*Journal of Dynamics, Monitoring and Diagnostics*, 2022, 1, 84-92
https://doi.org/10.37965/jdmd.2022.68

ISTP
RESEARCH ARTICLE

# A Modified Generative Adversarial Network for Fault Diagnosis in High-Speed Train Components with Imbalanced and Heterogeneous Monitoring Data

**Chong Wang,[1] Jie Liu,[1] and Enrico Zio[2]**
[1]School of Reliability and Systems Engineering, Beihang University, Beijing, China
[2]Energy Department, Politecnico di Milano, Milano, Italy

(Received 09 December 2021; Revised 28 March 2022; Accepted 30 March 2022; Published online 02 April 2022)

***Abstract:*** Data-driven methods are widely considered for fault diagnosis in complex systems. However, in practice, the between-class imbalance due to limited faulty samples may deteriorate their classification performance. To address this issue, synthetic minority methods for enhancing data have been proved to be effective in many applications. Generative adversarial networks (GANs), capable of automatic features extraction, can also be adopted for augmenting the faulty samples. However, the monitoring data of a complex system may include not only continuous signals but also discrete/categorical signals. Since the current GAN methods still have some challenges in handling such heterogeneous monitoring data, a Mixed Dual Discriminator GAN (noted as M-D2GAN) is proposed in this work. In order to render the expanded fault samples more aligned with the real situation and improve the accuracy and robustness of the fault diagnosis model, different types of variables are generated in different ways, including floating-point, integer, categorical, and hierarchical. For effectively considering the class imbalance problem, proper modifications are made to the GAN model, where a normal class discriminator is added. A practical case study concerning the braking system of a high-speed train is carried out to verify the effectiveness of the proposed framework. Compared to the classic GAN, the proposed framework achieves better results with respect to F-measure and G-mean metrics.

***Key words:*** braking system; fault diagnosis; generative adversarial network; heterogeneous data; high-speed train; imbalanced data

## 1. INTRODUCTION

With the development of information technology and the increase of automation, industrial equipment are becoming not only more functional but also more complex. Complexity is such that a small fault may cause the loss of the whole system functionality, with possible large economic losses, environmental damages, and injuries to personnel and people from the public [1]. For preventing this to occur, condition monitoring and fault diagnosis technologies are becoming increasingly applied to system health and safety management [2,3]. In particular with the digitization of industry and the associated availability of large amounts of data from system operation, data-driven methods are widely considered for fault diagnosis in complex systems [4,5].

Among the data-driven methods, those of the neural network family have been widely used for fault diagnosis in various complex systems due to other powerful nonlinear modeling ability. For example, using the deep convolutional neural network (DCNN), an efficient fault diagnosis model was proposed in [6] for rotating machinery. By comparison, it was shown that the proposed model is superior to other models in terms of accuracy, memory space, computational complexity, transfer performance, etc. Reference [7] presented a condition monitoring and fault diagnosis method for wind turbine generator (WTG), using artificial neural network (ANN), and empirical mode decomposition (EMD). The proposed approach can identify the differently imbalanced faults in WTG. Again, the deep neural network (DNN) was adopted to recognize faults in high-speed train bogies in [8]. This fault diagnosis model achieved very high diagnostic accuracy when applied to high-speed trains with different speeds and different faults. In [9], an efficient fault diagnosis method for smart grid systems was designed based on long short-term memory (LSTM) recurrent neural networks. And, using echo-state networks, the multiclass classification task can be processed by application of different dimensionality reduction techniques.

However, neural networks typically need to be trained based on large amounts of relevant data to obtain a high-precision model. In practice, the monitoring data from real applications are imbalanced, especially for high-reliability systems. Due to the large contribution of the majority class samples in model fitting, the classification hyperplane of a deep learning-based fault diagnosis model may be biased toward the majority class with imbalanced data, resulting in a quite poor classification accuracy on the minority class. For fault diagnosis, the data related to faults are scarce and with such imbalanced data, satisfactory results cannot be achieved.

Nowadays, there are three main categories of methods to deal with imbalanced data [10,11]:

(1) Resampling methods, such as undersampling [12], oversampling [13,14], and heterogeneous sampling [15];
(2) Algorithm-level methods, such as modifying loss function [16], modifying classification threshold [17], and cost-sensitive learning [18];

Corresponding author: Jie Liu (e-mail: liujie805@buaa.edu.cn)

(3) Ensemble learning, such as using boosting ensembles to conduct iterative training [19].

These methods improve the classification accuracy, as verified in various fields [11]. Among them, the Synthetic Minority Over-sampling TEchnique (SMOTE) is a popular method to adjust the data distribution by adding synthetic minority class samples, so as to improve the classification performance.

The recently proposed generative adversarial network (GAN) has also been used to generate synthetic samples. Differently from SMOTE and its variants, which rely on expert knowledge for designing synthetic minority generation rules, a GAN method can learn automatically the inherent distribution and generate minority samples similar to the real ones. A GAN includes two variable networks: a generator and a discriminator (respectively denoted as G and D), which are trained in a mutual game [20]. The generated samples of G are judged and evaluated by D, and then G is optimized according to the evaluation results. By so doing, the efficiency and quality of the sample generation process can be greatly improved [20]. At present, GAN and its variants have been successfully applied in many fields, such as image inpainting [21], scene synthesis [22], and face recognition [23].

For imbalanced data in fault diagnosis, the research based on GAN has been developed gradually in recent years. For example, a new fault diagnosis method based on GANs was proposed in [24] to process insufficient vibration monitoring data of rotating machinery. As the actual fault data of bearing are limited, an imbalanced fault diagnosis model based on GAN was established in [25], and a detailed comparative study was conducted. Reference [26] proved that a new bearing fault diagnosis method based on Switchable Normalization Semi-Supervised GAN (SN-SSGAN) is effective. Based on an improved GAN, the dataset for fault diagnosis of reciprocating machines was balanced in [27]. In [28], synthesized data generated by a Dual-Discriminator Conditional GANs (D2CGANs) was used for data expansion, which evidently improved the accuracy of fault diagnosis for rolling bearing. An Auxiliary Classier Wasserstein GAN with Gradient Penalty (ACWGAN-GP) was proposed in [29], which can stably generate high-quality transmission gear fault samples. Reference [30] combined the Semi-Supervised GAN with Wavelet Transform to design a new fault diagnosis model (noted as WT-SSGAN) for application to rotating machinery. Reference [31] proposed a method for generating fault samples of rolling bearings based on GAN and autoencoder (AE) and verified the effectiveness of this method. GAN was effectively employed to balance the training dataset of automatic fault detection and diagnosis (AFDD) for chillers in [32]. In [33], a method based on Conditional Variational AE and GAN (CVAE-GAN) was proposed for imbalanced fault diagnosis of planetary gearbox. Reference [34] designed a multilabel one-dimensional GAN (1-DGAN) fault diagnosis framework and confirmed that the generated data are informative. Other similar works have been published, and interested readers can refer to the review of [35].

The previous work focuses mainly on continuous and high sampling frequency data, whereas the actual data often contain also both (low sampling frequency) numeric and categorical variables, which make the so-called heterogeneous data. For example, factors that affect the braking system of high-speed trains include numerical variables such as voltage and current, as well as categorical variables, such as operating mode and braking status. The categorical and some of the numerical variables are discrete. Proper GAN structure for generating discrete values remains a challenge, as pointed out in [35]. Although some GAN models have been proposed for discrete variables, as in [36], some difficulties still need to be solved. To the best knowledge of the authors, no published work has reported a solution for the heterogeneity of the monitoring variables with GAN. The common way is to encode directly the discrete variables into numeric ones by one-hot code [37]. However, the direct use of one-hot code is likely to cause information loss, and the values of the samples generated by GAN with one-hot code are no longer discrete and may exceed the values range of the original discrete values, with no engineering interpretation. In addition to the data heterogeneity, the previous work assumes the availability of enough reconstructed fault data for GAN training, with a sampling frequency as high as several thousand hertz. This is because training a GAN model on limited fault samples may cause severe over-fitting or under-fitting problems. Normal/majority samples which are large in size can be borrowed for improving the performance of a GAN model.

In this paper, a fault sample generation method with heterogeneous imbalanced monitoring data is proposed by a modified GAN (Mixed Dual Discriminator GAN, M-D2GAN). Based on the proposed M-D2GAN, simulation samples can be generated with the same data structure and distribution as the actual fault samples and without aggravating the class overlapping, which can allow effectively processing imbalanced data when used for fault diagnosis. Also, the addition of these generated samples will not affect the impact of the fault diagnosis model on the real data structure. To make the generated fault samples more in line with the actual situation, different types of variables, such as floating-point type, integer type, indifferent category type, and hierarchical category type, are output in the last layer of the generator G in different coding and output forms. A double discriminator framework is used to avoid the generated samples to enhance class intersectionality. The first discriminator D is used to discriminate whether the generated sample is real, and the second discriminator F is used to discriminate whether the generated sample is a faulty sample. Firstly, F is trained based on the real normal samples and, then, G and D are trained iteratively based on real fault samples and F. Using the real monitoring data from a high-speed train braking system, the proposed method is proved to be effective in generating fault samples. A comparative experiment is performed in two steps. In the first step, it is verified that the samples generated by M-D2GAN can effectively improve the accuracy of fault diagnosis models. In the second step, it is verified that the proposed model improves performance compared to the classical GAN.

The rest of this manuscript is structured as follows. In Section 2, the proposed M-D2GAN for generating heterogeneous fault samples is introduced in detail. Comparative experiments based on real monitoring data from a high-speed train are considered in Section 3. Section 4 draws the conclusions and gives some perspective.

## 2. THE PROPOSED M-D2GAN FOR FAULT DIAGNOSIS WITH HETEROGENEOUS AND IMBALANCED MONITORING DATA

### A. BRIEF INTRODUCTION OF GAN

GAN is proposed in [20] to generate samples by an adversarial process in which two networks, i.e. a generator G and a discriminator D, are trained simultaneously and compete against each other. The structure of a classic GAN is shown in Fig. 1.

In [20], the two competing networks are of the multilayer perceptron (MLP) type and can be trained with back propagation. G is trained to generate samples similar to the real data, whereas D is used to calculate the probability that a generated sample comes from the same distribution of the real samples. By capturing the distribution of real data, G generates real-like samples. The training goal of G is to make D unable to distinguish the generated samples from the real ones, i.e. the probability calculated by D approaches 0.5.

For the real sample $x$ and a Gaussian random noise $z$, the generated sample can be defined as $G(z)$. Then, the objective of GAN is a min-max optimization with objective value function $V$(G,D) as follows:

$$\min_{\mathrm{G}} \max_{\mathrm{D}} V(\mathrm{D,G}) = \mathrm{E}_{x\sim p_{data}}[\log D(x)] + \mathrm{E}_{z\sim p_z}[\log(1 - D(G(z)))] \quad (1)$$

where $D(*)$ is the probability that a sample comes from the same distribution as the real data rather than G, $\mathrm{E}(*)$ represents the calculated expectation, and $p_{data}$ and $p_z$ are, respectively, the distributions of $x$ and $z$.

GAN is trained in an iterative manner, so to optimize G and D alternatively. First one can fix G and optimize D by:

$$\max_{\mathrm{D}} V(\mathrm{D,G}) = \mathrm{E}_{x\sim p_{data}}[\log D(x)] + \mathrm{E}_{z\sim p_z}[\log(1 - D(G(z)))] \quad (2)$$

After the optimization of D, G can be optimized by:

$$\min_{\mathrm{G}} V(\mathrm{D,G}) = \mathrm{E}_{z\sim p_z}[\log(1 - D(G(z)))] \quad (3)$$

This is repeated iteratively updating the two networks until the GAN system reaches a dynamic equilibrium (as Nash equilibrium). By setting the output structure of G, several new samples can be generated that follow the $x$ distribution.

In fact, the generator G and the discriminator D are not trained alternately, rather D is trained $t$-times and G is trained once, simultaneously. This is because in the early training period, the generated sample $G(z)$ is easily negated by D, i.e. $D(G(z))$ is very small, resulting in $\log(1 - D(G(z)))$ quite close to 0. In this case, G should be optimized by $\max_G[\log(D(G(z)))]$ instead of $\min_G[\log(1 - D(G(z)))]$.

### B. THE PROPOSED FAULT DIAGNOSIS FRAMEWORK WITH M-D2GAN

As discussed in the Introduction, there are several challenges for building a GAN model for heterogeneous and imbalanced data. A M-D2GAN is proposed in this work for tackling the heterogeneous imbalanced data problem. The structure of the proposed method is shown in Fig. 2.

For the original heterogeneous data with $N$ samples $X = (X_1, X_2, \ldots, X_K)$, $X_i = (x_{i1}, x_{i2}, \ldots, x_{iN})^{\mathrm{T}}$ $(i = 1,2,\ldots,K)$, different types of variables need different encoding strategies before they can be input into the fault diagnosis model. Thus, in the encoded data, one column refers to a certain original variable. The last layer of G needs to generate samples of the same size as the samples in the encoded data. Let the columns represent $X_i$ in the encoded data or generated data as the corresponding columns of $X_i$ (hereinafter referred to as the columns corresponding to $X_i$). The proposed M-D2GAN develops in two steps:

(1) Identify different types of variables and design the output rules of the last layer in the generator G for the columns corresponding to the different types of variables. In general, the heterogeneous monitoring data can be divided into the following four types: floating-point type, integer type, indifferent category type, and hierarchical category type. The different softmax formats for columns corresponding to the different types of variables are as follows: for floating-point variables, the output columns are not processed; for integer and hierarchical category variables, the ones are rounded to the nearest whole number; the Gumbel-softmax applied in [38] is used to generate indifferent category variables.

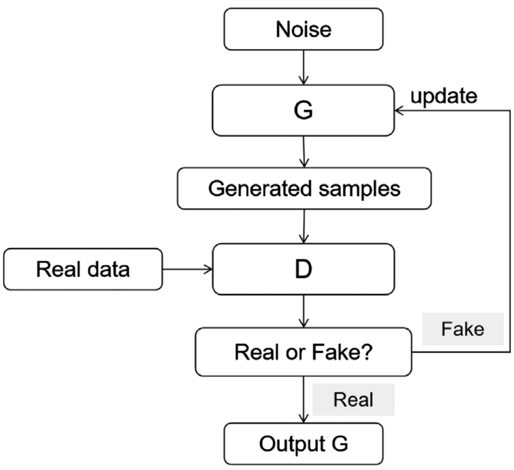


**Fig. 1.** The structure of a classic GAN.

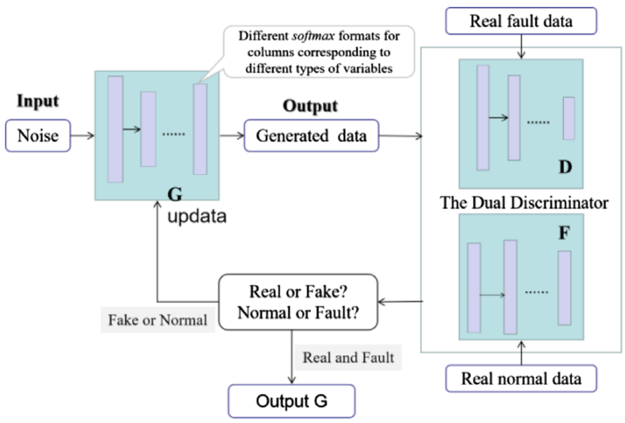


**Fig. 2.** The structure of M-D2GAN.

(2) Build a double discriminator to discriminate the generated samples, in which the discriminator D distinguishes whether the generated samples are real and F distinguishes whether the generated samples are normal. Then, the training goals of G are to make D unable to distinguish the generated samples from the real ones and to make F able to distinguish the generated samples from normal ones. The calculated probability of D approaches 0.5, and the calculated probability of F approaches 0.

## C. CODING AND OUTPUT RULES FOR HETEROGENOUS DATA

The coding and output rules for different types of variables are designed as follows:

(1) Floating-point variables are directly input into D, F, and the fault diagnosis model. When generating samples in G, the corresponding columns do not need to be processed;

(2) Integer variables are also directly input into D, F and the fault diagnosis model. When generating samples in G, the corresponding columns in G are rounded;

(3) Hierarchical category variables are input to the models after label coding, according to the grade from small to large. When generating samples in G, set the output range for the corresponding columns and round them;

(4) Indifferent category variables are input to the models after one-hot coding. When generating samples in G, *Gumbel-softmax* is used to obtain the columns corresponding to the coded variables.

The classical GAN can handle well continuous data, and the optimal model can be obtained by first-order differential calculations. For discrete variables, the sampling process is nondifferentiable, so the parameters are difficult to update. Therefore, the columns corresponding to continuous variables do not need to be processed in the G output layer. Integer variables can also be directly input into the models, but the corresponding columns need to be rounded when generating samples. For category variables with levels, label coding according to the grade from small to large is applied as a numerical method. And, the corresponding columns need to be rounded uniformly with integer variables in the output layer of G. Moreover, the differentiable *Gumbel-softmax* in [38] can be used to obtain indifference discrete variables. To improve the efficiency and make the generated data conform to practical significance, the output of the generator G can be designed to remain within a value range. Here, the value range of each variable can be obtained according to experience and the distribution of the existing data.

In order to prevent information loss, based on one-hot encoding technology, indifference discrete variables are converted into multidimensional vectors, suitable for efficient Boolean constraint propagation [39]. Thus, one-hot encoding is used in this work for encoding the indifference discrete variables. Similarly, to generate indifference discrete variables, the output of G needs to work with one-hot vectors, usually obtained by implementing the *softmax* function in the output layer. Instead of the classical *softmax* function, the *Gumbel-softmax* function is appended after the decoder in this paper.

For the $d$-dimensional vector $h = (h_1,h_2,\ldots,h_d)$, the corresponding $d$-dimensional one-hot vector $y = (y_1,y_2,\ldots,y_d)$ can be determined by the maximum value of the $h$'s components. For example, for (1,5,8,0.2), whose third component is the largest, the corresponding one-hot vector is (0,0,1,0), which can represent a certain category. Obviously, this method cannot calculate the gradient nor can it update the network. The usual improvement method is to normalize the vector by a *softmax* function so that the gradient can be calculated and the obtained value can represent the probability:

$$p = softmax(h) = (p_1,p_2,\ldots,p_d) \tag{4}$$

where $p_i = \exp(h_i)/\Sigma_{j=i}^{d}\exp(h_j), i = 1,2,\ldots,d$.

The *softmax* function tends to make the probability of the maximum value significantly larger than other values, but the vector $p$ which represents the probability has no probability meaning and the sampling process of output from *softmax* is nondifferentiable during the model training process. In this case, based on *softmax* transformation, a differentiable function named as *Gumbel-softmax* is proposed to make the generated data more realistic [36]:

$$y\prime = softmax\left(\frac{1}{\tau(h+g)}\right) \tag{5}$$

where $g = (g_1,g_2,\ldots,g_d)$ and $g_i(I = 1,2,\ldots,d)$, are independent of each other and conform to a *Gumbel* distribution, and $\tau$ is a control parameter of the soft degree in the *softmax* function. When $\tau \to 0$, $y\prime \to y$, and when $\tau \to \infty$, $y\prime$ approximately satisfies a uniform distribution.

Based on the processing method of different types of variables described in the previous paragraphs, samples heterogeneous as the real samples can be generated. Then, in GAN training and fault diagnosing, the generated samples can allow achieving higher performance.

In addition, in this paper, the most commonly used rounding method is adopted for getting integer and hierarchical category variables, i.e. the nearest integer value.

***1) GAN FOR IMBALANCED DATA.*** In this work, the GAN takes a dual discriminator for tackling imbalanced data. In the current use of GAN for fault diagnosis, convolution and deconvolution layers are widely adopted for tackling high-frequency monitoring data [35]. Actually, the structures of G, D, and F are all flexible, and the model can be adjusted according to the actual needs of practice. In this work, MLP network is selected to build G, D, and F. The input of G is random Gaussian noise, and the output is synthetic faulty samples. The synthetic faulty samples are the input of D and F. Considering the limited fault data, ANNs, instead of a deep learning layout are adopted for building D and F. ANNs can recognize the complex nonlinear relationship between input and output data and have been widely applied to solve complex practical problems [40]. Moreover, for these models, the number of nodes and hidden layers are gradually increased for the considered problem until the model's performance does not change or decline.

In this case, the generator G is a neural network whose input is noise $z$ and output is a group of simulated samples following the same structure as the actual data. The number of output samples is usually set to be the same as the number of real normal samples participating in training D, so as to avoid D being affected by data imbalance problems. Thus, G can be considered as mapping $\tilde{x} = G(z)$, and the training goal of G is to make $\tilde{x}$ obey the distribution of $x$.

For fault diagnosis in a high-reliability system, the generated samples should not only be similar to the real fault data but also different from the real normal data. Thus, an extra item is added to the classic loss function in Eq. (1), to prevent the generated data from exacerbating the class overlapping problem. Finally, the objective function in this work is

$$\min_G \max_D V(D,G) = E_{x\sim p_{data}}[\log D(x)] + E_{z\sim p_z}[\log(1 - D(G(z)))] + E_{z\sim p_z}[\log(F(G(z)))] \quad (6)$$

where $F(*)$ is the probability that a generated sample is in normal state, $D(*)$ is the probability that a sample comes from the same distribution as the real data rather than G, $E(*)$ represents the calculated expectation, and $p_{data}$ and $p_z$ are, respectively, the distributions of $x$ and $z$.

Since the normal samples are large in size, we can get a good F just by training the MLP model with the normal samples. The MLP encoder trained with back propagation based on normal samples is considered to be a one-class learning model, which can judge whether new samples belong to the normal class. Considering that if the generated samples or fault samples are added to the training model F, the generated samples may locate closely to the actual fault samples, resulting in failure diagnosis unable to identify random faults and nonoccurring faults. In this work, the generated samples are expected to deviate from the normal samples. Thus, the model F is no longer optimized in the process of training M-D2GAN. Moreover, M-D2GAN is also trained in an iterative manner, and the objective function can optimize G and D alternatively. When G is fixed, the loss function of D is

$$\max_D V(\mathrm{D,G}) = E_{x\sim p_{data}}[\log D(x)] + E_{z\sim p_z}[\log(1 - D(G(z)))] \quad (7)$$

And, when D is fixed, G can be optimized by:

$$\min_G V(\mathrm{D,G}) = E_{z\sim p_z}[\log(1 - D(G(z)))] + E_{z\sim p_z}[\log F(G(z))] \quad (8)$$

In this work, Adam algorithm [41], which is insensitive to hyperparameters, is used for model optimization. The process of generated data is offline, without the need of consideration on the amount of calculation. Therefore, in order to prevent the model from reaching local optimization, we set a low learning rate as 0.001. Moreover, same as in the classic GAN training process introduced in Section A, actually, D is trained t-times and G is simultaneously trained once in M-D2GAN.

## 3. APPLICATION RESULTS

The braking system is one of the most important components of high-speed trains, ensuring their effective deceleration during operation. Due to the requirements of high reliability, there have been many related fault diagnosis studies for high-speed train braking systems, including expert systems [42,43], physical mechanism analysis [44,45], and data-driven modeling techniques [46,47]. Since high-speed train braking system is a complex system, it is difficult to analyze the function subdivision and corresponding failure mechanisms. Using traditional fault diagnosis methods consumes a lot of manpower and material resources and often fails to achieve good results. With the development of information technology, sensors for system monitoring are widely used, and data-driven fault diagnosis is developing rapidly.

Data related to the state of the braking system can be collected by arranging sensors and capturing system state information. Data-driven fault diagnosis technology has made some breakthroughs in high-speed train braking systems. For example, based on support vector machine (SVM), high precision and stability fault diagnosis frameworks for high-speed train braking systems were proposed in [48,49]. These optimized SVM models can achieve good results in the classification of highly imbalanced data, which is helpful for fault diagnosis of complex systems.

Based on the monitoring dataset of a high-speed train braking system in 1-year operation, in this section, a comparative experiment is designed to verify that the samples generated by the proposed method can significantly improve the fault diagnosis accuracy. In addition, to show the improvement of the proposed method on GAN, the classic GAN with one-hot code is considered as a comparison benchmark in the experiment. The classic GAN here refers to the model with one generator G and one discriminator D. G and D are all MLP networks. The process of this comparative experiment is shown in Fig. 3.

In order to avoid the influence of random factors on the results of the comparison, different commonly used classification algorithms are considered as the fault diagnosis model in this experiment, including logistic regression (LR), K-nearest neighbor (KNN), MLP, and convolutional neural network (CNN). A fivefold cross-validation is used to compare the average generalization accuracy of each fault diagnosis model. The actual training data for the fault diagnosis model under fivefold cross-validation are training data and generated data. All the models are optimized by gradually increasing structural complexity and iterations, until the generalization accuracy remains stable or decreases.

The monitoring dataset contains 43 variables related to brake system failure, including 18 floating-point variables,

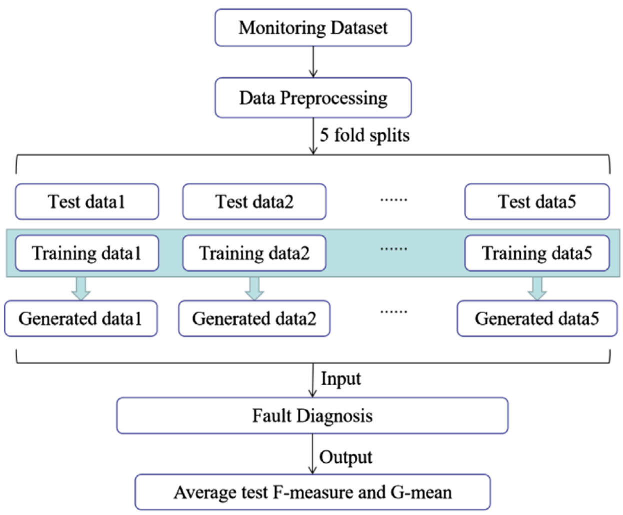


**Fig. 3.** Flow chart of the comparative experiment.

3 integer variables, 2 hierarchical category variables, and 20 indifferent category variables, such as speed, voltage, current, temperature, traction effort, level of access, operation mode, and braking state. Due to the restriction of confidentiality, we do not list these variables with their names and represent them as V1, V2 . . . V43. These samples are marked as failure or normal state without distinguishing different faults. Before generating new fault samples by different methods, necessary data processing, such as data cleaning and standardization, are performed.

Figure 4 is a 2D projection of the dataset, based on t-distributed stochastic neighbor embedding (t-SNE), where red 'o' represents fault samples and gray '*' represents normal ones. The t-SNE is a common algorithm for exploratory analysis of high-dimensional data. It can project high-dimensional data into 2-D or 3-D spaces by converting the similarity between data points into probability, so as to realize the visualization of high-dimensional data [50]. Moreover, only 10% of the normal samples are used in the t-SNE projection, because otherwise the fault samples would be completely covered when using all normal data. It has been confirmed that the projection distribution of the 10% normal samples is roughly consistent with that of the whole normal data.

There are 28 837 normal samples and 159 fault samples in the original monitoring dataset, which means that the data are highly imbalanced (this can also be seen from Fig. 4). Thus, the classification interface of the fault diagnosis model will seriously deviate to the normal state, resulting in the prediction accuracy of fault class being extremely low. Although we pay more attention to the fault state than to the normal state, too few fault samples have little effect on the overall accuracy. Therefore, the overall generalization accuracy cannot be the indicator for comparing generating methods; rather, only the fault state should be focused to meet the actual needs. The precision and recall of the fault state can be used as the evaluation indexes for the generated samples.

However, from Fig. 4, we can see that the data of different states are largely overlapped. Therefore, high precision and high recall are impossible to be achieved simultaneously. In this case, *F-measure* and *G-mean* are applied as the comparison metrics to comprehensively evaluate the effects of different methods.

Whereas *TP*, *TN*, *FP*, and *FN*, respectively, represent the number of true positives, true negatives, fault positives, and fault negatives, *F-measure* and *G-mean* can be calculated by:

$$F - measure = 2 * \frac{precision * recall}{precision + recall} \tag{9}$$

$$G - mean = \sqrt{TPR * TNR} \tag{10}$$

with $precision = TP/(TP + FP)$, $recall = TPR = TP/(TP + FN)$, and $TNR = TN/(TN + FP)$.

As shown in Fig. 3, in the fivefold cross-validation, M-D2GAN is trained based on each training data to obtain a satisfactory generator, and then sufficient generated samples are generated to reduce the imbalance between classes. Here, the generator and discriminators are MLPs with three hidden layers, and the number of nodes are set as the commonly used ones, i.e. 16, 32, 64, 128, etc., and the parameters are fine-tuned for different training data. The original 43 variables are encoded by the coding method presented in the previous section and transformed into 45 new variables. Thus, the input layers of D and F contain 45 nodes, and the outputs are one-dimensional values. The input layer of G contains one node, and the output is a $127 \times 45$-dimensional data group. Based on each training data, after using M-D2GAN to obtain an excellent generator in the distribution of $z$, several batches of generated samples can be obtained by inputting different $z$ to eliminate the imbalance between classes in the training data.

Different classification algorithms are used as the fault diagnosis model to verify the effectiveness of the proposed method. After comparison, the data in this paper can get better feature extraction effect by using the network with three hidden layers. Therefore, MLP and CNN fault diagnosis models are also three-layer hidden layer structures. In the same operating environment, the experimental results are reported in Table I.

It can be seen that imbalanced data will seriously affect the fault diagnosis result, and both the proposed method and the classical GAN can improve the fault prediction accuracy by adding fault samples. Among them, the proposed M-D2GAN can further improve the fault diagnosis results. It is because of this that the samples generated by M-D2GAN are close to the distribution of real fault samples and significantly different from the normal ones, which makes the new samples to be more easily identified.

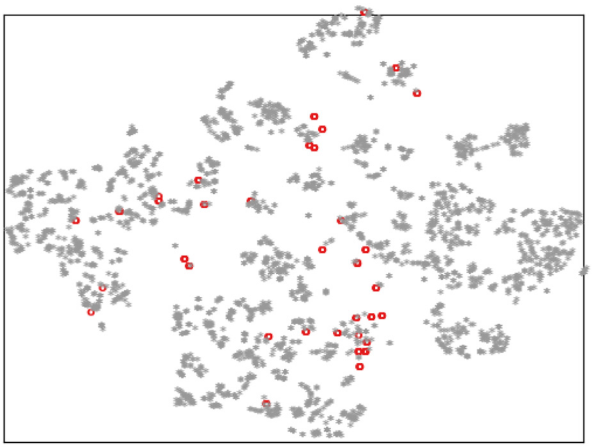

**Fig. 4.** The 2-D t-SNE projection for the dataset.

**Table I** Comparison of different generating methods

| ***Model*** | ***Method*** | ***F-measure*** | ***G-mean*** |
|---|---|---|---|
| LR | Untreated | 0.21884 | 0.42733 |
| | GAN | 0.42424 | 0.79723 |
| | M-D2GAN | 0.56111 | 0.82432 |
| KNN | Untreated | 0.21429 | 0.47034 |
| | GAN | 0.66316 | 0.78208 |
| | M-D2GAN | 0.72414 | 0.82305 |
| MLP | Untreated | 0.53020 | 0.68825 |
| | GAN | 0.66949 | 0.78591 |
| | M-D2GAN | 0.75238 | 0.84135 |
| CNN | Untreated | 0.38095 | 0.81494 |
| | GAN | 0.77396 | 0.8559 |
| | M-D2GAN | 0.82245 | 0.88741 |

It can also be seen from the t-SNE projections of the datasets balanced with different methods that

(1) M-D2GAN and GAN in the experiment can improve the separability of the elements in the dataset, thereby improving the accuracy of the fault diagnosis model;

(2) Comparing Figs. 5 and 6, one finds out that the conventional GAN may overfit the scarce fault data and that it generates synthetic samples close to each the original fault data, showing a large number of clusters; on the contrary, the proposed M-D2GAN generates synthetic samples with a better generalization capability, which can also be seen from the comparison results in Table I;

(3) Fundamentally, the samples generated by M-D2GAN have the same heterogeneous structure as the original ones, i.e. discrete, nominal, and continuous variables can be generated within the same value ranges, while using multisource information fusion technology, the effective features of different types can be extracted by different activation functions and then fused into the diagnosis model.

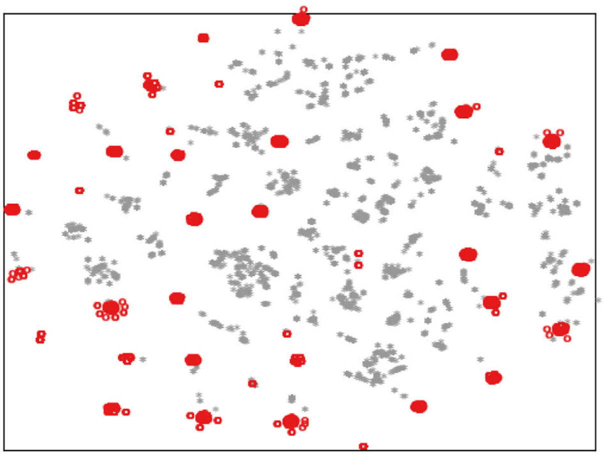

**Fig. 5.** The 2-D t-SNE projection for the dataset balanced with GAN.

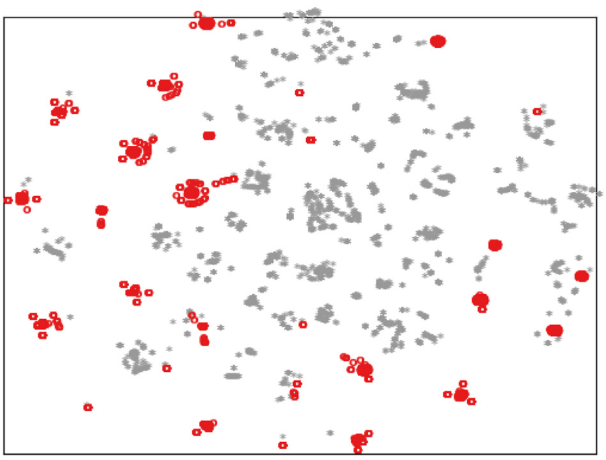

**Fig. 6.** The 2-D t-SNE projection for the dataset balanced with M-D2GAN.

## 4. CONCLUSIONS

In Prognostics and Health Management (PHM) of high-reliability systems, imbalanced data appear, which may seriously deteriorate the fault diagnosis performance of a data-driven method. In addition to the between-class imbalance caused by limited fault samples, the monitored signals from a complex system are usually heterogeneous, i. e. including both numeric and categorical variables. To deal with imbalanced heterogeneous data, a modified GAN fault diagnosis method is proposed in this work. In this method, to make the generated fault samples more in line with the actual situation, different types of variables are generated in different ways, including floating-point, integer, categorical, and hierarchical. Furthermore, the dual discriminator is designed to avoid the generated samples to exacerbate class overlapping, thus improving the effectiveness for imbalanced data. The fault data generated by M-D2GAN can be transformed to the heterogeneous variables as in the raw data, and the fault samples are enriched to improve the fault diagnosis performance. In comparison with classic GAN in a real case study concerning a high-speed train braking system, M-D2GAN has been shown to generate higher-quality fault samples, thus improving the prediction accuracy of the data-driven fault diagnosis model. In this paper, almost all the models used to verify the generated data are ready-made models. In fact, the design of fault diagnosis model often adopts different activation functions for different types of data before feature fusion. Therefore, it is necessary to keep the data types of simulation samples consistent with those of real samples.

Finally, for other PHM analyses of high-reliability complex systems, M-D2GAN has also potential value in fault prediction and regression. Wasserstein GAN (WGAN) can also be used to further optimize the generating method.

## CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

## References


[1] E. Zio, "Some challenges and opportunities in reliability engineering," *IEEE Trans. Reliab.*, pp. 1–14, 2016. DOI: 10.1109/TR.2016.2591504.

[2] E. Zio, "Prognostics and health management (PHM): where are we and where do we (need to) go in theory and practice," *Reliab. Eng. Syst. Saf.*, vol. 218, p. 108119, 2021. DOI: 10.1016/j.ress.2021.108119.

[3] Y. Hu, X. Miao, Y. Si, E. Pan, and E. Zio, "Prognostics and health management: a review from the perspectives of design, development and decision," *Reliab. Eng. Syst. Saf.*, vol. 217, p. 108063, 2021. 10.1016/j.ress.2021.108063.

[4] A. Shokry, P. Baraldi, A. Castellano, L. Serio, and E. Zio, "Identification of critical components in the complex technical infrastructure of the large Hadron Collider using relief feature ranking and support vector machines," *Energies*, vol. 14, 2021. DOI: 10.3390/en14186000.

[5] K. Tidriri, N. Chatti, S. Verron, et al., "Bridging data-driven and model-based approaches for process fault diagnosis and health monitoring: a review of researches and future challenges-ScienceDirect," *Annu. Rev. Control*, vol. 42, pp. 63–81, 2016.

[6] S. Ma, W. Cai, et al., "A lighted deep convolutional neural network based fault diagnosis of rotating machinery," *Sensors (Basel, Switzerland)*, vol. 19, no. 10, pp. 2381, 2019.

[7] H. Malik and S. Mishra, "Artificial neural network and empirical mode decomposition based imbalance fault diagnosis of wind turbine using TurbSim, FAST and Simulink," *IET Renew. Power Gener.*, vol. 11, no. 6, pp. 889–902, 2017.

[8] H. Hu, T. Bo, X. Gong, et al., "Intelligent fault diagnosis of the high-speed train with big data based on deep neural networks," *IEEE Trans. Ind. Inf.*, vol. 13, no. 4, pp. 2106–2116, 2017.

[9] H. Hassani, R. Razavi-Far, M. Saif, and E. Zio, "Deep learning with long short-term memory networks for diagnosing faults in smart grids," *Eur. J. Secur. Res.*, pp. 1–19, 2021. DOI: 10.1007/s41125-021-00074-4.

[10] B. Krawczyk, "Learning from imbalanced data: open challenges and future directions," *Prog. Artif. Intell.*, vol. 5, no. 4, pp. 221–232, 2016.

[11] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, October 2009.

[12] P. Zhou, X. Wang, Y. Liu, F. Liu, and J. Zhao, "Condition monitoring and fault diagnosis of motor bearings using undersampled vibration signals from a wireless sensor network," *J. Sound Vib.*, p. 414, 2017. DOI: 10.1016/j.jsv.2017.11.007.

[13] Y. Zhang, X. Li, L. Gao, et al., "Imbalanced data fault diagnosis of rotating machinery using synthetic oversampling and feature learning," *J. Manuf. Syst.*, vol. 48, no. Part C, pp. 34–50, 2018.

[14] J. Zhang and W. W. Y. Ng, "Stochastic sensitivity measure-based noise filtering and oversampling method for imbalanced classification problems," in *2018 IEEE Int. Conf. Syst., Man, and Cybernetics (SMC)*, IEEE, Miyazaki, Japan, pp. 403–408, 2018. DOI: 10.1109/SMC.2018.00078.

[15] G. U. Ping and O. Y. Yuan-You, "Classification research for unbalanced data based on heterogeneous-sampling," *Appl. Res. Comput.*, vol. 32, no. 2, pp. 379–381, 2015.

[16] K.-H. Yang and G.-L Shan, and L.-L. Zhao, "Steam turbine fault diagnosis based on least squares support vector machine," vol. 22, pp. 778–782, 2007.

[17] H. Liu, C. Lu, W. Hou, and S.-P. Wang, "An adaptive threshold based on support vector machine for fault diagnosis," pp. 907–911, 2009. DOI: 10.1109/ICRMS.2009.5269966.

[18] J. Liu and E. Zio, "Integration of feature vector selection and support vector machine for classification of imbalanced data," *Appl. Soft Comput.*, vol. 75, pp. 702–711, 2018.

[19] P. Karimi, "Comparing the fault diagnosis performances of single neural networks and two ensemble neural networks based on the boosting methods," *J. Chem. Eng. China*, vol. 2, no. 1, pp. 21–32, 2014.

[20] I. Goodfellow, J. Pouget-Abadie, M. Mirza, et al., "Generative adversarial nets," in: *Advances in Neural Information Processing Systems*, Z. Ghahramani, M. Welling, C. Cortes, et al. Eds. Red Hook, NY: Curran Associates, Inc., 2014, pp. 2672–2680.

[21] C. Ledig, L. Theis, F. Huszar, et al., Photo-realistic single image super-resolution using a generative adversarial network," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017*, Honolulu, HI, USA. pp. 105–114, 2016. DOI: 10.1109/CVPR.2017.19.

[22] S. Sharafi, B. Majidi, and A. Movaghar, "Low altitude aerial scene synthesis using generative adversarial networks for autonomous natural resource management," in *IEEE Conf. Knowl. Based Eng. Innov., Department of Computer Engineering, Khatam University*, Tehran, Iran; Department of Computer Engineering, Sharif University of Technology, pp. 322–326, 2019. DOI: 10.1109/KBEI.2019.8734904.

[23] M. Sharif, S. Bhagavatula, L. Bauer, et al., "Adversarial generative nets: neural network attacks on state-of-the-art face recognition," arXiv:1801.00349, 2017.

[24] Y. Ding, L. Ma, J. Ma, et al., "A generative adversarial network-based intelligent fault diagnosis method for rotating machinery under small sample size conditions," *IEEE Access*, vol. 7, pp. 149736–149749, 2019. DOI: 10.1109/ACCESS.2019.2947194.

[25] W. Mao, Y. Liu, L. Ding, et al., "Imbalanced fault diagnosis of rolling bearing based on generative adversarial network: a comparative study," *IEEE Access*, vol. 7, pp. 9515–9530, 2019. DOI: 10.1109/ACCESS.2018.2890693.

[26] D. Zhao, F. Liu, and H. Meng, "Bearing fault diagnosis based on the switchable normalization SSGAN with 1-D representation of vibration signals as input," *Sensors*, vol. 19, no. 9, pp. 2000, 2019. DOI: 10.3390/s19092000.

[27] D. Cabrera, F. Sancho, J. Long, et al., "Generative adversarial networks selection approach for extremely imbalanced fault diagnosis of reciprocating machinery," *IEEE Access*, vol. 99, p. 1, 2019.

[28] T. Zheng, L. Song, J. Wang, et al., "Data synthesis using dual discriminator conditional generative adversarial networks for imbalanced fault diagnosis of rolling bearings," *Measurement*, vol. 158, 107741, 2020.

[29] Z. Li, T. Zheng, Y. Wang, et al., "A novel method for imbalanced fault diagnosis of rotating machinery based on generative adversarial networks," *IEEE Trans Instrum Meas.*, vol. 70, pp. 1–17, 2021, Art no. 3500417, DOI: 10.1109/TIM.2020.3009343.

[30] P. Liang, C. Deng, J. Wu, et al., "Intelligent fault diagnosis via semisupervised generative adversarial nets and wavelet transform," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 7, pp. 4659–4671, 2020.

[31] F. Zhou, S. Yang, H. Fujita, et al., "Deep learning fault diagnosis method based on global optimization GAN for unbalanced data," *Knowl.-Based Syst.*, vol. 187, no. Jan., pp. 104837.1–104837.19, 2020.

[32] K. Yan, A. Chong, and Y. Mo, "Generative adversarial network for fault detection diagnosis of chillers," *Build. Environ.*, vol. 172, p. 106698, 2020.

[33] Y. R. Wang, G. D. Sun, Q. Jin, "Imbalanced sample fault diagnosis of rotating machinery using conditional variational auto-encoder generative adversarial network," *Appl. Soft*, vol. 92, p. 106333, 2020.

[34] Q. Guo, Y. Li, Y. Song, et al., "Intelligent fault diagnosis method based on full 1-D convolutional generative adversarial network," *IEEE Trans. Ind.*, vol. 16, no. 3, pp. 2044–2053, 2020.

[35] J. Gui, Z. Sun, Y. Wen, D. Tao, and J. Ye, "A review on generative adversarial networks: algorithms, theory, and applications," in *IEEE Trans. Knowledge Data Eng.*, (Early Access) pp. 1–1. 2021. DOI: 10.1109/TKDE.2021.3130191.

[36] L. Yu, W. Zhang, J. Wang, et al., "SeqGAN: sequence generative adversarial nets with policy gradient," CoRR, vol. abs/1609.05473, 2016.

[37] M. Cassel and F. L. Kastensmidt, "Evaluating one-hot encoding finite state machines for SEU reliability in SRAM-based FPGAs," in *12th IEEE Inter. On-Line Testing Symp. (IOLTS 2006)*, 10–12 July 2006, Como, Italy. 2006. DOI: 10.1109/IOLTS.2006.32.

[38] M. J. Kusner and J. M. Hernández-Lobato, "GANS for sequences of discrete elements with the Gumbel-softmax distribution," *arXiv*, 2016.

[39] Y. Matsunaga, "Accelerating SAT-based Boolean matching for heterogeneous FPGAs using one-hot encoding and CE-GAR technique," in *20th Asia and South Pacific Design Automation Conference, ASP-DAC 2015*, pp. 255–260, 2015. DOI: 10.1109/ASPDAC.2015.7059014.

[40] I. A. Basheer and M N. Hajmeer, "Artificial neural networks: fundamentals, computing, design, and application," *J. Microbiol. Methods*, vol. 43, no. 1, pp. 3–31, 2000.

[41] D. Kingma and J. Ba, "Adam: a method for stochastic optimization," *Comput. Sci.*, arXiv:1412.6980 [cs.LG], 2014.

[42] S. Lee, G. Kim, and S. Kim, "Self-adaptive and dynamic clustering for online anomaly detection," *Expert Syst. Appl.*, vol. 38, no. 12, pp. 14891–14898, 2011.

[43] M. Nakano, A. Takahashi, and S. Takahashi, "Fuzzy logic-based Portfolio selection with particle filtering and anomaly detection," *Knowl.-Based Syst.*, vol. 131, pp. 113–124, 2017.

[44] N. Patil, S. Menon, D. Das, et al., "Anomaly detection of non punch through insulated gate bipolar transistors (IGBT) by robust covariance estimation techniques," in *Int. Conf. Reliab.*, Safety and Hazard - Risk-Based Technologies and Physics-of-Failure Methods (ICRESH) Mumbai, India, pp. 68–72, 2010. DOI: 10.1109/ICRESH.2010.5779635.

[45] Z. Mao, G. Tao, B. Jiang, and X.-G. Yan, "Adaptive compensation of traction system actuator failures for high-speed trains," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, pp. 2950–2963, 2017.

[46] M. A. Zuluaga, E. J. F. D. Leyton, M. Hernández Hoyos, et al., "Feature selection for SVM-based vascular anomaly detection," in Menze, B., Langs, G., Tu, Z., Criminisi, A. (eds), Medical Computer Vision. Recognition Techniques and Applications in Medical Imaging. MCV 2010. Lecture Notes in Computer Science, vol. 6533. Springer, Berlin, Heidelberg. 2018. DOI: 10.1007/978-3-642-18421-5_14.

[47] M. Du, F. Li, G. Zheng, and V. Srikumar " DeepLog: anomaly detection and diagnosis from system logs through deep learning," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS '17). Association for Computing Machinery*, New York, NY, USA, 1285–1298, 2017. DOI: 10.1145/3133956.3134015.

[48] J. Liu, Y.-F. Li, and E. Zio, "A SVM framework for fault detection of the braking system in a high speed train," *Mech. Syst. Signal Process.*, vol. 87, pp. 401–409, 2017.

[49] J. Liu and E. Zio, "A scalable fuzzy support vector machine for fault detection in transportation systems," *Exp. Syst. Appl.*, p. 102, 2018. DOI: 10.1016/j.eswa.2018.02.017.

[50] M. Nguyen, S. Purushotham, H. To, et al.. "m-TSNE: a framework for visualizing high-dimensional multivariate time series," in *Proceedings of the 2016 Workshop on Visual Analytics in Healthcare, 2016*, Chicago, IL. pp. 22–29, 2017.