IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 17, 2020, accepted April 3, 2020, date of publication April 7, 2020, date of current version April 24, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2986356

Data Fusion Generative Adversarial Network for Multi-Class Imbalanced Fault Diagnosis of Rotating Machinery

QIANJUN LIU[1], (Member, IEEE), GUIJUN MA[2], AND CHENG CHENG[1]
[1]School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, China
[2]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 Key Research and Development Program of China under Grant 2018YFB1701202, and in part by the National Natural Science Foundation of China under Grant 51905197.

ABSTRACT For the fault diagnosis problems of rotating machinery in the real industrial practice, measurement data with imbalanced class distributions negatively affect the diagnostic performance of most conventional machine learning classification algorithms since equal cost weights are assigned to different fault classes. Meanwhile, the widely used traditional data generation methods for the imbalanced data problem are limited by data dependencies over time continuity. To fill this research gap, this paper develops a new diagnostic framework based on the adversarial neural networks (GAN) and multi-sensor data fusion technique to generate new synthetic data for data compensation purpose. Two different practice modes are designed based on this framework according to the position logic of the data fusion, namely a Pre-fusion GAN mode and a Post-fusion GAN mode. More concretely, without data pre-processing, the designed generator generates synthetic data to puzzle the discriminator and the synthetic data that out-trick the discriminator can be used to compensate the minor class. To avoid data dependency and to ensure the generality of the proposed framework, the network modelling are trained with a more practical approach where the training and test data are obtained under different rotating speeds. Two imbalanced data sets on the rotating machinery, one benchmark public rolling bearing data set and another gear box data set acquired in our lab, are used to validate the proposed method. The performance is examined through a wide range of data imbalanced ratios (as high as 30:1), and compared with other state-of-the-art methods. The experiment results conclude that the proposed Pre-fusion GAN and Post-fusion GAN frameworks both have good performance on the imbalanced fault diagnosis of rotating machinery.

INDEX TERMS Generative adversarial networks, imbalanced fault diagnosis, data continuity, rotating machinery.

I. INTRODUCTION

Fault diagnosis of the rotating machinery is of vital importance in manufacturing facilities to ensure a safe working condition and prevent the loss from escalating damage [1]. As more sensory data on current, temperature and vibration signals have become available over the past few years, many data-driven algorithms have been investigated to solve the fault diagnosis problem [2], which can be generally categorized into statistical model-based methods and machine learning methods. Statistical model-based methods are mainly refer to the Wiener process [3], particle filters [4], and Kalman filters [5], etc. In recent years, machine learning methods including convolutional neural network (CNN), recurrent neural network (RNN) [6] [7], and deep auto-encoders [8] have been widely applied in the fault diagnosis of rolling machinery. These above diagnosis methods have achieved promising performance on rotating machinery when their proposed models are trained to exam a balanced data set.

However, the imbalanced data have frequently occurred since most of the instances in industrial practice of rotating machinery systems are performed under normal or the healthy working condition (normal class) [9]. Thus,

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

the previous aforementioned models may suffer from biases towards the normal class as the data have a highly imbalanced ratio among normal class and faulty classes [10]. Several techniques have been proposed to address this issue. The two most widely used conventional synthetic data up-sampling/generation techniques, synthetic minority over-sampling technique (SMOTE) [11] and Adaptive synthetic sampling approach (ADASYN) [12], aim at linearly interpolating virtual data into data set based on the neighbor relationship (see [11]–[13] and the references therein). However, the construction of balanced data sets does not significantly improve the performance of these two techniques because the synthetic data generated by simple linear up-sampling may fail to provide sufficient feature information for the classifier. In contrast, machine learning methods prove to be more advanced in feature extraction. For example, relevance vector machine (RVM) [14] and auto-encoders (AEs) [15] are employed to propose models to solve the data imbalance problem by adjusting the weights of minor data without the data generation process. Hence, they offer a good alternative to data generation methods, but at the expense of tuning the weights of a model for a specific application. This paper attempts to deal with class imbalances by generating the minority fault classes and integrate the advantages of data generation and feature extraction.

The approach proposed in this paper is based on the recent developed Generative Adversarial Networks (GAN) which is composed of a generator and a discriminator. The discriminator is designed to distinguish the generated samples from real samples as exhaustively as possible. After multiple iterations, the generator generates data that can deceive the discriminator, that is, the synthesized data cannot be recognized by the discriminator. Starting with the original idea proposed in [16] by Goodfellow *et al.*, several extensions of this network have been proposed for image anomaly diagnosis tasks (e.g. DCGAN [17] and bidirectional generative adversarial networks (BiGAN) [18]). This yielded a set of techniques associated with the disciplines of GAN-based works for intelligent fault diagnosis: machinery and electronic systems, including GAN with Adaboost classifier [19], Wasserstein GAN (WGAN) with gradient penalty [20], Auxiliary Classifier GAN [21], GAN network for cross-domain fault diagnosis problems [22], [23], research on the loss of GAN [24] and WGAN with stacked auto-encoders [25]. Nevertheless, considering the fact that multiple fault types and extreme high imbalanced ratios (e.g. $>$20:1) of normal to fault sample number are very common in real industrial applications but have rarely been investigated in the previous literature, this study will extend the GAN-based fault diagnosis problems to more realistic situations by dealing with multiple fault types under high imbalanced ratios.

In addition, unlike the above-mentioned studies which mostly utilize sensory measurements recorded from a single sensor. Our approached framework fuses all appropriate sensor data of the tested machinery to provide more information and features to the GAN or the diagnosis model to further enhance the diagnostic accuracy. Based on this, two different modes are designed according to the position logic of the data fusion algorithm in our work, namely a Pre-fusion GAN mode where sensor data are fused before the GAN and a Post-fusion GAN mode where sensor data are fused after the GAN.

This paper presents a new GAN-based diagnostic framework that integrates multi-sensor data fusion technique and data generation technique for the data compensation purpose. Measured data of different sensor types are fused before or after the GAN network (i.e., Pre-fusion GAN or Post-fusion GAN) to improve the diagnosis performance. A convolutional neural network (CNN) is then designed to achieve multi-fault classification. In the model training and testing stages, we thereby propose a more practical approach that the training and testing data are acquired under different working conditions, which prevents the time dependencies over time continuity and thus increases the generality of the trained model. A benchmark CWRU rolling bearing data set and another gear box data set recorded by our lab are used to verify the feasibility and effectiveness of the proposed approach. To further validate the effectiveness of the proposed method, we also compare the current mainstream data generation methods, such as ADSYN, SMOTENC, KMeansSMOTE, and SVMSMOTE.

The main contributions of this paper are as follows: 1) The diagnosis framework in this study combines the benefits of data generation and feature extraction by using the GAN-based model, the data fusion logic, and the CNN model. 2) As high as 30:1 normal to fault class-imbalanced ratio is investigated for multi-class fault diagnosis problem, which is closer to the real industrial scenarios. 3) The training and validation datasets are undergoing different operating conditions, thereby preventing the time dependency problem in time-series forecasting and prediction problems, which increases the robustness of the model and demonstrates the generalizability of the diagnostic model. 4) The proposed approach will be first verified on a benchmark public rolling bearing data set. Additionally, a new gear box data set acquired by our lab will be used to prove the model effectiveness on general rotating machinery.

This paper consists of five sections. Section II reviews the basic structure of the GAN and CNN. In Section III, the problem of this study is formulated and the general diagnostic framework is explained. Section IV presents the experiment results based on two rotating machines, with their performance compared with other mainstream class-imbalanced methods. Section V summarises the paper and discusses future works.

The following notations will be used throughout this work: the symbol $\mathbb{R}$ stands for the real number set and $\mathbb{Z}$ for the positive integer set, and $\bar{(\cdot)}$ represents the fused data.

II. PRELIMINARIES

This section will show the basic principle of GAN, CNN, and their structure components including convolutional layer,

activation function, fully connected layer, pooling layer, and dropout.

A. NETWORK LAYERS IN GAN AND CNN

Convolutional layer (CONV) is the key module of CNN and GAN, The CONV layer is made up with a set of learnable filters that set the width and height of its structural volume. Each filter performs a convolutional operation on the width and height based on the structure volume to obtain the dot product between the filter entry and any position input in forward transmission. As the filter slides across the width and height of the input, a activation map is generated that shows the filter's response at each spatial location: There is a complete set of filters in each CONV layer, and each filter will generate a separate activation map. The activation maps will be stacked along the depth dimension to form the output.

$$(d_n)_m = (Q_n * x^i_{(n-1)})_m + b_n \tag{1}$$

where $n = 1, \ldots, n$ represents the n-th feature map and m represents the index of filter of the n-th feature map. $(d_n)_m$ is the output of the layer n at the m-th position. $x^i_{(n-1)}$ is input from the previous layer $n-1$ at the i-th position. '$*$' means the convolutional operation. Q^n denote the trainable parameters of filter and b^n is the bias of the n-th feature map.

Activation function is introduced here is the Rectified linear unit (ReLU), the sigmoid and the softmax. Due to its piecewise constant gradient characteristic, ReLU is selected as the non-linear activation function:

$$ReLU(d) = max(0, d) \tag{2}$$

When dealing with dual classification problems, the sigmoid is always chosen as the activation function for output:

$$sigmoid(d) = \frac{1}{1 + e^{-x}} \tag{3}$$

In this paper, the sigmoid activation function is used in the the discriminator of GAN as the activation function of the discriminator. The softmax is able to be used as an activation function for multi-classification problems for output:

$$softmax(d) = \frac{\exp d}{\sum \exp d} \tag{4}$$

Pooling layer will be regularly inserted between consecutive CONV layers to implement the function of gradually reducing the size of the representation space, so that parameters and calculations in the network are reduced to control over-fitting. The pooling layer runs independently for each depth slice of the input, and often the max operation resizes spatially:

$$A^j_n = \max_{\gamma=1,\ldots,\omega^j} u^{j-1}_{\gamma+(n-1)I^j_{al}} \tag{5}$$

where $\omega^j \in \mathbb{Z}$ is the pooling size, and I^j_{al} is the stride in the pooling layer. A^j_n is the output of the pooling layer, and the $u^{j-1}_{\gamma+(n-1)I^j_{al}}$ is the input of the pooling layer.

Dropout [26] is a Google patented regularization technology that reduces over-fitting of neural networks by preventing complex collaborative adjustments to training data.

Fully-connected layer is fully connected to all activation in the previous layer, that is, the same architecture as the conventional multi-layer neural network, and its activation calculation method is achieved by matrix multiplication combined with offset.

B. GAN

The synthesis of virtual artificial data is based on GAN. GAN is an unsupervised learning method that learns by having two neural networks interact with each other. This method was proposed by Goodfellow et al in [16]. After that, the coupled generative adversarial network [27], wasserstein generative adversarial networks [17], energy-based generative adversarial networks [28], deep convolutional generative adversarial networks [29], and the least squares generative adversarial network [30] are followed. The GAN consists of a generator denoted by G and a discriminator denoted by D. As shown in Fig. 2, the role of G as a generator is to generate synthetic data that may be similar to the real sample, and the function of D as a discriminator is to distinguish real samples from false samples as much as possible. Initially, G will randomly sample from the latent space, and then adjust the parameters continuously until data produced by G cannot distinguish by D, that is, the synthetic data can deceive the discriminator.

C. CNN

CNN will be the multi-category imbalanced fault classifier in this study. In deep learning, the CNN is a regularized version of the multi-layer perception which always refers to a full connected layer. CNN can also be defined as a class of deep neural networks [31], [32]. The regularization is able to achieved by adding some form of weight measurement to the loss function. For normalization CNN are using hierarchical patterns in data meanwhile combining simple patterns with complex patterns for better results. The outstanding advantage of CNN is that it can mine the inherent relationships of the same type of data and make the boundaries between different types of data clear. This not only overcomes the problem of gradient dilution and the problem of being easily trapped into local minimums in traditional shallow fault diagnosis algorithms, but also shortens training time.

III. DATA FUSION GAN FOR MULTI-CLASS IMBALANCED FAULT DIAGNOSIS

The proposed framework for the multi-class imbalanced fault diagnosis problem using data fusion and GAN can be generally divided by three stages: the data fusion process, GAN for synthetic data generation, and CNN for multi-class fault diagnosis. The schematic of the proposed diagnostic framework is presented in Fig. 1. In the following subsections, these three stages will be explained in details in Section III-A, Section III-B, and Section III-C, respectively.

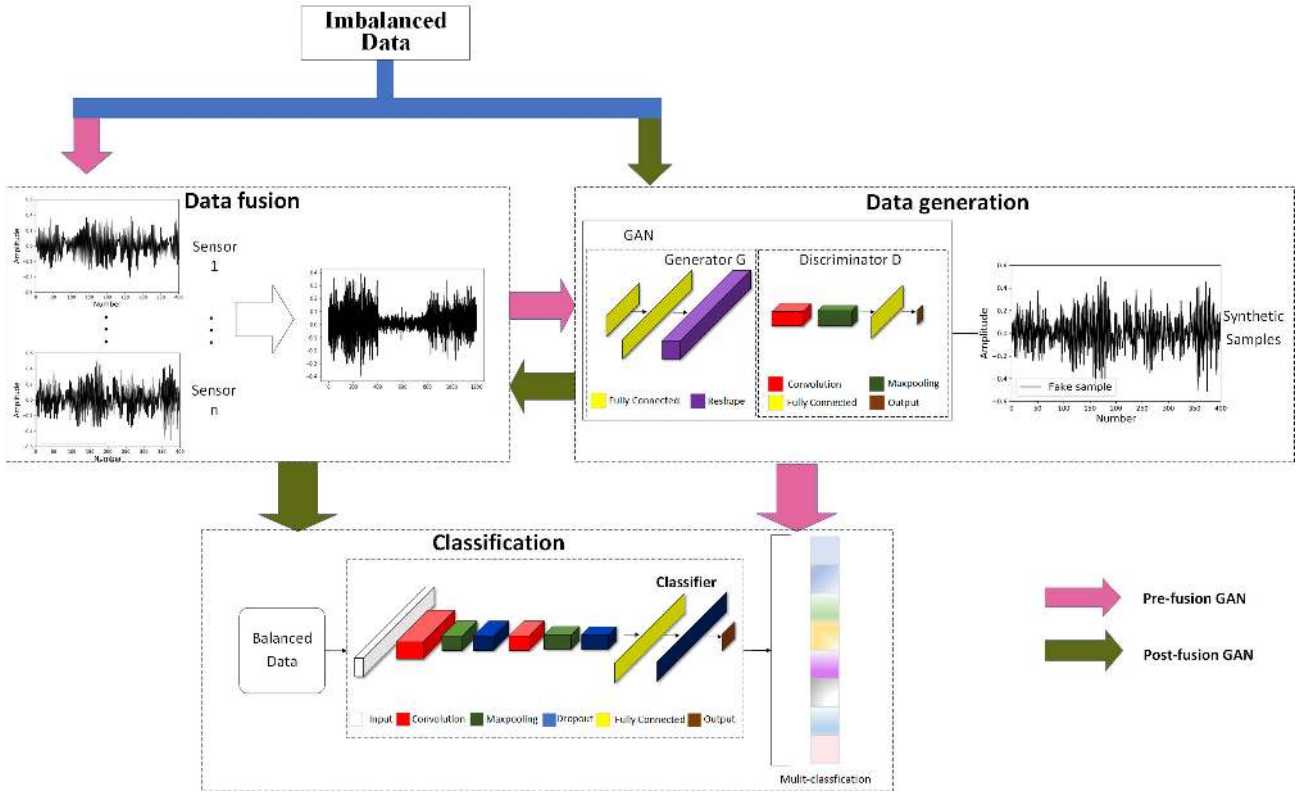


FIGURE 1. Imbalanced fault diagnosis of rotating machinery estimation framework.

A. DATA FUSION

In this study, data fusion is proceed to merge the data information/features from different sensors before data flows into the GAN generator or the CNN classifier. As shown in Fig. 1, data fusion can be expected before Stage 2 or after Stage 2, in this framework, we denote the former case as Pre-fusion GAN and the later case as Post-fusion GAN. More specifically, the Pre-fusioin GAN deals with the raw sensor signals, on the other hand, the Post-fusion GAN deals with signals after data data generation by the GAN generator. We will compare the classification enhancement of both cases in the later experiments on rotating machinery. Since expressions of formulas are the same, here we address the data fusion as a unified problem regardless of which fusion case will be used.

Consider a time series data $\mathbf{S}_{i,j} = \{S^p_{i,j}\}^P_{p=1} \in \mathbb{R}^{l\times P}$ for senor $i \in \mathbb{Z}$ and class $j \in \mathbb{Z}$, where P is the sample number of class j and l is the measurement number of each sample. Thus, one of the sample $S^p_{i,j}$ is with the form:

$$S^p_{i,j} = [s^p_{i,j}(1), s^p_{i,j}(2), \ldots, s^p_{i,j}(l)]^T. \tag{6}$$

For data fusion of class j, multiple time-series data streams corresponding to senors $i = [1, 2, \ldots, I]$ are concatenated in series to form a fused data set $\bar{\mathbf{S}}_j$ for data generation or classification, i.e.

$$\bar{\mathbf{S}}_j = \begin{bmatrix} \mathbf{S}_{1,j} \\ \mathbf{S}_{2,j} \\ \vdots \\ \mathbf{S}_{I,j} \end{bmatrix} = \begin{bmatrix} s^1_{1,j}(1) & s^2_{1,j}(1) & \ldots & s^P_{1,j}(1) \\ s^1_{1,j}(2) & s^2_{1,j}(2) & \ldots & s^P_{1,j}(2) \\ \vdots & \vdots & \ddots & \vdots \\ s^1_{1,j}(l) & s^2_{1,j}(l) & \ldots & s^P_{1,j}(l) \\ \vdots & \vdots & \vdots & \vdots \\ s^1_{I,j}(1) & s^2_{I,j}(1) & \ldots & s^P_{I,j}(1) \\ s^1_{I,j}(2) & s^2_{I,j}(2) & \ldots & s^P_{I,j}(2) \\ \vdots & \vdots & \ddots & \vdots \\ s^1_{I,j}(l) & s^2_{I,j}(l) & \ldots & s^P_{I,j}(l) \end{bmatrix} \in \mathbb{R}^{lI\times P}. \tag{7}$$

Instead of dealing with a binary class-imbalanced diagnosis problem, multi-class fault diagnosis problem in this study involves several representative fault types on the rotating machinery, regarding to different fault locations, fault roughness levels, and rotating speeds. Hence, the classification model should process multiple classes corresponding to $j = [1, 2, \ldots, J]$, where J is the number of classes including normal class and all fault types. Considering the actual industrial problems, the ratios of sample number P in the normal case $j = 1$ to P in each fault case $j = \{2, \ldots, J\}$ are very high, resulting in very high normal to fault ratios.

B. GAN-BASED DATA GENERATION

As shown in Fig. 2, the network structure of the GAN in the proposed method is consists of a generator G and a discriminator D.

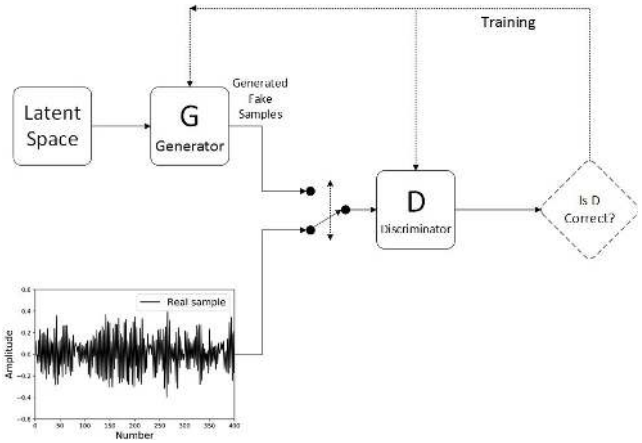


FIGURE 2. The schematic diagram of GAN.

1) GENERATOR

The G is used to generate synthetic samples as similar as model input generated by the Stage 1 in Fig. 1.

For the Pre-fusion case, the training data set $\bar{F}^{j}_{train} = \bar{\mathbf{S}}_j$ of fault class j for $j = \{2, 3, \ldots, J\}$. In total $J-1$ GAN models will be trained for data generation.

For the Post-fusion case, the training data set $F^{i,j}_{train} = \mathbf{S}_{i,j}$ of fault class j for $j = \{2, 3, \ldots, J\}$ and sensor number i for $i = \{1, 2, \ldots, I\}$. In total $(J-1)I$ GAN models will be trained for data generation.

The generator G consists of fully-connected layers, batch-norm layers, and leaky *ReLU* activation layers. The specific layers in the experiment are shown in Table 4 and Table 5. Randomly sampled M from the latent space are initialized to generate the synthetic data with the same dimension and size as $\bar{F}^{j}_{train}$ or $F^{i,j}_{train}$. M, where for the Pre-fusion case

$$\bar{F}^{j}_{synthetic} = G(M), \tag{8}$$

or for the Post-fusion case

$$F^{i,j}_{synthetic} = G(M). \tag{9}$$

2) DISCRIMINATOR

The generator makes that both the feature of the original training set and the pattern of the latent space vector are learned during the multi-epochs.

The discriminator D is used to distinguish whether input data is real or generated. The loss function used here is binary cross-entropy as loss function. *Adam* [33] is used here as the optimizer.

C. CNN AS A CLASSIFIER

After fault data sets has been generated, new balanced data sets can be construed. For the Pre-fusion case, the new balanced data set of a fault class j is $\bar{F}^{j}_{pre} = [\bar{F}^{j}_{synthetic}, \bar{F}^{j}_{train}]$. Together with the normal sample, CNN classification model used to be trained for multi-class fault diagnosis (see classification step in Fig. 1). For the Post-fusion case, the generated $F^{i,j}_{synthetic}$ and $F^{i,j}_{train}$ from $i = 1$ to I will be concatenated according to construct Eq. (7) to the balanced fused data $\bar{F}^{j}_{post}$ for CNN training. The loss function used in CNN is categorical crossentropy. *Adam* [33] is used here as the optimizer.

IV. EXPERIMENTAL SETUP

This section shows experimental setup to verify the effectiveness and feasibility of the proposed method for the diagnosis of multi-class imbalanced rotating machinery on two representative data sets: a rolling bearing data from Case Western Reserve University (CWRU) [34] and a gear box data collected from our laboratory [35]. Both data sets are balanced datasets since originally generated. Therefore, we are able to manually adjust the data set composition for the purpose of validating the effectiveness of the proposed method across small to high imbalanced ratios (i.e., ratios from 1:10 to 1:30 in this research).

A. DATA SET DESCRIPTION

1) ROLLING BEARING DATA

As shown in Fig. 3, this experimental apparatus includes a motor (motor shaft is supported by test bearings), a torque transducer/encoder, and a dynamo-meter. Four kinds of motor loads 0, 1, 2, and 3 hp correspond to four motor rotating speeds 1797 rpm, 1772rpm, 1750 rpm, and 1730 rpm, respectively. Accelerometers with a sampling rate of 12 kHz are used for collecting vibration signals. In this task, four fault diameters (0.007, 0.014, 0.021, and 0.028 inches (1 inch = 25.4 mm)) of three fault types (outer race (OR) fault, inner race (IR) fault, and ball (B) fault) together with one normal condition, a total of 10 classes, are considered in this work (see Table 1). Three accelerometers associated with different positions (drive end (DE), fan end (FE), and base (BA)) are conducted for this 10-way classification problem.

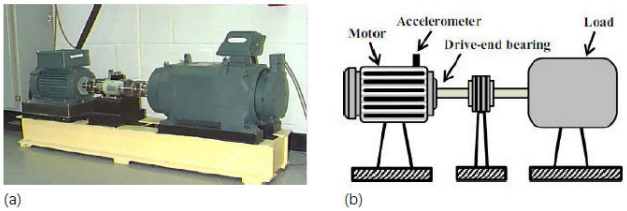


FIGURE 3. The CWRU data acquisition environment and equipment: (a) physical picture [36] and (b) rolling bearing experiment setup. [15].

2) GEAR BOX DATA

The schematic diagram of the gear box experiment is shown in the Fig. 4, including four key gears: one input high-speed

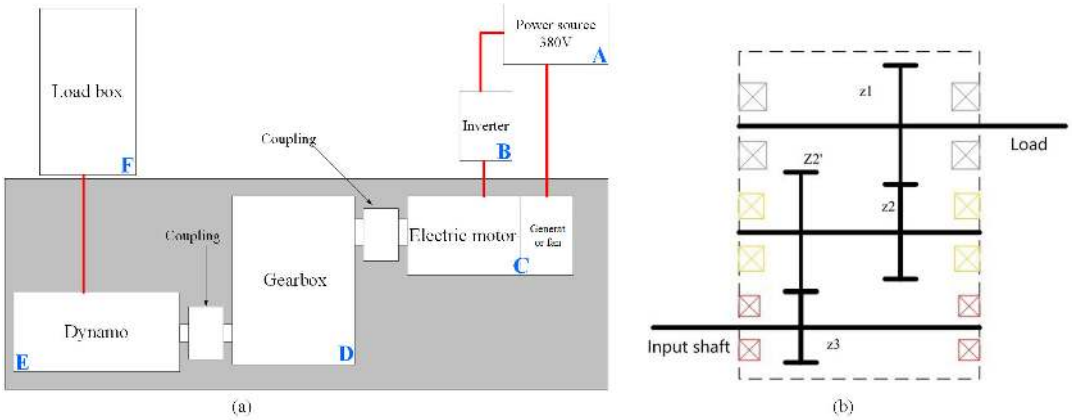


FIGURE 4. (a) The Structure of the experimental platform [35] of the data set from our lab (b) The internal schematic of the gearbox used in the experiment.

TABLE 1. Description of ten rolling bearing working conditions.

Working condition	Fault diameter (In.)	OR fault orientation	Labels
Normal	0	-	1
Ball fault	0.007	-	2
Ball fault	0.014	-	3
Ball fault	0.021	-	4
Inner race fault	0.007	-	5
Inner race fault	0.014	-	6
Inner race fault	0.021	-	7
Outer race fault	0.007	Center @6:00	8
Outer race fault	0.014	Center @6:00	9
Outer race fault	0.021	Center @6:00	10

gear, one output low-speed gear, and two intermediate gears. Diverse vibration signals are collected by two accelerometers on the vertical and horizontal directions, and a current sensor is used to measure current sensor. Three types of faults and normal data constitute a four-way classification problem. More specifically, three types of fault conditions are 1) one tooth broke in high-speed shaft side, 2) one tooth broke in low-speed shaft side, and 3) the high-speed shaft gear breaks one tooth, while the middle shaft gear breaks one tooth. Their labels are presented in Table 3. The sampling rate of the acceleromters and sensor are 10 kHz and 1 kHz respectively.

B. DATA PROCESSING AND IMPLEMENTATION DETAILS

1) ROLLING BEARING DATA

As shown in Table 2, ten working conditions including nine faulty conditions and one normal condition are considered, where each condition contains $P = 300$ samples and each sample is a selected vibration signal segment with a length of $l = 400$.

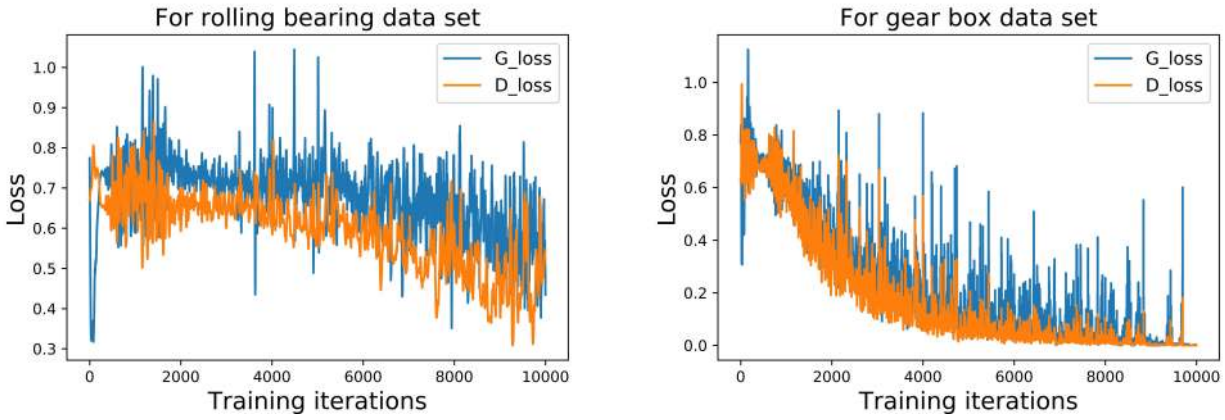
For the Pre-fusion GAN, we first consider construct the data set under different imbalanced ratios (i.e., 1:10, 1:20, and 1:30), then merged the data collected by the FE, DE, and BA accelerometers according to the time, so as to obtain fused data with a length of 1200. After that, Pre-fusion GAN

TABLE 2. The number of training/testing samples from three different sensors.

Bearing working condition	Drive end accelerometer data (DE)	Fan end accelerometer data (FE)	Base accelerometer data (BA)	Labels
Normal	300(0 hp)/300(2 hp)	-	300(0 hp)/300(2 hp)	1
Ball fault	300(0 hp)/300(2 hp)	300(0 hp)/300(2 hp)	300(0 hp)/300(2 hp)	2
Ball fault	300(0 hp)/300(2 hp)	300(0 hp)/300(2 hp)	300(0 hp)/300(2 hp)	3
Ball fault	300(0 hp)/300(2 hp)	300(0 hp)/300(2 hp)	300(0 hp)/300(2 hp)	4
Inner race fault	300(0 hp)/300(2 hp)	300(0 hp)/300(2 hp)	300(0 hp)/300(2 hp)	5
Inner race fault	300(0 hp)/300(2 hp)	300(0 hp)/300(2 hp)	300(0 hp)/300(2 hp)	6
Inner race fault	300(0 hp)/300(2 hp)	300(0 hp)/300(2 hp)	300(0 hp)/300(2 hp)	7
Outer race fault	300(0 hp)/300(2 hp)	300(0 hp)/300(2 hp)	300(0 hp)/300(2 hp)	8
Outer race fault	300(0 hp)/300(2 hp)	300(0 hp)/300(2 hp)	300(0 hp)/300(2 hp)	9
Outer race fault	300(0 hp)/300(2 hp)	300(0 hp)/300(2 hp)	300(0 hp)/300(2 hp)	10

TABLE 3. The number of samples from two different sensors.

Gear working condition	The numbers of vibration data	The numbers of current data	Labels
Tooth broke in high-speed shaft	450(200 r/min)/450(400 r/min)/450(600 r/min)	450(200 r/min)/450(400 r/min)/450(600 r/min)	0
Tooth broke in low-speed shaft	450(200 r/min)/450(400 r/min)/450(600 r/min)	450(200 r/min)/450(400 r/min)/450(600 r/min)	1
Tooth broke in high-speed shaft and middle shaft	450(200 r/min)/450(400 r/min)/450(600 r/min)	450(200 r/min)/450(400 r/min)/450(600 r/min)	2
Normal	450(200 r/min)/450(400 r/min)/450(600 r/min)	450(200 r/min)/450(400 r/min)/450(600 r/min)	3



(a) (b)

FIGURE 5. The loss of Generator/The loss of Discriminator of GAN model.

TABLE 4. Network architecture for rolling bearing data set.

Layer	Filters	Ks/S[1]	Ps/Dr[2]	Output size
Generator				
Input	-	-	-/-	m/n[3]
FC(Linear)	-	-	-/-	1x100
FC(Linear)	-	-	-/-	m/n[3]
Reshape	-	-	- /-	p/q[4]
Discriminator				
Input	-	-	-/-	p/q[4]
Conv (ReLU)	64	50/50	-/-	1x8x64
Maxpooling	-	-	2/-	1x4x64
FC(Linear)	-	-	-	1x256
FC (Sigmoid)	-	-	-/-	1x1
CNN for imbalanced data set				
Input	-	-	-/-	p/q[4]
Conv (ReLU)	64	20/10	-	1x39x64
Maxpooling	-	-	2/-	1x19x64
Dropout	-	-	-/0.7	1x18x64
Conv (ReLU)	32	2/1	-	1x18x64
Maxpooling	-	-	2/-	1x9x64
Dropout	-	-	-/0.7	1x9x64
FC(Linear)	-	-	-	1x576
FC (Softmax)	-	-	-/-	1x10
CNN for balanced data set				
Input	-	-	-/-	p/q[4]
Conv (ReLU)	64	50/40	-	1x39x64
Maxpooling	-	-	4/-	1x19x64
Dropout	-	-	-/0.5	1x18x64
Conv (ReLU)	32	2/1	-	1x18x64
Maxpooling	-	-	4/-	1x9x64
Dropout	-	-	-/0.5	1x9x64
FC(Linear)	-	-	-	1x576
FC (Softmax)	-	-	-/-	1x10

[1] Kernel size/Stride
[2] Pooling size/Dropout rate
[3] m = 1x400, where m is the output size of Post-fusion GAN, n = 1x1200 ,where n is the the output size of Pre-fusion GAN
[4] p = 1x400x1, where m is the output size of Post-fusion GAN, q = 1x1200x1 ,where n is the output size of Pre-fusion GAN

TABLE 5. Network architecture for gear box data set from our lab.

Layer	Filters	Ks/S[1]	Ps/Dr[2]	Output size
Generator				
Input	-	-	-/-	m/n[3]
FC(Linear)	-	-	-/-	1x100
FC(Linear)	-	-	-/-	m/n[3]
Reshape	-	-	- /-	p/q[4]
Discriminator				
Input	-	-	-/-	p/q[4]
Conv (ReLU)	64	50/50	-/-	1x8x64
Maxpooling	-	-	2/-	1x4x64
FC(Linear)	-	-	-	1x256
FC (Sigmoid)	-	-	-/-	1x1
CNN for imbalanced data set				
Input	-	-	-/-	p/q[4]
Conv (ReLU)	64	30/20	-	1x59x64
Conv (ReLU)	32	2/1	-	1x59x64
Maxpooling	-	-	2/-	1x29x64
FC(Linear)	-	-	-	1x1856
FC (Softmax)	-	-	-/-	1x4
CNN for balanced data set				
Input	-	-	-/-	p/q[4]
Conv (ReLU)	64	50/50	-	1x12x64
Conv (ReLU)	32	2/1	-	1x12x64
Maxpooling	-	-	2/-	1x6x64
FC(Linear)	-	-	-	1x384
FC (Softmax)	-	-	-/-	1x4

[1] Kernel size/Stride
[2] Pooling size/Dropout rate
[3] m =1x600, where m is the output size of Post-fusion GAN, n = 1x1200 ,where n is the the output size of Pre-fusion GAN
[4] p = 1x600x1, where m is the output size of Post-fusion GAN, q = 1x1200x1 ,where n is the the output size of Pre-fusion GAN

is used to generate new data for classed that have relatively few samples. For instance, when the imbalanced ratio is 1:10, we first fused the sensory data for each sample, then a data set was obtained by reserving 300 normal samples and 30 faulty samples in each faulty type. Next, the GAN network is used to generate synthetic data to compensate minor classes. For Post-fusion GAN, according to the imbalanced ratio, we generated the synthetic data for each of the three accelerometers of nine faulty types that have a small amount of samples, then the data of three accelerometers are fused, and 300 samples with a length of 1200 in each of ten classes are finally obtained for CNN classifier.

It should be noted that data of the fan end accelerometer is vacant in the normal condition (see Table 2), in order to satisfy the requirement of network training, 300 zero padding of the same length are added in this part to make the normal data is consistent with the fault data in length.

In order to avoid the time-series dependency of segmented data [37], the sensor data under 1797 rpm (0 hp) are used for training and the data under 1799 rpm (2 hp) are used for test.

TABLE 6. Comparison of results between proposed approaches and existing major approaches on rolling bearing data set.

IR[1]/Accuarcy(%)	Unprocessed	ADSYN	KMeansSMOTE	SMOTENC	SVMSMOTE	Fusion only	Pre-fusion GAN	Post-fusion GAN
Balanced	86.01±0.04	-	-	-	-	**91.96±0.10**	-	-
1:10	68.43±0.12	70.26±0.13	70.28±0.11	69.96±0.11	70.98±0.12	**87.10±0.17**	83.91±0.18	85.72±0.39
1:20	56.01±0.06	56.22±0.07	56.31±0.11	56.84±0.07	57.15±0.09	75.07±0.14	75.90±0.15	**77.23±0.12**
1:30	51.17±0.30	52.33±0.15	53.50±0.25	53.62±0.14	54.27±0.23	65.40±0.21	72.37±0.12	**75.47±0.23**
IR[1]/Precision(%)	Unprocessed	ADSYN	KMeansSMOTE	SMOTENC	SVMSMOTE	Fusion only	Pre-fusion GAN	Post-fusion GAN
Balanced	85.81±0.15	-	-	-	-	**91.25±0.16**	-	-
1:10	65.96±0.06	68.71±0.12	70.31±0.17	68.34±0.26	70.52±0.29	**86.54±0.12**	83.87±0.15	85.83±0.21
1:20	56.63±0.23	56.72±0.09	57.01±0.21	56.98±0.27	57.11±0.24	73.25±0.11	75.83±0.21	**77.25±0.17**
1:30	52.51±0.19	52.15±0.11	53.07±0.07	53.92±0.19	53.99±0.21	64.37±0.16	71.86±0.23	**75.33±0.19**
IR[1]/Recall(%)	Unprocessed	ADSYN	KMeansSMOTE	SMOTENC	SVMSMOTE	Fusion only	Pre-fusion GAN	Post-fusion GAN
Balanced	84.43±0.08	-	-	-	-	**90.86±0.07**	-	-
1:10	66.25±0.14	67.25±0.38	65.94±0.26	68.01±0.29	69.24±0.17	**86.88±0.23**	84.01±0.22	86.27±0.23
1:20	57.74±0.25	55.32±0.28	56.17±0.15	55.3±0.08	56.49±0.06	72.64±0.14	74.39±0.16	**77.68±0.08**
1:30	52.57±0.27	51.23±0.24	51.26±0.18	55.25±0.18	53.97±0.14	63.28±0.28	71.02±0.07	**75.26±0.31**
IR[1]/F1-value(%)	Unprocessed	ADSYN	KMeansSMOTE	SMOTENC	SVMSMOTE	Fusion only	Pre-fusion GAN	Post-fusion GAN
Balanced	84.43±0.08	-	-	-	-	**90.86±0.07**	-	-
1:10	66.1±0.15	67.97±0.31	68.05± 0.12	68.17±0.22	69.87±0.24	**86.71±0.17**	83.94±0.11	86.05±0.25
1:20	57.19±0.23	56.01±0.08	56.59±0.14	56.13±0.15	56.8±0.16	72.94±0.21	75.1±0.15	**77.46±0.21**
1:30	52.54±0.21	51.68±0.18	52.15±0.11	54.57±0.16	53.98±0.14	63.82±0.13	71.44±0.16	**75.29±0.17**
IR[1]/AUC	Unprocessed	ADSYN	KMeansSMOTE	SMOTENC	SVMSMOTE	Fusion only	Pre-fusion GAN	Post-fusion GAN
Balanced	91.35±0.08	-	-	-	-	**95.22±0.25**	-	-
1:10	81.63± 0.17	82.3 ±0.15	83.77± 0.21	83.56± 0.36	84.31± 0.06	**92.17± 0.26**	90.87± 0.08	91.85± 0.16
1:20	73.36± 0.06	75.02± 0.26	75.68± 0.11	74.98± 0.33	77.28± 0.15	85.63± 0.16	86.56± 0.15	**88.32± 0.07**
1:30	70.21± 0.35	71.66± 0.18	73.91± 0.23	72.15± 0.29	75.02± 0.13	82.59± 0.32	84.78± 0.21	**86.2 ±0.23**

[1] Imbalanced ratio

TABLE 7. Comparison of results between proposed approaches and existing major approaches on gear box data set.

IR[1]/Accuarcy(%)	Unprocessed	ADSYN	KMeansSMOTE	SMOTENC	SVMSMOTE	Fusion only	Pre-fusion GAN	Post-fusion GAN
Balanced	95.96±0.15	-	-	-	-	**98.80±0.21**	-	-
1:10	88.97±0.13	89.01±0.56	89.28±0.11	89.99±0.41	90.98±0.22	91.15±0.22	91.26±0.32	**92.07±0.11**
1:20	82.00±0.40	82.98±0.26	82.15±0.13	82.56±0.27	83.08±0.19	84.30±0.31	84.42±0.27	**86.21±0.11**
1:30	79.02±0.21	80.15±0.38	81.00±0.21	80.96±0.17	81.39±0.55	81.29±0.17	82.77±0.23	**84.11±0.15**
IR[1]/Precision(%)	Unprocessed	ADSYN	KMeansSMOTE	SMOTENC	SVMSMOTE	Fusion only	Pre-fusion GAN	Post-fusion GAN
Balanced	95.21±0.14	-	-	-	-	**97.95±0.35**	-	-
1:10	87.52± 0.23	88.15± 0.28	89.12± 0.11	89.51± 0.23	90.53± 0.31	90.87± 0.33	91.14± 0.12	**92.26± 0.33**
1:20	81.6 ±0.07	81.45± 0.06	81.23± 0.16	81.65± 0.19	82.37± 0.36	84.17± 0.31	84.58± 0.14	**86.48± 0.17**
1:30	78.54± 0.28	79.63± 0.17	80.58± 0.08	80.77± 0.08	81.67± 0.35	81.2 ±0.21	82.07± 0.11	**84.32± 0.31**
IR[1]/Recall(%)	Unprocessed	ADSYN	KMeansSMOTE	SMOTENC	SVMSMOTE	Fusion only	Pre-fusion GAN	Post-fusion GAN
Balanced	94.83±0.23	-	-	-	-	**98.83±0.17**	-	-
1:10	88.45± 0.21	89.34± 0.15	88.62± 0.12	89.32± 0.17	90.66± 0.33	90.62± 0.19	91.25± 0.08	**91.84± 0.26**
1:20	81.86± 0.08	82.07± 0.26	81.33± 0.29	81.46± 0.21	82.97± 0.17	83.56± 0.25	83.98± 0.15	**85.53± 0.15**
1:30	78.13± 0.22	78.35± 0.33	79.62± 0.31	78.96± 0.07	81.45± 0.35	80.79± 0.33	82.69± 0.19	**83.27± 0.18**
IR[1]/F-value(%)	Unprocessed	ADSYN	KMeansSMOTE	SMOTENC	SVMSMOTE	Fusion only	Pre-fusion GAN	Post-fusion GAN
Balanced	95.02±0.15	-	-	-	-	**98.39±0.14**	-	-
1:10	87.98± 0.16	88.74± 0.15	88.87± 0.07	89.41± 0.32	90.59± 0.31	90.74± 0.32	91.19± 0.23	**92.05± 0.25**
1:20	81.73± 0.23	81.76± 0.14	81.28± 0.28	81.55± 0.09	82.67± 0.08	83.86± 0.09	84.28± 0.19	**86.01± 0.27**
1:30	78.33± 0.31	78.98± 0.21	80.1 ±0.23	79.85± 0.13	81.56± 0.18	80.99± 0.18	82.38± 0.08	**83.79± 0.16**
IR[1]/AUC	Unprocessed	ADSYN	KMeansSMOTE	SMOTENC	SVMSMOTE	Fusion only	Pre-fusion GAN	Post-fusion GAN
Balanced	97.88±0.29	-	-	-	-	**99.28±0.24**	-	-
1:10	94.15± 0.16	94.23± 0.11	94.32± 0.17	94.28± 0.23	94.86± 0.15	95.32± 0.26	95.98± 0.26	**96.63± 0.12**
1:20	90.27± 0.12	90.36± 0.15	90.45± 0.29	91.03± 0.21	91.25± 0.06	91.93± 0.16	92.51± 0.29	**92.86± 0.11**
1:30	89.63± 0.11	89.72± 0.16	89.43± 0.16	89.54± 0.08	90.21± 0.25	90.63± 0.32	91.14± 0.21	**91.75± 0.16**

[1] Imbalanced ratio

2) GEAR BOX DATA

As shown in Table 3, four working conditions (i.e., tooth broken in high-speed shaft, tooth broken in low-speed shaft, tooth broken in high-speed shaft and middle shaft, and normal condition) are used for imbalanced classification. We choose one of two vibration signals and the current signal for investigation. There are $P = 450$ two-channel signal samples ($I = 2$) with a sample length of $l = 600$ are used to constitute

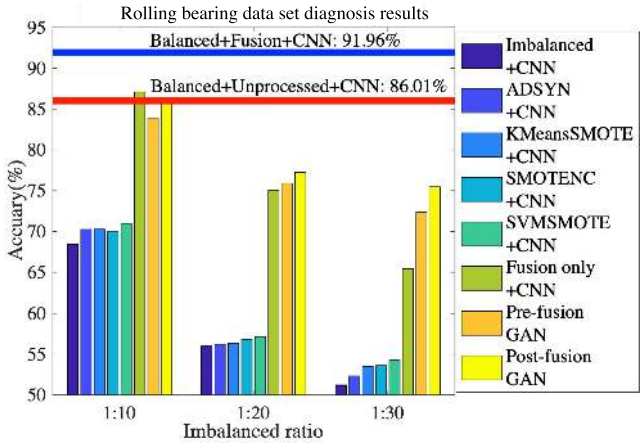


FIGURE 6. Comparison of accuracies between proposed approaches and existing major approaches on the rolling bearing data set.

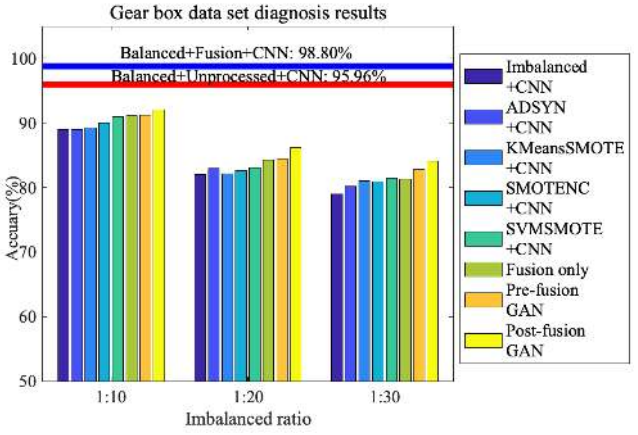


FIGURE 7. Comparison of accuracies between proposed approaches and existing major approaches on gear box data set.

the initial imbalanced data sets with three different imbalanced ratio (1:10, 1:20, and 1:30). According to Eq. (7), for both the Pre-fusion GAN and Post-fusion cases, 450 samples with a sample length of 1200 are fused before classification.

In order to avoid the time-series dependency of segmented data [37], the data under motor speeds of 400 r/min and 600 r/min of the electric motor are used for training, while the data under a motor speed of 200 r/min are used for test.

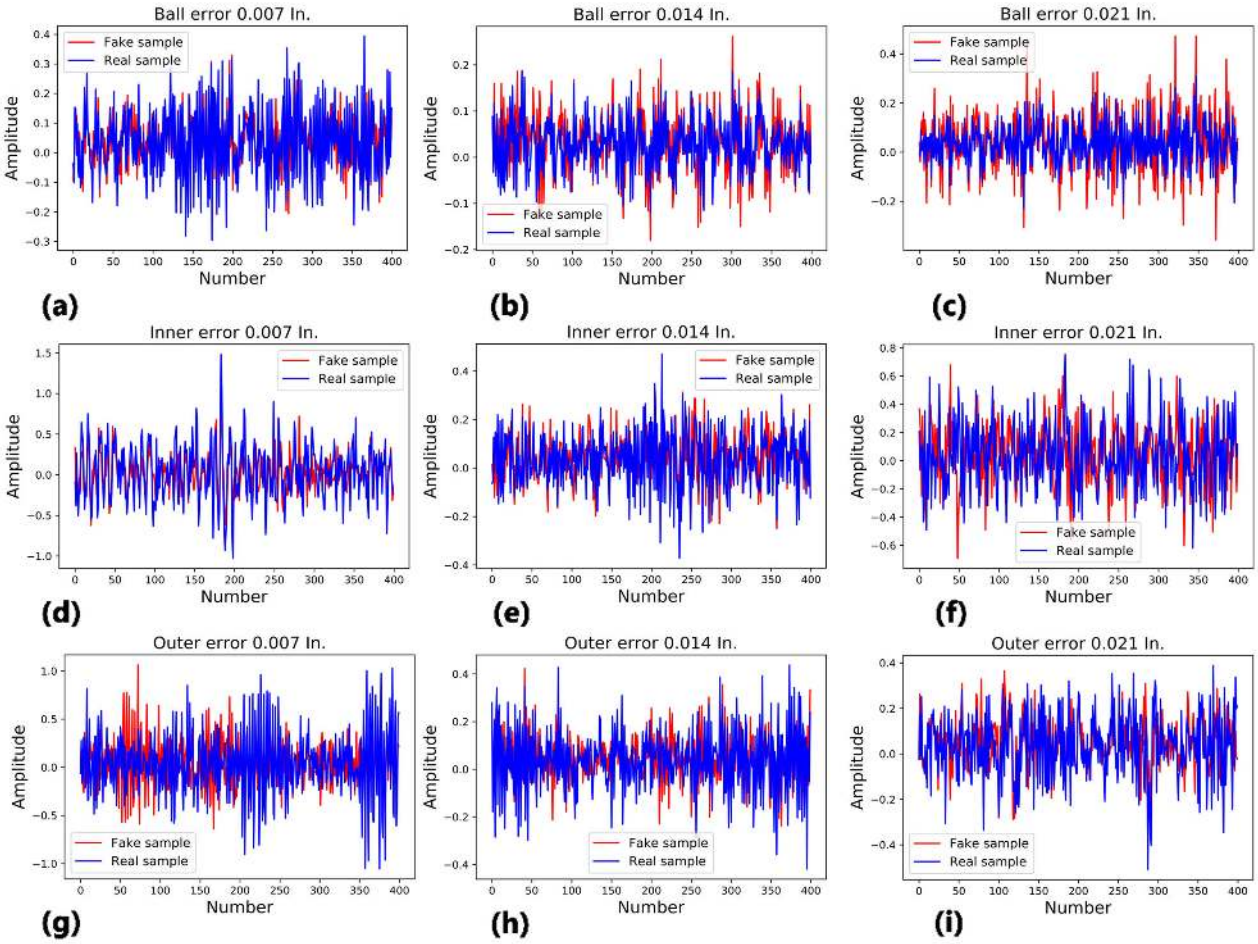


FIGURE 8. Comparative example of synthetic samples generated by GAN and original sample from accelerometer data on fan end (FE): inner race fault with fault diameters of (a) 0.007In., (b) 0.014In., (c) 0.021In., outer race fault with fault diameters of (d) 0.007In., (e) 0.014In., (f) 0.021In., ball fault with fault diameters of (g) 0.007In., (h) 0.014In., (i) 0.021In.

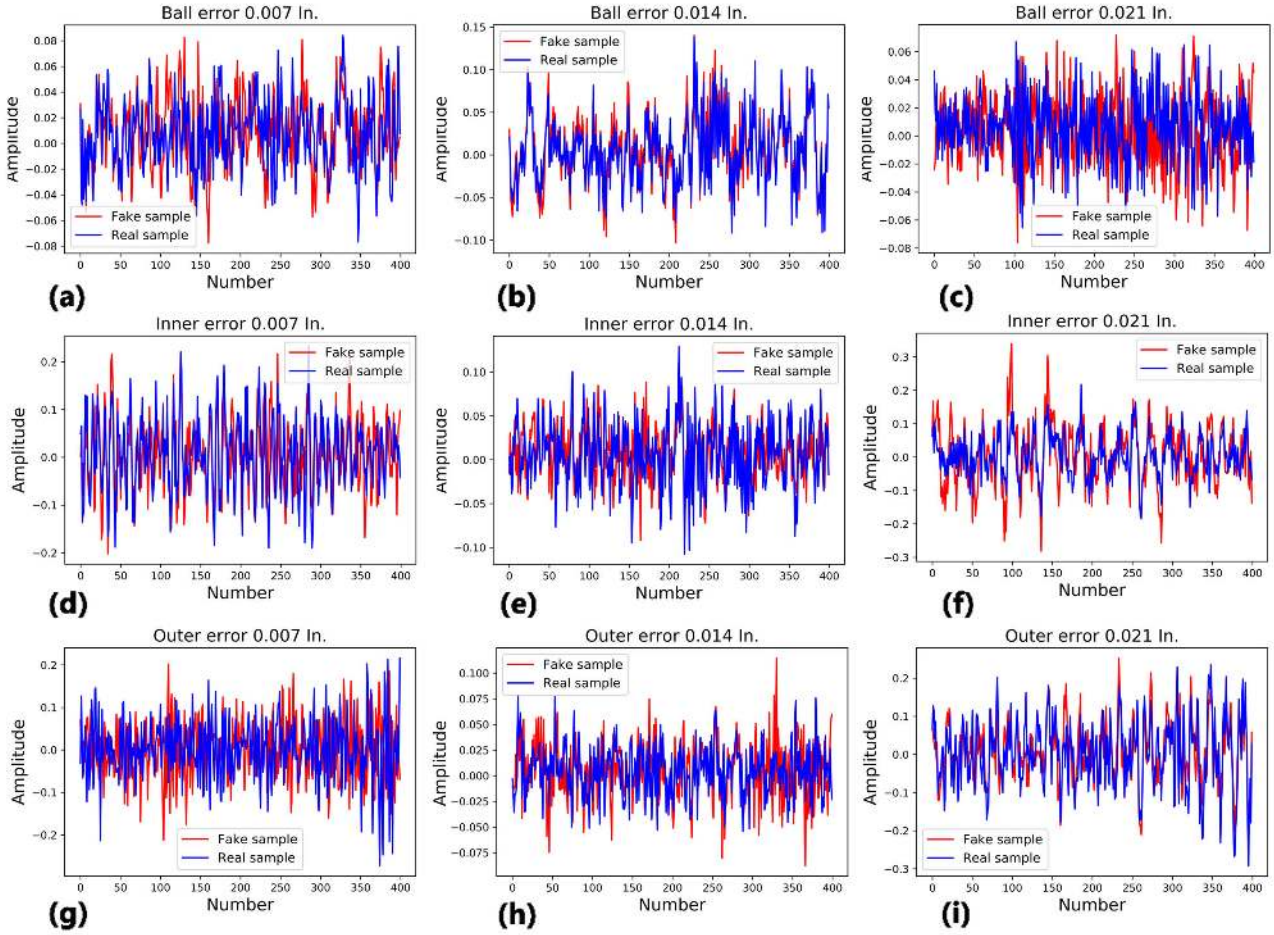


FIGURE 9. Comparative example of synthetic samples generated by GAN and original sample from accelerometer data on base (BA): inner race fault with fault diameters of (a) 0.007In., (b) 0.014In., (c) 0.021In., outer race fault with fault diameters of (d) 0.007In., (e) 0.014In., (f) 0.021In., ball fault with fault diameters of (g) 0.007In., (h) 0.014In., (i) 0.021In.

C. EXPERIMENTAL RESULTS AND ANALYSIS

1) THE LOSSES OF GAN

As shown in Fig. 5, as the losses of generator and discriminator decrease and stabilize with the increasing steps, generated data can be seen as a good representative of original data.

2) ROLLING BEARING DATA

The diagnosis performances of three imbalanced ratios (1:10, 1:20, and 1:30) of different diagnosis approaches are summarized in Table 6 and Fig. 6, this result is obtained from the network architecture as Table 4. In order to better evaluate the proposed algorithm, more metrics such as Area Under Curve (AUC) for Receiver Operating Characteristic (ROC), precision, recall rate, and F-value are applied to assess this algorithm, results are shown in Table 4.

Use imbalanced ratio of 1:10 as a representative case, the accuracy using the proposed Pre-fusion GAN is 83.91%, while the accuracy using the proposed Post-fusion GAN is 85.72%. Compared with the performance of original imbalanced model (unprocessed), more than 10% accuracy increase is achieved. Other diagnosis approaches, such as ADSYN, KMeansSMOTE, SMOTENC, and SVMSOMTE, are used for comparisons with the proposed methods.

In terms of the imbalanced ratios of 1:20 and 1:30, results in Table 6 and Fig. 7 show that the proposed Post-fusion GAN has the best performance. Especially, the post-fusion GAN achieved the highest experimental accuracy compared to other approaches under 1:30 the 1:30 imbalanced. The time series evolution of the representative synthetic vibration data of FE, BA, and drive DE using the proposed GAN network are shown in the Appendix Fig. 8, Fig. 9 and Fig. 10 respectively.

In order to show the effect of the generated data more clearly, the probability distribution function is used to display synthetic and original samples as shown in Appendix Fig. 13. To be noticed, when the imbalanced ratio is 1:10, the results of Pre-fusion GAN and Post-fusion GAN are a little worse than fusion-only method, we think the reason is the imbalanced problem is weakened after the data fusion when imbalanced ratio is small, while synthetic data play a role of noise that reduces the accuracy here.

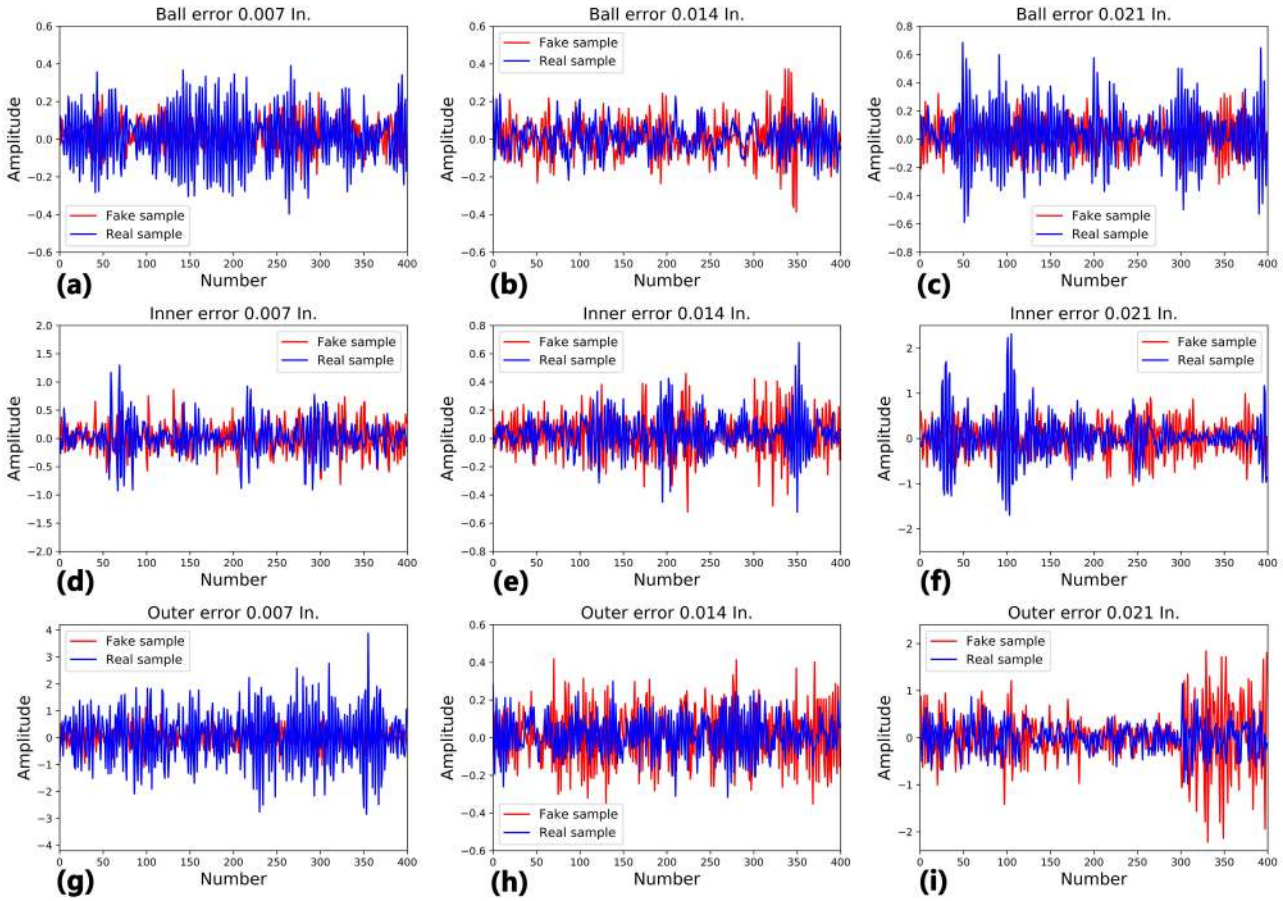


FIGURE 10. Comparative example of synthetic samples generated by GAN and original sample from accelerometer data on drive end (DE): inner race fault with fault diameters of (a) 0.007In., (b) 0.014In., (c) 0.021In., outer race fault with fault diameters of (d) 0.007In., (e) 0.014In., (f) 0.021In., ball fault with fault diameters of (g) 0.007In., (h) 0.014In., (i) 0.021In.

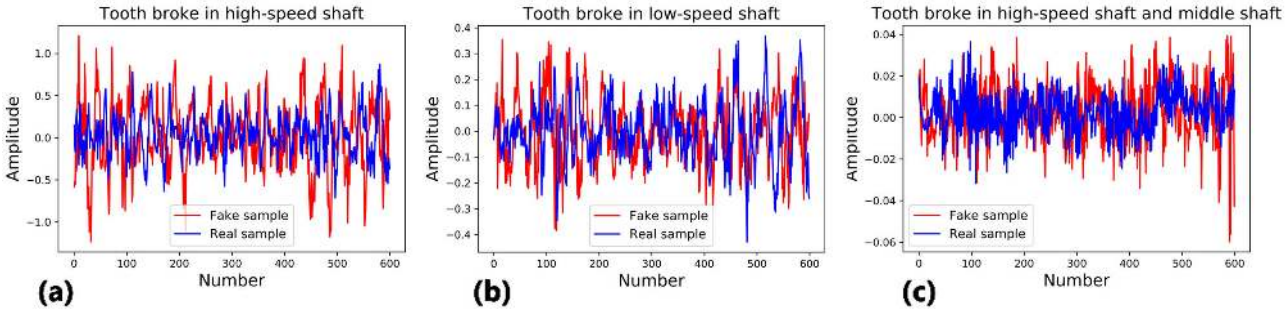


FIGURE 11. Comparative example of synthetic samples generated by GAN and original sample as vibration data: (a) Tooth broke in high-speed shaft gear, (b) Tooth broke in low-speed shaft gear, (c) Tooth broke in high-speed shaft gear and middle shaft gear breaks one tooth.

In addition, with the increase of the imbalanced ratio, the accuracy improvement brought by the proposed methods is more obvious than other methods. From this study, it can be concluded that the more serious the imbalanced problem is, the accuracy difference between the balanced data and the imbalanced data will be greater, which confirms that the compensated data generated by the GAN network captures accurate hidden feature of the fused or sensor signals.

3) GEAR BOX DATA

The prediction performances of three imbalanced ratios and different prediction approaches are summarized in Table 6 and Fig. 7. These results are obtained based on the network architecture shown in Table 5.

The results show that the proposed Pre-fusion GAN and Post-fusion GAN perform well in the cases of 1:10, 1:20, and 1:30, and the Post-fusion GAN performs best among all

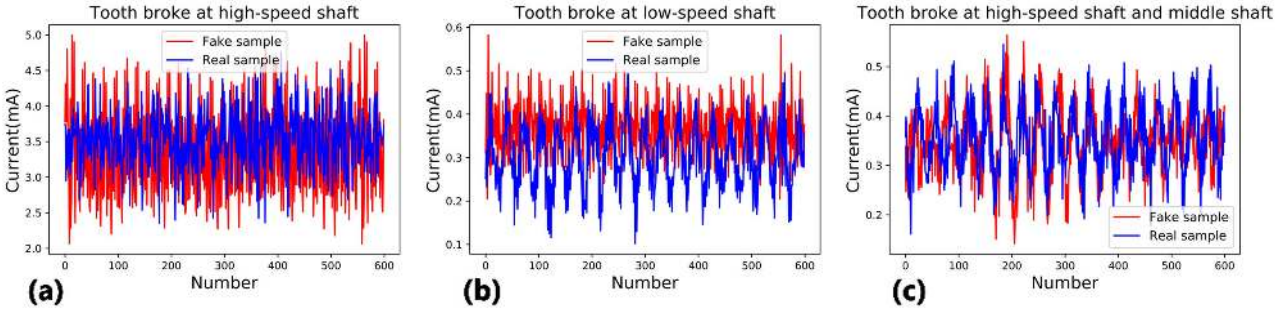


FIGURE 12. Comparative example of synthetic samples generated by GAN and original sample as current data: (a) Tooth broke in high-speed shaft gear, (b) Tooth broke in low-speed shaft gear, (c) Tooth broke in high-speed shaft gear and middle shaft gear breaks one tooth.

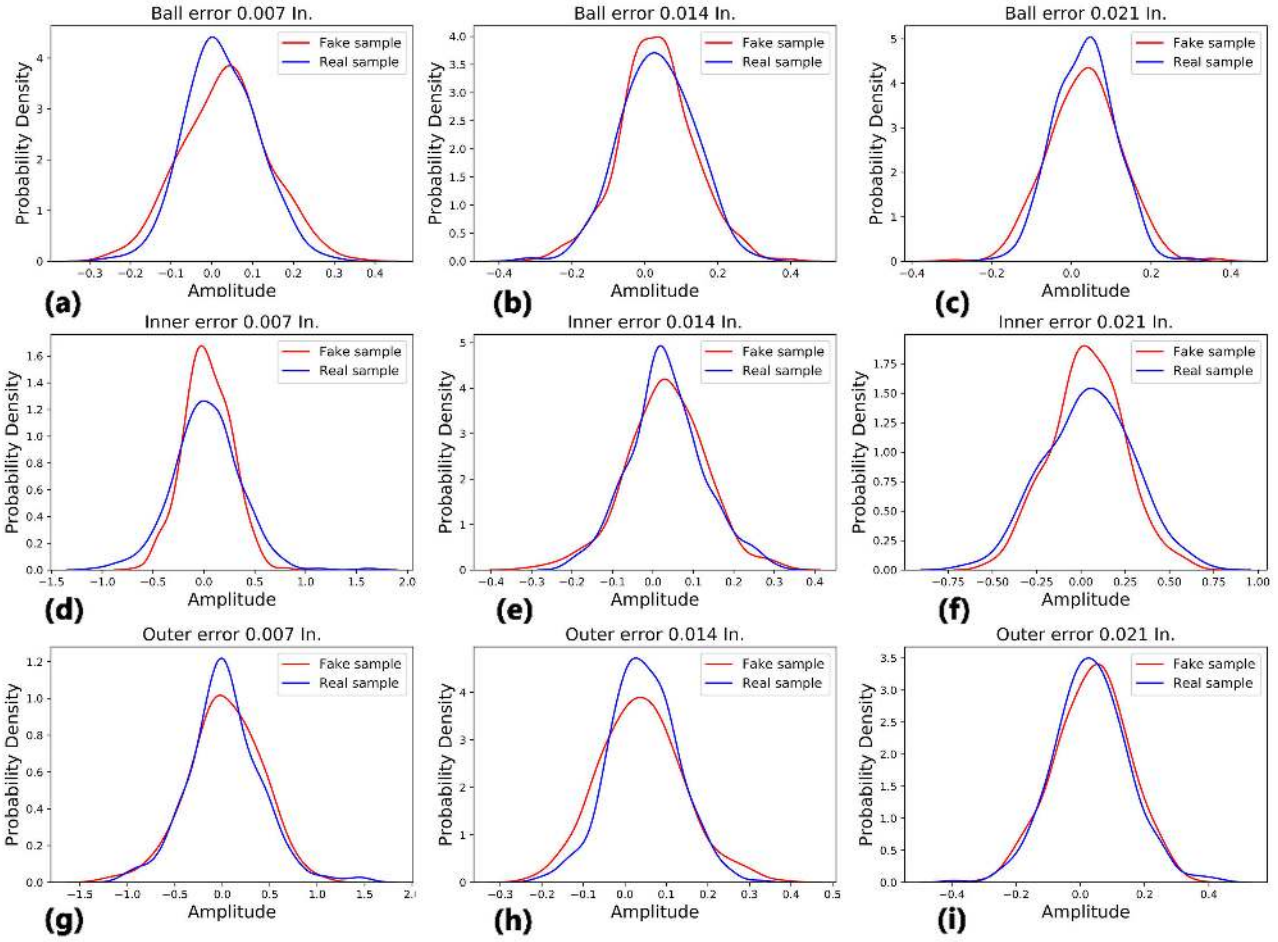


FIGURE 13. Comparative example for probability distribution of synthetic samples generated by GAN and original sample from accelerometer data on fan end (FE): inner race fault with fault diameters of (a) 0.007In., (b) 0.014In., (c) 0.021In., outer race fault with fault diameters of (d) 0.007In., (e) 0.014In., (f) 0.021In., ball fault with fault diameters of (g) 0.007In., (h) 0.014In., (i) 0.021In.

the diagnosis methods. Moreover, with the increase of the imbalanced ratio, the accuracy improvement brought by the proposed methods is more significant than other methods. For the post-fusion GAN with the highest experimental accuracy, the representative synthetic vibration data and current data using the proposed GAN network are shown in the Appendix Fig. 11, and Appendix Fig. 12. In order to show the effect of the generated data more clearly, the probability distribution function is used to display synthetic and original samples, see results in Appendix Fig. 14.

V. CONCLUSION AND FUTURE WORK

In this paper, a new diagnostic framework based on adversarial neural networks (GAN) that integrates a multi-sensor

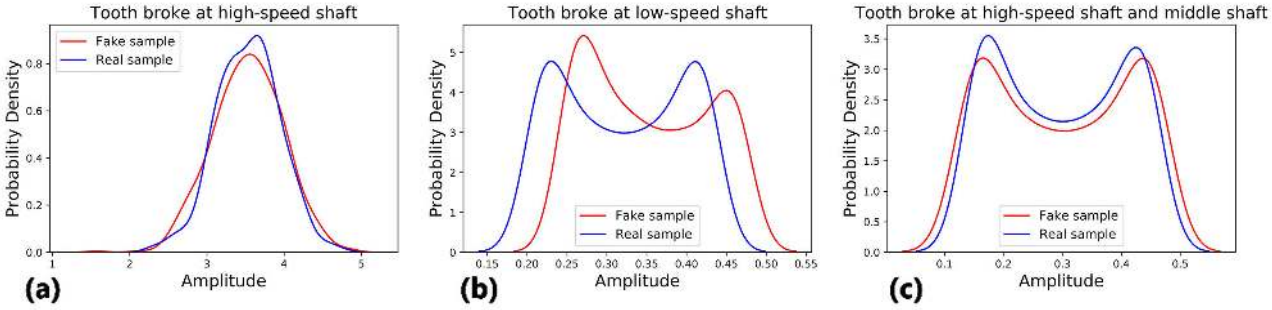


FIGURE 14. Comparative example for probability distribution of synthetic samples generated by GAN and original sample as current data: (a) Tooth broke in high-speed shaft gear, (b) Tooth broke in low-speed shaft gear, (c) Tooth broke in high-speed shaft gear and middle shaft gear breaks one tooth.

data fusion technique to generate synthetic data for data compensation purpose was proposed. Two different practice modes are designed based on this framework according to the position of the data fusion logic. The experiments on different imbalanced ratios are designed on the premise of avoiding the effect of time continuity on the data. The experiments are implemented on two different rotating machinery multi-channel data sets including a rolling bearing data set of CWRU and a data set of the rotating gear box from our laboratory. The multi-channel data generated by different sensors are fused to derive experimental model. The proposed approaches are compared with other widely used data generation methods to verify its effectiveness.

For the future work, we aim to apply the proposed algorithm on more rotating machinery systems to investigate its generalization, and modify the GAN-based network architecture to produce better synthetic data.

APPENDIX

For the experiment on the rolling bearing data set, examples of the time series vibration of the the synthetic data based on the vibration data from accelerometers on FE, BA and DE are shown in as Fig. 8, Fig. 9 and Fig. 10. For the experiment on gear box data from our lab, the synthetic vibration data and current data generated by the GAN network are shown as Fig. 11 and Fig. 12. In order to show the effect of the generated data more clearly, the probability distribution function is used to display synthetic and original samples as shown in Appendix Fig. 13 and Fig. 14.

ACKNOWLEDGMENT

The authors would like to thank Prof. Y. Yuan for his valuable comments and helpful suggestions.

REFERENCES

[1] 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.

[2] Y. Yuan, X. Tang, W. Zhou, W. Pan, X. Li, H.-T. Zhang, H. Ding, and J. Goncalves, "Data driven discovery of cyber physical systems," *Nature Commun.*, vol. 10, no. 1, pp. 1–9, Dec. 2019.

[3] S. Orey and W. E. Pruitt, "Sample functions of the N-parameter Wiener process," *Ann. Probab.*, vol. 1, no. 1, pp. 138–163, Feb. 1973.

[4] F. Gustafsson, F. Gunnarsson, N. Bergman, U. Forssell, J. Jansson, R. Karlsson, and P.-J. Nordlund, "Particle filters for positioning, navigation, and tracking," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 425–437, 2002.

[5] I. Arasaratnam and S. Haykin, "Cubature Kalman filters," *IEEE Trans. Autom. Control*, vol. 54, no. 6, pp. 1254–1269, Jun. 2009.

[6] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006.

[7] G. Ma, Y. Zhang, C. Cheng, B. Zhou, P. Hu, and Y. Yuan, "Remaining useful life prediction of lithium-ion batteries based on false nearest neighbors and a hybrid neural network," *Appl. Energy*, vol. 253, Nov. 2019, Art. no. 113626.

[8] S. Lange and M. Riedmiller, "Deep auto-encoder neural networks in reinforcement learning," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2010, pp. 1–8.

[9] T.-Y. Liu and G.-Z. Li, "The imbalanced data problem in the fault diagnosis of rolling bearing," *Comput. Eng. Sci.*, vol. 32, no. 5, pp. 150–153, 2010.

[10] 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., 7th Int. Symp. Cyberspace Saf. Secur., IEEE 12th Int. Conf. Embedded Softw. Syst.*, Aug. 2015, pp. 1314–1319.

[11] 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, pp. 321–357, Jun. 2002.

[12] 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., IEEE World Congr. Comput. Intell.*, Jun. 2008, pp. 1322–1328.

[13] E. Ramentol, Y. Caballero, R. Bello, and F. Herrera, "SMOTE-RSB: A hybrid preprocessing approach based on oversampling and undersampling for high imbalanced data-sets using SMOTE and rough sets theory," *Knowl. Inf. Syst.*, vol. 33, no. 2, pp. 245–265, 2012.

[14] V. T. Tran, B.-S. Yang, F. Gu, and A. Ball, "Thermal image enhancement using bi-dimensional empirical mode decomposition in combination with relevance vector machine for rotating machinery fault diagnosis," *Mech. Syst. Signal Process.*, vol. 38, no. 2, pp. 601–614, Jul. 2013.

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

[16] 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.*, vol2014, pp. 2672–2680.

[17] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," 2017, *arXiv:1701.07875*. [Online]. Available: http://arxiv.org/abs/1701.07875

[18] J. Donahue, P. Krähenbühl, and T. Darrell, "Adversarial feature learning," 2016, *arXiv:1605.09782*. [Online]. Available: http://arxiv.org/abs/1605.09782

[19] T. Zhang, K. Zhu, and D. Niyato, "A generative adversarial learning-based approach for cell outage detection in self-organizing cellular networks," *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 171–174, Feb. 2020.

[20] X. Gao, F. Deng, and X. Yue, "Data augmentation in fault diagnosis based on the Wasserstein generative adversarial network with gradient penalty," *Neurocomputing*, early access, Apr. 24, 2019, doi: 10.1016/j.neucom.2018.10.109.
[21] Q. Guo, Y. Li, Y. Song, D. Wang, and W. Chen, "Intelligent fault diagnosis method based on full 1-D convolutional generative adversarial network," *IEEE Trans. Ind. Informat.*, vol. 16, no. 3, pp. 2044–2053, Mar. 2020.
[22] Z. Chai and C. Zhao, "A fine-grained adversarial network method for cross-domain industrial fault diagnosis," *IEEE Trans. Autom. Sci. Eng.*, early access, Jan. 6, 2020, doi: 10.1109/TASE.2019.2957232.
[23] J. Jiao, M. Zhao, and J. Lin, "Unsupervised adversarial adaptation network for intelligent fault diagnosis," *IEEE Trans. Ind. Electron.*, early access, Dec. 5, 2019, doi: 10.1109/TIE.2019.2956366.
[24] C. Zhao, C. Chen, Z. Cai, M. Shi, X. Du, and M. Guizani, "Classification of small UAVs based on auxiliary classifier wasserstein GANs," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2018, pp. 206–212.
[25] J. Wang, S. Li, B. Han, Z. An, H. Bao, and S. Ji, "Generalization of deep neural networks for imbalanced fault classification of machinery using generative adversarial networks," *IEEE Access*, vol. 7, pp. 111168–111180, 2019.
[26] G. E. Hinton, A. Krizhevsky, I. Sutskever, and N. Srivastva, "System and method for addressing overfitting in a neural network," U.S. Patent 9 406 017, Aug. 2, 2016.
[27] M.-Y. Liu and O. Tuzel, "Coupled generative adversarial networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol2016, pp. 469–477.
[28] J. Zhao, M. Mathieu, and Y. LeCun, "Energy-based generative adversarial network," 2016, *arXiv:1609.03126*. [Online]. Available: http://arxiv.org/abs/1609.03126
[29] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, *arXiv:1511.06434*. [Online]. Available: http://arxiv.org/abs/1511.06434
[30] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2794–2802.
[31] W. Zhang, "Shift-invariant pattern recognition neural network and its optical architecture," in *Proc. Annu. Conf. Jpn. Soc. Appl. Phys.*, 1988, p. 734, Paper 6P-M-14.
[32] R. Collobert and J. Weston, "A unified architecture for natural language processing: Deep neural networks with multitask learning," in *Proc. 25th Int. Conf. Mach. Learn. (ICML)*, 2008, pp. 160–167.
[33] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980
[34] *Bearing Data Center: Seeded Fault Test Data*. Accessed: Jun. 14, 2019. [Online]. Available: http://csegroups.case.edu/bearingdatacenter/home
[35] *Mad Net*. Accessed: Jun. 14, 2019. [Online]. Available: http://madnet.org:8765
[36] K. Loparo. (2012). *Case Western Reserve University Bearing Data Center*. Accessed: May 11, 2011. [Online]. Available: http://esegroups, case.edu/bearingdata-center/pages/12k-drive-end-bearing-fault-data
[37] Y. Yuan, G. Ma, C. Cheng, B. Zhou, H. Zhao, H.-T. Zhang, and H. Ding, "A general end-to-end diagnosis framework for manufacturing systems," *Nat. Sci. Rev.*, vol. 7, no. 2, pp. 418–429, Feb. 2020.

QIANJUN LIU (Member, IEEE) received the bachelor's degree from the Huazhong University of Science and Technology, Wuhan, China, in 2017, and the M.S. degree from the University of California, Riverside. His research interests include multiagents control, machine fault diagnosis, and deep learning applications.

GUIJUN MA received the bachelor's degree from the Huazhong University of Science and Technology, Wuhan, China, in 2017, where he is currently pursuing the Ph.D. degree in mechanical engineering with the School of Mechanical Science and Engineering. His research interests include machine fault diagnosis and remaining useful life prediction.

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 Imperial College London, U.K. Since 2018, she has been a Postdoctoral Researcher with the Huazhong University of Science and Technology, China. Her research interests include robust control, vehicle dynamics, mechatronic systems modeling and simulation, and deep learning applications.