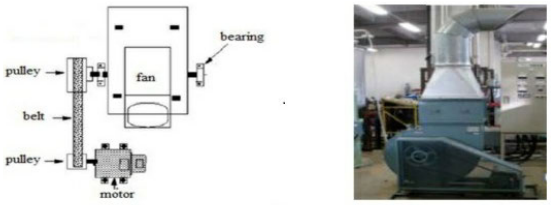
Input: Training set $\mathbf{D_{train}}$ (only normal samples).
Initialize: Initialize all parameters of the whole network.
Output: Trained model $\mathcal{M}$.
Training :
1: $\mathbf{F_{train}} \leftarrow$ Input $\mathbf{D_{train}}$ into feature extractor.
2: Forward Propagation as following:
  2.1: $\mathbf{Z} \leftarrow$ Downscale $\mathbf{F_{train}}$ using Eq. (3).
  2.2: $\hat{\mathbf{F}} \leftarrow$ Upscale $\mathbf{Z}$ using Eq. (4).
  2.3: $\hat{\mathbf{Z}} \leftarrow$ Downscale $\hat{\mathbf{F}}$ using Eq. (5).
  2.4: $\mathbf{L_g} \leftarrow$ Use Eq. (6)-(9).
  2.5: $\mathbf{L_d} \leftarrow$ Use Eq. (10).
3: Backward Propagation:
  Update network parameters using *Adam* [41] optimizer.
4: Trained model $\mathcal{M} \leftarrow$ Repeat **Step 2** and **Step 3** until $\mathbf{L_g}$ and $\mathbf{L_d}$ converge.
Testing :
1: $\mathbf{F_{test}} \leftarrow$ Input $\mathbf{D_{test}}$ into feature extractor.
2: $\mathbf{L_a}$ and $\mathbf{L_l} \leftarrow$ Input $\mathbf{F_{test}}$ into trained generator of $\mathcal{M}$.
3: $\mathbf{A} \leftarrow$ Use Eq. (11).

FIGURE 3. Testing bed of CWRU.

the outer raceway (OR). Fault bearings were reinstalled into the test motor and vibration data was recorded for motor loads from 0 to 3 horsepower (motor speeds from 1720 to 1797 rpm). Datasets were collected at 12 kHz and at 48 kHz, respectively. Each dataset file consists of three types of signal, namely the accelerometer signal at the drive end, the accelerometer signal at the fan end, and the base accelerometer signal. Testing bed of CWRU is shown in Figure 3.

Rolling bearing data from our laboratory [43]. Our experiment uses normal and faulty bearings installed into the testing motor whose voltage signals are acquired. Fault bearings with a 0.055 inch fault diameter are damaged using electro-discharge machining. As in the CWRU dataset, faults are introduced to B, OR and IR respectively. Signals are collected with 50 kHz sampling frequency undergoing three motor speeds (i.e., 600 rpm, 800 rpm, and 1000 rpm). To ensure the validity of the recorded experimental data, each operational condition is repeated three times with the same experimental setting. Our test rig is shown in Figure 4.

Details of two datasets are summarized in Table 1. To examine the robustness of our proposed GAN-based method against environment noise, a small load is applied to the bearing on our testing bed to obtain lower signal to noise ratio (SNR) signals for evaluation.



(a) Configuration (b) Real platform

FIGURE 4. Our experimental test rig.

TABLE 1. Details of two rolling bearing datasets.

	CWRU	Our Lab
Signal	vibration	voltage
Assembly difficulty	hard	easy
SNR	high	low
Frequency (kHz)	12/48	50
Motor speed (rmp)	1730, 1750, 1772, 1797	600, 800, 1000
Fault diameter (inch)	0.007, 0.014, 0.021,0.028	0.055
Fault type	IR, OR, B	IR, OR, B

B. DATA PRE-PROCESSING

We divide normal samples into a training set and a testing set respectively. For rolling bearing dataset from CWRU, the sample dimension is set to $t = 3136$. And for the rolling bearing dataset from our lab, $t = 12000$.

To make our results comparable to most published results for the benchmark CWRU dataset, we use the original data without feature extraction for comparison. For the CWRU dataset, we use $b = 400$ samples under normal operation condition to train the anomaly detector. $v = 541$ and $u = 384$ samples without labels are used in testing set.

For the rolling bearing data from our lab, sixteen distinguishing features are extracted ($q = 16$). Expressions of these features are listed in Table 2. We use $b = 100$ samples under normal operation condition to train the anomaly detector. $v = 375$ and $u = 225$ samples without labels are used in the testing set.

C. IMPLEMENTATION DETAILS

1) COMPARE WITH DBN, ANOGAN, AND BIGAN

For comparison, other three networks are employed: namely deep belief network (DBN) [44], AnoGAN [15], and bidirectional generative adversarial networks (BiGAN) [16]. Specifically, DBN is a multilayer generative model which is composed by Restricted Boltzmann Machines (RBMs), which is efficient for extracting high-dimensional temporal datas. AnoGAN uses an anomaly score to indicate how samples fit the learned distribution. Anomalous data can be correctly identified with a high probability by using this method. BiGAN learns the inverse mapping which means the data is projected back into the latent space. The learned feature

TABLE 2. Extracted features.

Feature	Equation
Maximum Value	$f_{1i} = max(X(i))$
Mean Value	$f_{2i} = \frac{1}{N}\sum_{i=1}^{N} X(i)$
Minimum Value	$f_{3i} = min(X(i))$
Standard Value	$f_{4i} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X(i) - f_{1i})^2}$
Peak to Peak Value	$f_{5i} = f_{1i} - f_{3i}$
Mean Amplitude	$f_{6i} = \frac{1}{N}\sum_{i=1}^{N} \lvert X(i)\rvert$
Root Mean Square Value	$f_{7i} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} X(i)^2}$
Skewness Value	$f_{8i} = \frac{1}{N}\sum_{i=1}^{N} X(i)^3$
Kurtosis Value	$f_{9i} = \frac{1}{N}\sum_{i=1}^{N} X(i)^4$
Waveform Indicator	$f_{10i} = \frac{f_{7i}}{f_{6i}}$
Pulse Indicator	$f_{11i} = \frac{f_{1i}}{f_{6i}}$
Kurtosis Index	$f_{12i} = \frac{f_{9i}}{f_{7i}}$
Peak Index	$f_{13i} = \frac{f_{1i}}{f_{7i}}$
Square Root Amplitude	$f_{14i} = (\frac{1}{N}\sum_{i=1}^{N}\sqrt{\lvert X(i)\rvert})^2$
Margin Indicato	$f_{15i} = \frac{f_{1i}}{f_{14i}}$
Skewness Indicator	$f_{16i} = \frac{f_{8i}}{f_{7i}^4}$

TABLE 3. Network architecture of our approach.

Layer	Filters	Ks/S[1]	Output size
Encoder			
Input	-	-	4x4x1
Conv (LeakyReLU/Batchnorm)	4	2/1	3x3x4
Conv	8	2/1	2x2x8
Decoder			
Input	-	-	2x2x8
ConvTran (ReLU)	-	2/1	3x3x4
ConvTran (Tanh)	-	2/1	4x4x1
Discriminator			
Input	-	-	4x4x1
Conv (LeakyReLU/Batchnorm)	4	2/1	3x3x4
Conv (LeakyReLU/Batchnorm)	8	2/1	2x2x8
FC (Sigmoid)	-	-	1

[1] Kernel size/Stride

representation is widely used for anomaly detection. We train DBN under the condition that the ratio of the number of normal samples to abnormal samples is 5:1, while AnoGAN and BiGAN are trained only with normal samples.

We implement our approach in PyTorch [45] by optimizing the networks using Adam [41]. Initial learning rate $lr = 1e^{-4}$, and momentums $\beta_1 = 0.5$, $\beta_2 = 0.999$ are used. Each model is trained 50 epochs with 16 batch-size for both datasets. Detailed network architecture of our approach is shown in Table 3. And, detailed network architectures of AnoGAN and BiGAN, are presented in Table 6 and Table 7 in Appendix. In addition, for hyper-parameters in equation (9), we first consider grid search and cross validation using the method of GridSearchCV [46]. Afterwards, we made fine adjustments with trial and error method based on the results of GridSearchCV, and determined the best parameters according to the minimum generator loss value L_g.

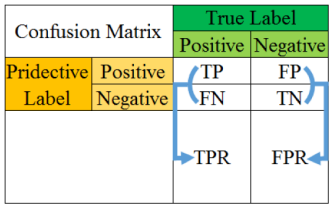


FIGURE 5. Confusion matrix.

2) COMPARE WITH CLASSICAL ALGORITHMS FOR SOLVING UNBALANCED DATASET

Different from our approach, classic algorithms for solving unbalanced dataset, such as SMOTE [28] and ADASYN [47], aim to obtaining a balanced samples by various sampling strategies. To further verify the feasibility, AUC acquired by classic algorithms including SMOTE [28], Borderline SMOTE [48], ADASYN [47], SMOTEENN [49], SMOTE-Tomek [49], and RandomOverSample [50] are compared with that of our proposed method. Conveniently, we selected the most common SVC and random forest as classification methods.

In addition, to measure imbalance degree, an unbalanced factor is defined as the ratio of the number of normal samples to abnormal samples. The larger the unbalanced factor, the smaller the number of abnormal samples in both the training and testing sets. We implement above classic algorithms under two different unbalanced factors (5:1 and 20:1) and get AUC results.

D. EVALUATION METRICES

To verify if the proposed approach will achieve good performance, the anomaly scores of testing set will be calculated.

After calculating anomaly scores, the overall performance will be compared by using the area under curve (AUC) of the receiver operating characteristic (ROC). In ROC, the true positive rate (TPR) is as a function of the false positive rate (FPR). TPR and FPR are defined as follows,

$$TPR = \frac{TP}{TP + FN}, \tag{12}$$

$$FPR = \frac{FP}{FP + TN}, \tag{13}$$

where TP is the number of positive samples predicted to be positive. FN is the number of positive samples predicted to be negative. FP is the number of negative samples predicted to be positive, and TN is the number of negative samples predicted to be negative. Above four numbers also construct the confusion matrix (Figure 5), which is one of the evaluation indexes of various classification models.

V. EXPERIMENTAL RESULTS

A. RESULTS OF THE TWO DATASETS

1) RESULTS OF ROLLING BEARING DATASET FROM CWRU

Scores of CWRU dataset of the propsoed method are shown in Figure 6. As can be seen from this picture, normal and three

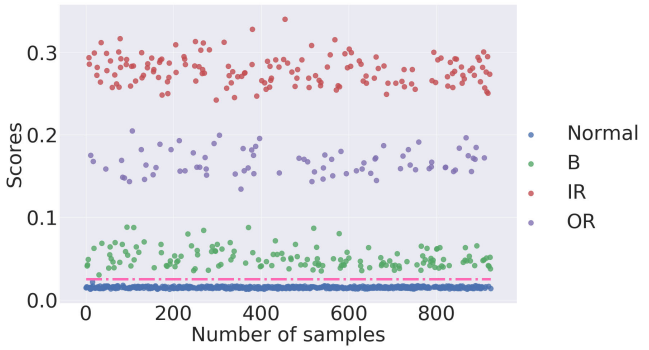


FIGURE 6. Different types of faulty on the CWRU dataset. Different levels of scores represent different states of rolling bearings (Normal, IR, OR, B). Pink dotted line separates the normal and abnormal samples.

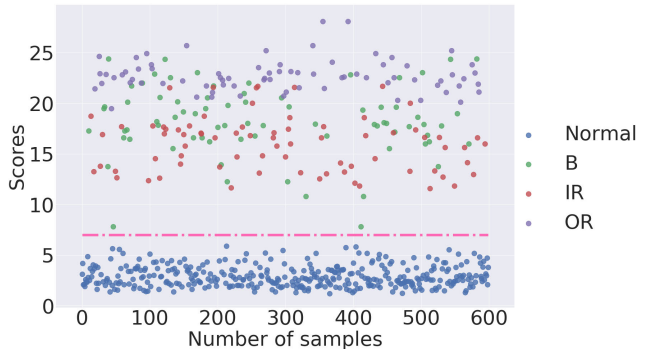


FIGURE 7. Anomaly detection performance on dataset from our lab. Pink dotted line separates the normal and abnormal samples.

fault conditions of rolling bearings are clearly separated with different levels of scores. Additionally, different types of failures can be surprisingly recognized (which is not our initial objective in this study). Pink dotted line seperates normal and abnormal samples. Moreover, the results also show that the data pattern of ball faults (B in Figure 6) are closest to that of normal, while the inner race faults (IR in Figure 6) are the most distinguished from the normality.

2) RESULTS OF THE ROLLING BEARING DATASET FROM OUR LABORATORY

For this dataset, we focus on fault signals acquired with lower SNR. Therefore, the robustness of our proposed method can be investigated with higher noise level compared to the previous case. The results are presented in Figure 7. It clearly show that the fault samples (above the pink dotted line) and normal samples are well separated, and fault samples obtain much higher anomaly scores than normal samples, compared to the results in Figure 6. Thus, we confirm that the proposed method achieves excellent classification results using our own rolling bearing dataset even the signals are highly noised.

However, although the data distribution of faulty in outer race still farthest deviate from normal samples, anomaly scores of three kinds of faulty are randomly mixed together. The future study will explore more complex GANomaly method to achieve multi-class diagnosis.

B. COMPARE WITH OTHER METHODS

1) COMPARE WITH DBN, ANOGAN, AND BIGAN

To ensures a fair comparison across different methods, hyper-parameters and architectures of each method are well tuned to obtain its best diagnosis performance for each dataset. AUC of each methods are shown in Table 4.

TABLE 4. AUC results for different networks on two datasets.

Networks	CWRU[1]	OUR LAB[2]
DBN[44]	0.5544	0.6409
AnoGAN[15]	0.9989	0.8142
BiGAN[16]	0.9977	0.8058
Our Approach	**1.000**	**1.000**

[1] W/O feature extractor
[2] W/ feature extractor

For the dataset from CWRU, AUCs of three GAN-based methods are almost the same, yet our approach shows a completely correct anomaly detection results. For our own dataset, compared to other approaches, our architecture using the model trained only with normal sample lead to the superior performance (AUC=1).

2) COMPARE WITH CLASSIC ALGORITHMS

Table 5 shows AUC results of classic algorithms under different unbalanced factors. From this table, we can conclude that classical methods for unbalanced datasets can not achieve the AUC of 1 on both datasets, especially worse results are carried out on the dataset collected from our laboratory when compared to our proposed approach. Although the accuracy of Borderline SMOTE and SMOTEENN for the dataset from CWRU can reach AUC=1 under the condition that the classification algorithm is SVC with an unbalanced factor of 5:1, when the unbalanced factor increases to 20:1, the accuracy decreases. Furthermore, the precondition for these classical algorithms is that abnormal samples are necessary in both the training and testing sets, indicating that models can not be trained without abnormal samples. However, our approach only need to be trained on normal samples, meanwhile, 100% accurate anomaly detection can be achieved on the both datasets.

C. RESULTS ANALYSIS

1) DATA VISUALIZATION

The proposed method aims to training the model that generates scores as small as possible for samples under normal conditions. To explore the effectiveness of the trained model, it is necessary to visualize the difference between the scores of normal and fault samples. We randomly select a normal sample and three fault samples in the CWRU dataset for investigation. Figure 8 shows the temporal vibrations of raw samples X (green lines) and reconstructed samples $\hat{X}$ (red lines) of each condition. It can be observed that the temporal vibrations of normal samples in Figure 8(a)

TABLE 5. AUC results for classic algorithms on two datasets.

Algorithms	CWRU[1]				OUR LAB[2]			
	5:1		20:1		5:1		20:1	
	SVC	RandomForest	SVC	RandomForest	SVC	RandomForest	SVC	RandomForest
SMOTE[28]	0.9984	0.8950	0.9655	0.7425	0.9099	0.8875	0.9945	0.9925
Borderline SMOTE [48]	1.0000	0.7850	1.0000	0.6475	0.9207	0.8975	0.9935	0.9450
ADASYN [47]	0.9668	0.7950	0.9565	0.6875	0.8960	0.8850	0.9780	0.8875
SMOTEENN [49]	1.0000	0.8300	0.9240	0.7400	0.9043	0.8950	0.9855	0.9850
SMOTETomek [49]	0.9998	0.8750	0.9835	0.6775	0.9063	0.8800	0.9965	0.9450
RandomOverSample [50]	0.9652	0.8200	0.8785	0.5475	0.9108	0.8900	0.9950	0.9450

[1] W/O feature extractor
[2] W/ feature extractor

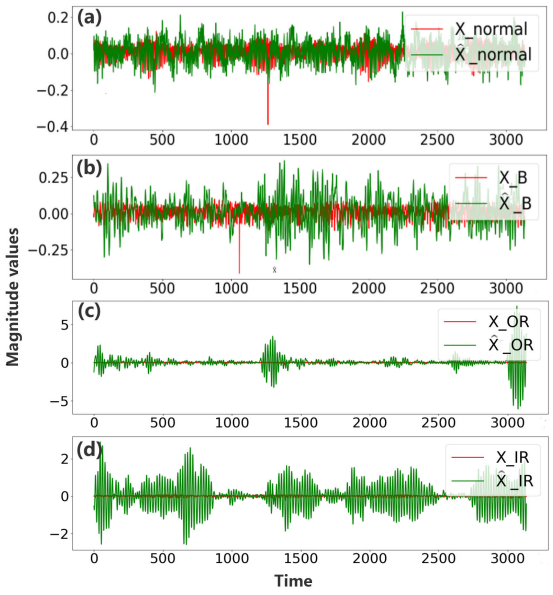


FIGURE 8. Examples of temporal vibrations of raw sample X and re-engineered sample $\hat{X}$ of (a) normal condition, (b) ball fault, (c) outer race fault, and (d) inner race fault on the CWRU dataset.

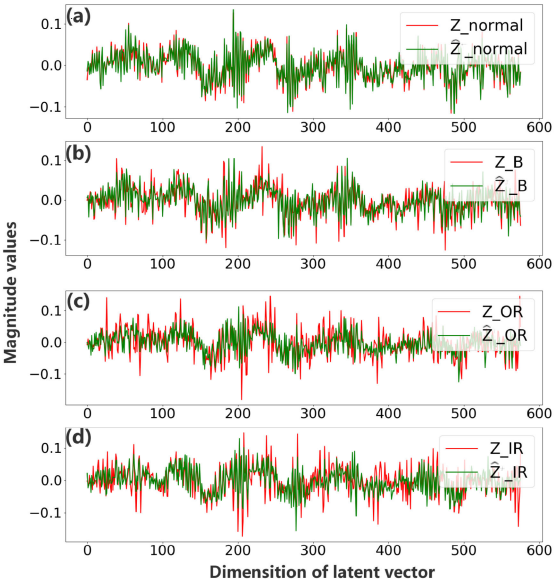


FIGURE 9. Examples of temporal vibrations of raw latent vector Z and reengineered latent $\hat{Z}$ of (a) normal condition, (b) ball fault, (c) outer race fault, and (d) inner race fault on CWRU dataset.

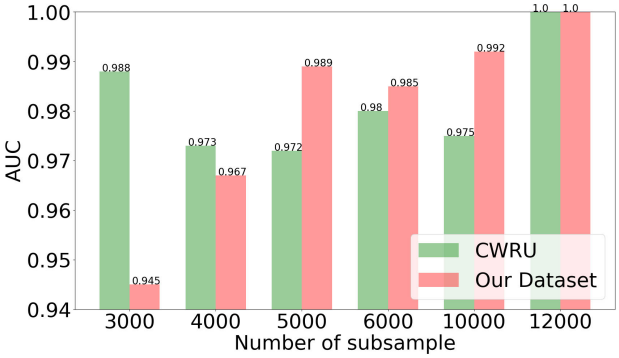
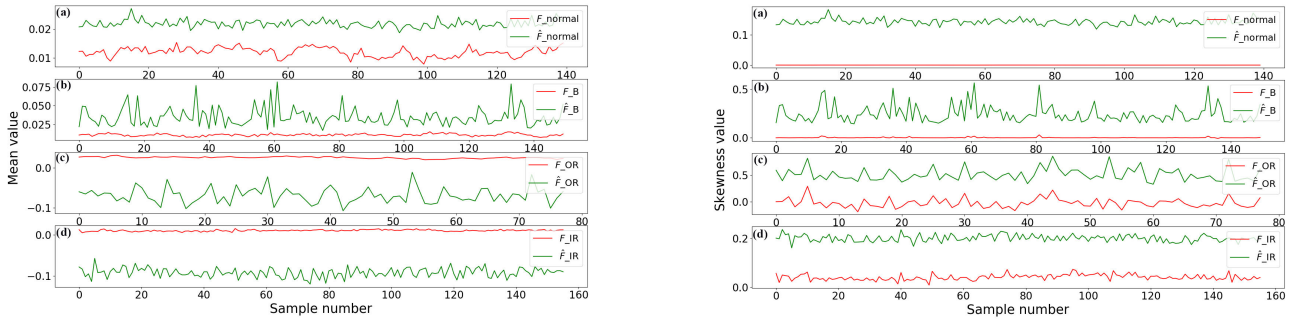


FIGURE 10. Overall performance of our model based on varying size of the subsample.

are very similar to each other. In contrast, Figures 8(b)-(d) show that the difference between $\hat{X}$ and X of each fault condition are obvious. We further investigate the difference between the latent vectors Z (original) and $\hat{Z}$ (re-engineered) learned by the model, and the results are shown in Figure 9. Not surprisingly, temporal vibrations of the fault conditions in Figure 9(b)-(d) are discriminated very well and temporal vibrations of the normal samples in Figures 9(a) are very similar.

In addition, in our approach, 16 most representative features are acquired after feature extractor. To intuitively observe the difference between the temporal vibration of feature F and the reconstructed feature $\hat{F}$ under various operating modes, we present the most representative two (mean value and skewness value) of the features in Fig. 11. It can be seen that the original extracted features of abnormal samples are sometimes difficult to be distinguished (magnitudes of red lines). After processing the abnormal samples according to the encoding-decoding-encoding patterns from the generator, the reconstructed feature values deviate a lot from the original compared to that of normal samples.



(a) (b)

FIGURE 11. Two examples of temporal vibrations of raw sample F and re-engineered sample $\hat{F}$ of (a) normal condition, (b) ball fault, (c) outer race fault, and (d) inner race fault on the CWRU dataset.

These results further demonstrate that the proposed method effectively learns the hidden patterns of the normal samples; consequently, fault samples can be recognized effectively.

2) THE IMPACT OF SAMPLE SEGMENTATION

In Figure 10, it can be observed that optimal performance is achieved when the length of the subsample is 12000 for each dataset. Considering the sampling frequency is 50kHz for voltage signals collected in our lab, we are able to infer that the potential patterns and spectral complexities are hidden in every 0.24 seconds. Similarly, the patterns of vibration signals from CWRU can be seen in every 0.25 seconds.

VI. CONCLUSION

In this paper, we put forward an innovative GAN-based architecture for anomaly detection involving imbalanced industrial time series; this architecture requires only normal samples for training. We elaborately design a feature extractor based on the characteristics of time series before the anomaly detector, and an encoder-decoder-encoder generator helps to detect the existence of abnormal samples by outputting larger anomaly scores. This network architecture achieves 100% accuracy (AUC=1) for both rolling bearing data from CWRU and our lab. With advances in both experimental techniques and equipment, and computational power and storage capacity, sensors in industrial fields can now collect more multivariate time-series data than ever before. Therefore, future work should include combining the information between different dimensions of multivariate time series to achieve better diagnostic effects.

APPENDIX

Detailed network architectures of AnoGAN and BiGAN are shown in Table 6 and Table 7.

ACKNOWLEDGMENT

Authors would like to thank Prof. Ye Yuan for his valuable comments and helpful suggestions.

TABLE 6. Network architecture of AnoGAN.

Layer	Output size (F/Ks/S)[1]
Generator	
Input	50
FC1 (ReLU/Batchnorm)	1024
FC2 (ReLU/Batchnorm)	128
ConvTran (ReLU)	4x4x1 (1/4/1)
Discriminator	
Input	4x4x1
Conv (LeakyReLU)	1x1x64 (64/4/1)
FC1 (LeakyReLU/Batchnorm)	1024
FC2 (Sigmoid)	1

[1] Filters/Kernel size/Stride

TABLE 7. Network architecture of BiGAN.

Layer	Output size (F/Ks/S)[1]
Encoder	
Input	4x4x1
Conv1 (LeakyReLU/Batchnorm)	4x4x32 (32/3/1)
Conv2 (LeakyReLU/Batchnorm)	4x4x64 (64/3/1)
Conv3 (LeakyReLU/Batchnorm)	1x1x128 (128/4/1)
FC	200
Decoder	
Input	200
FC1 (ReLU/Batchnorm)	1024
FC2 (ReLU/Batchnorm)	128
ConvTran2d (ReLU)	4x4x64 (64/4/1)
ConvTran2d (Tanh)	4x4x1 (1/3/1)
Discriminator (D)[2]	
Input	4x4x1
Conv1 (LeakyReLU)	4x4x64 (64/3/1)
Conv2 (LeakyReLU)	1x1x64 (64/4/1)
Discriminator (L)[3]	
Input	200
FC (LeakyReLU)	16
Discriminator[4]	
Input	64+16
FC1 (LeakyReLU)	16
FC2 (Sigmoid)	1

[1] Filters/Kernel size/Stride
[2] Inputs are original and generated samples
[3] Inputs are latent vectors
[4] Inputs are the combination of samples and their latent vectors

REFERENCES

[1] R. Liu, B. Yang, E. Zio, and X. Chen, "Artificial intelligence for fault diagnosis of rotating machinery: A review," *Mech. Syst. Signal Process.*, vol. 108, pp. 33–47, Aug. 2018.

[2] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. N. Kavuri, "A review of process fault detection and diagnosis: Part I: Quantitative model-based methods," *Comput. Chem. Eng.*, vol. 27, no. 3, pp. 293–311, 2003.

[3] Z. Wu, W. Lin, and Y. Ji, "An integrated ensemble learning model for imbalanced fault diagnostics and prognostics," *IEEE Access*, vol. 6, pp. 8394–8402, 2018.
[4] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, vol. 7. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[5] W. Mao, W. Feng, X. Liang, and X. Zhang, "A novel deep output kernel learning method for bearing fault structural diagnosis," *Mech. Syst. Signal Process.*, vol. 117, no. 2, pp. 293–318, 2019.
[6] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mech. Syst. Signal Process.*, vol. 20, no. 7, pp. 1483–1510, 2006.
[7] H. Yang, J. Mathew, and L. Ma, "Fault diagnosis of rolling element bearings using basis pursuit," *Mech. Syst. Signal Process.*, vol. 19, no. 2, pp. 341–356, Mar. 2005.
[8] P. Baraldi, L. Podofillini, L. Mkrtchyan, E. Zio, and V. N. Dang, "Comparing the treatment of uncertainty in Bayesian networks and fuzzy expert systems used for a human reliability analysis application," *Rel. Eng. Syst. Saf.*, vol. 138, pp. 176–193, Jun. 2015.
[9] V. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer, 2013.
[10] T. Han, D. Jiang, Q. Zhao, L. Wang, and K. Yin, "Comparison of random forest, artificial neural networks and support vector machine for intelligent diagnosis of rotating machinery," *Trans. Inst. Meas. Control*, vol. 40, no. 8, pp. 2681–2693, Jun. 2018.
[11] S. Haykin, *Neural Networks: A Comprehensive Foundation*. Upper Saddle River, NJ, USA: Prentice-Hall, 1994.
[12] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, "An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data," *IEEE Trans. Ind. Electron.*, vol. 63, no. 5, pp. 3137–3147, May 2016.
[13] H. Yin and K. Gai, "An empirical study on preprocessing high-dimensional class-imbalanced data for classification," in *Proc. IEEE 17th Int. Conf. High Perform. Comput. Commun. IEEE 7th Int. Symp. Cyberspace Saf. Secur. IEEE 12th Int. Conf. Embedded Softw. Syst.*, Aug. 2015, pp. 1314–1319.
[14] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.
[15] T. Schlegl, P. Seeböck, S. M. Waldstein, U. Schmidt-Erfurth, and G. Langs, "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery," in *Proc. Int. Conf. Inf. Process. Med. Imag.* New York, NY, USA: Springer, 2017, pp. 146–157.
[16] J. Donahue, P. Krähenbühl, and T. Darrell, "Adversarial feature learning," 2016, *arXiv:1605.09782*. [Online]. Available: https://arxiv.org/abs/1605.09782
[17] S. Akcay, A. Atapour-Abarghouei, and T. P. Breckon, "Ganomaly: Semi-supervised anomaly detection via adversarial training," 2018, *arXiv:1805.06725*. [Online]. Available: https://arxiv.org/abs/1805.06725
[18] J. Liu, F. Qu, X. Hong, and H. Zhang, "A small-sample wind turbine fault detection method with synthetic fault data using generative adversarial nets," *IEEE Trans. Ind. Informat.*, vol. 15, no. 7, pp. 3877–3888, Jul. 2019.
[19] T. Han, C. Liu, W. Yang, and D. Jiang, "A novel adversarial learning framework in deep convolutional neural network for intelligent diagnosis of mechanical faults," *Knowl.-Based Syst.*, vol. 165, pp. 474–487, Feb. 2019.
[20] W. Mao, Y. Liu, L. Ding, and Y. Li, "Imbalanced fault diagnosis of rolling bearing based on generative adversarial network: A comparative study," *IEEE Access*, vol. 7, pp. 9515–9530, 2019.
[21] D. Li, D. Chen, J. Goh, and S.-K. Ng, "Anomaly detection with generative adversarial networks for multivariate time series," 2018, *arXiv:1809.04758*. [Online]. Available: https://arxiv.org/abs/1809.04758
[22] C. Cheng, G. Ma, Y. Zhang, M. Sun, F. Teng, H. Ding, and Y. Yuan, "Online bearing remaining useful life prediction based on a novel degradation indicator and convolutional neural networks," 2018, *arXiv:1812.03315*. [Online]. Available: https://arxiv.org/abs/1812.03315
[23] A. Taylor, S. Leblanc, and N. Japkowicz, "Anomaly detection in automobile control network data with long short-term memory networks," in *Proc. IEEE Int. Conf. Data Sci. Adv. Anal. (DSAA)*, Oct. 2016, pp. 130–139.
[24] T. Guo, Z. Xu, X. Yao, H. Chen, K. Aberer, and K. Funaya, "Robust online time series prediction with recurrent neural networks," in *Proc. IEEE Int. Conf. Data Sci. Adv. Anal. (DSAA)*, Oct. 2016, pp. 816–825.
[25] S. Kanarachos, S.-R. G. Christopoulos, A. Chroneos, and M. E. Fitzpatrick, "Detecting anomalies in time series data via a deep learning algorithm combining wavelets, neural networks and Hilbert transform," *Expert Syst. Appl.*, vol. 85, pp. 292–304, Nov. 2017.
[26] K. Veeramachaneni, I. Arnaldo, V. Korrapati, C. Bassias, and K. Li, "Ai^2: Training a big data machine to defend," in *Proc. IEEE 2nd Int. Conf. Big Data Secur. Cloud (BigDataSecurity), IEEE Int. Conf. High Perform. Smart Comput. (HPSC), IEEE Int. Conf. Intell. Data Secur. (IDS)*, Apr. 2016, pp. 49–54.
[27] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, "A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 42, no. 4, pp. 463–484, Jul. 2012.
[28] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, no. 1, pp. 321–357, 2002.
[29] N. Japkowicz, "The class imbalance problem: Significance and strategies," in *Proc. Int. Conf. Artif. Intell.*, 2000, pp. 1–7.
[30] L. Nanni, C. Fantozzi, and N. Lazzarani, "Coupling different methods for overcoming the class imbalance problem," *Neurocomputing*, vol. 158, pp. 48–61, Jun. 2015.
[31] L. Demidova and I. Klyueva, "SVM classification: Optimization with the SMOTE algorithm for the class imbalance problem," in *Proc. 6th Medit. Conf. Embedded Comput. (MECO)*, Jun. 2017, pp. 1–4.
[32] W. Mao, L. He, Y. Yan, and J. Wang, "Online sequential prediction of bearings imbalanced fault diagnosis by extreme learning machine," *Mech. Syst. Signal Process.*, vol. 83, pp. 450–473, Jan. 2017.
[33] *The GAN Zoo: A List of All Named GANs*. Accessed: Jun. 14, 2019. [Online]. Available: https://github.com/hindupuravinash/the-gan-zoo
[34] S. K. Lim, Y. Loo, N.-T. Tran, N.-M. Cheung, G. Roig, and Y. Elovici, "DOPING: Generative data augmentation for unsupervised anomaly detection with GAN," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2018, pp. 1122–1127.
[35] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, *arXiv:1511.06434*. [Online]. Available: https://arxiv.org/abs/1511.06434
[36] A. Nanopoulos, R. Alcock, and Y. Manolopoulos, "Feature-based classification of time-series data," *Int. J. Comput. Res.*, vol. 10, no. 3, pp. 49–61, 2001.
[37] X. Wang, K. Smith, and R. Hyndman, "Characteristic-based clustering for time series data," *Data Mining Knowl. Discovery*, vol. 13, no. 3, pp. 335–364, 2006.
[38] H. Deng, G. Runger, E. Tuv, and M. Vladimir, "A time series forest for classification and feature extraction," *Inf. Sci.*, vol. 239, pp. 142–153, Aug. 2013.
[39] X. Wang, A. Wirth, and L. Wang, "Structure-based statistical features and multivariate time series clustering," in *Proc. 7th IEEE Int. Conf. Data Mining (ICDM)*, Oct. 2007, pp. 351–360.
[40] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training GANs," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 2234–2242.
[41] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: https://arxiv.org/abs/1412.6980
[42] *Bearing Data Center: Seeded Fault Test Data*. Accessed: Jun. 14, 2019. [Online]. Available: http://csegroups.case.edu/bearingdatacenter/home
[43] *Mad Net*. Accessed: Jun. 14, 2019. [Online]. Available: http://madnet.org:8765
[44] H. D. Shao, H. K. Jiang, X. Zhang, and M. G. Niu, "Rolling bearing fault diagnosis using an optimization deep belief network," *Meas. Sci. Technol.*, vol. 26, no. 11, p. 115002, Nov. 2015.
[45] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," in *Proc. Neural Inf. Process. Syst.*, Long Beach, CA, USA, 2017.
[46] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2011.
[47] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Jun. 2008, pp. 1322–1328.
[48] H. Han, W.-Y. Wang, and B.-H. Mao, "Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning," in *Proc. Int. Conf. Intell. Comput.* New York, NY, USA: Springer, 2005, pp. 878–887.

[49] G. E. Batista, R. C. Prati, and M. Monard, "A study of the behavior of several methods for balancing machine learning training data," *ACM SIGKDD Explorations Newslett.*, vol. 6, no. 1, pp. 20–29, 2004.
[50] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, Sep. 2008.

WENQIAN JIANG received the B.Eng. degree from the Department of Automation, North China Electric Power University, China, in 2018. She is currently pursuing the degree with the Huazhong University of Science and Technology, China.

Her current research interests include statistical learning, deep learning applications, and big data in manufacturing and prognostics.

YANG HONG received the B.Eng. degree in artificial intelligence and automation from the Huazhong University of Science and Technology, China, in 2019.

His current research interests include statistical learning and robot.

BEITONG ZHOU received the B.Eng. degree in electrical engineering and automation from Chongqing University, China, in 2017. He is currently pursuing the degree with the Huazhong University of Science and Technology, China.

His current research interests include statistical learning, deep learning applications, and optimizations.

XIN HE received the B.A. degree from the Academy of Arts and Design, School of Law, Tsinghua University, China, in 2011, and the M.A. degree in history of design from the Royal College of Art, U.K., in 2013. Since 2018, she has been an Assistant Lecturer with the Huazhong University of Science and Technology, China.

Her current research interests include artificial intelligence and design, user experience measurement, and history of modern design.

CHENG CHENG received the B.Eng. degree in measurement, control technology, and instrument from Tianjin University, China, in 2012, and the M.Sc. and Ph.D. degrees in control systems from the Imperial College London, U.K., in 2013 and 2018, respectively. Since 2018, she has been a Postdoctoral Researcher with the Huazhong University of Science and Technology, China.

Her current research interests include robust control, mechatronic systems modeling and simulation, and deep learning applications.