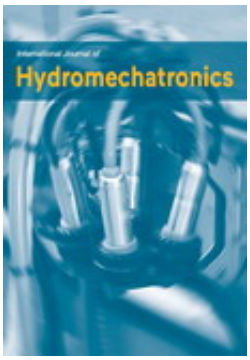


International Journal of Hydromechatronics

ISSN online: 2515-0472 - ISSN print: 2515-0464
https://www.inderscience.com/ijhm

An improved gated convolutional neural network for rolling bearing fault diagnosis with imbalanced data

Changsheng Xi, Jie Yang, Xiaoxia Liang, Rahizar Bin Ramli, Shaoning Tian, Guojin Feng, Dong Zhen

DOI: 10.1504/IJHM.2023.10055520

Article History:

Received:	30 August 2022
Last revised:	13 October 2022
Accepted:	08 November 2022
Published online:	25 April 2023

An improved gated convolutional neural network for rolling bearing fault diagnosis with imbalanced data

Changsheng Xi, Jie Yang and Xiaoxia Liang*

School of Mechanical Engineering,
Hebei University of Technology,
Tianjin 300401, China
Email: changsheng_xi@163.com
Email: 2010002@hebut.edu.cn
Email: xiaoxia.liang1@outlook.com
*Corresponding author

Rahizar Bin Ramli

Department of Mechanical Engineering,
Faculty of Engineering,
University of Malaya,
Kuala Lumpur 50603, Malaysia
Email: rahizar@um.edu.my

Shaoning Tian, Guojin Feng and Dong Zhen

School of Mechanical Engineering,
Hebei University of Technology,
Tianjin 300401, China
Email: shaoning_tian@163.com
Email: Guojin.Feng@outlook.com
Email: d.zhen@hebut.edu.cn

Abstract: To improve the ability of the deep learning model to handle imbalanced data, a fault diagnosis method based on improved gated convolutional neural network (IGCNN) is proposed. Firstly, an improved gated convolution layer is proposed for feature extraction, with the batch normalisation (BN) layer applied to adjust the data distribution and enhance the generalisation performance of the model. Then, the feature learned by multiple gated convolution layers and pooling layers is fed to the fully connected layer for fault type identification. Finally, the label-distribution-aware margin (LDAM) loss function is employed to adjust the model being more sensitive to the minority class and mitigate the influence of imbalanced data on the model. Experimental validation is conducted using two bearing datasets. Results show that the proposed method is more robust than other fault diagnosis methods, with higher recognition accuracy in severely imbalanced dataset.

Keywords: rolling bearings; fault diagnosis; imbalanced data; IGCNN; label-distribution-aware margin loss.

Reference to this paper should be made as follows: Xi, C., Yang, J., Liang, X., Bin Ramli, R., Tian, S., Feng, G. and Zhen, D. (2023) 'An improved gated convolutional neural network for rolling bearing fault diagnosis with imbalanced data', *Int. J. Hydromechatronics*, Vol. 6, No. 2, pp.108–132.

Biographical notes: Changsheng Xi received his BS degree in the Shandong University of Technology, Zibo, China, in 2020. He is currently working toward his Master's degree in the Hebei University of Technology, Tianjin, China. His research interests include mechanical fault diagnosis and deep learning.

Jie Yang received her MSc and PhD from the Hebei University of Technology, Tianjin, China, in 2003 and 2010, respectively. She is currently with the Hebei University of Technology, Tianjin, China. Her research interests include mechanical manufacturing and automation and intelligent robot technology.

Xiaoxia Liang received her MSc in Mechanical Design and Manufacturing and Automation and Mechanical Design and Theory from the Shandong University of Science and Technology, China, in 2015, and PhD in General Engineering in London South Bank University, UK, in 2021. She was the Project Leader in the Asset Life Management Team in The Welding Institute (TWI) Ltd. She is currently a post-Doctor in the Hebei University of Science and Technology. Her research interests include machinery fault detection, data analysis and risk-based assessment using machine/deep learning techniques.

Rahizar Bin Ramli received his PhD from the University of Leeds, UK, specialising in vehicle dynamics, semi-active control, and durability analysis for vehicle suspension systems. His current research interests include experimental and computational mechanics focusing on noise and vibration, vehicle dynamics, structural integrity, condition monitoring and engineering optimisation.

Shaoning Tian received his BS in Vehicle Engineering from the Henan University of Science and Technology, Luoyang, China, in 2018. He is currently working toward his PhD in Mechanical Engineering with the Hebei University of Technology, Tianjin, China. His research interests include mechanical system fault diagnosis and signal processing.

Guojin Feng received his BS and MS degrees from the Shandong University of Science and Technology, Qingdao, China, in 2009 and 2012, respectively, and PhD in Mechanical Engineering from University of Huddersfield, Huddersfield, UK, in 2016. His research interests include intelligent wireless condition monitoring system for industrial rotating machines and on-rotor micro-electro-mechanical systems (MEMS) sensing.

Dong Zhen received his BSc and MSc degrees from the Shandong University of Science and Technology, Qingdao, China, in 2006 and 2009, respectively, and PhD from University of Huddersfield, Huddersfield, UK, in 2012. He is currently with the Hebei University of Technology, Tianjin, China. His research interests include advanced signal processing and rotating machine fault diagnosis.

1 Introduction

As one critical component in rotating machinery, rolling bearing is prone to failure due to their harsh working environment (Hoang and Kang, 2019; Glowacz et al., 2018). Failed to detect fault in rolling bearing can cause damages to other components in the system, hence leading to breakdown of the system and more economic losses. Therefore, fault detection in rolling bearings has been extensively investigated. Nowadays, the field of machinery health monitoring has entered the big data era (Lei et al., 2016), and the deep learning-based fault diagnosis has achieved fruitful results in the field of condition monitoring and fault diagnosis (Hao et al., 2020; Sun et al., 2022; Guo et al., 2022). However, the good performance of these deep learning techniques is mainly based on relatively balanced datasets, without sufficient consideration of the influence of imbalanced data on the models. In practice, mechanical equipment mostly works under normal state with much less fault data, therefore, this causes imbalanced data samples, with a large amount of healthy data and limited types of fault data, which is the so-called long-tail distribution (Jia et al., 2018). This characteristic of the imbalanced data can have a severe influence on model performance. When the training data is imbalanced, the majority-class samples will be trained sufficiently, making the model more sensitive to them. Accordingly, the classification margin of the minority class will be narrowed and easily regarded by the model as the noise of the majority class. Moreover, the minority class is highly susceptible to overfitting, resulting in decreased generalisation capability of the model (Zhang et al., 2020). Therefore, this paper focuses on improving the model's capability to handle imbalanced data.

Generally, there are two types of methods to deal with imbalanced data, which are the data-based method (Li et al., 2022b) and the algorithm-based method (Xu et al., 2021). The data-based method is to expand the sample by increasing sampling or data generation, and thus converting the imbalanced problem into a balanced problem (Li et al., 2022b). For instance, Fan et al. (2019) expanded samples by using synthetic minority oversampling technique (SMOTE) and input them into the support vector machine (SVM) for fault diagnosis. Results of the experiments demonstrated that the method improved the model's performance for diagnosis on imbalanced datasets. Zhao et al. (2021) improved the diagnostic performance of the model under imbalanced data by using generative adversarial networks. Dixit and Verma (2020) presented a modified conditional variational auto-encoder (CVAE) for generating training samples, the generated samples had a high degree of similarity to the original samples. Although the data-based method alleviates the effects of imbalanced data, this method has three problems:

1 the sample information is not increased, which may lead to model overfitting (Li et al., 2022c)

2 the authenticity of the generated samples is questionable (Radford et al., 2015), which may affect the accuracy and precision of the machine learning model

3 with samples expanding, the computational effort of the model will be increased.

In contrast, the algorithm-based method increases the weights of minority classes by re-weighting to make the model more sensitive to minority class samples (Xu et al., 2021). Jia et al. (2018) proposed the deep normalised convolutional neural network (DNCNN) and weighted softmax loss to enable the model to be trained effectively on

imbalanced datasets. Zhang et al. (2019) improved the deep belief network (DBN) performance for imbalanced data with cost-sensitive learning. Dong et al. (2020) assigned different misclassification costs to each class using the cost adaptive loss function, which effectively solved the data imbalance problem. Note that, these models can be greatly affected by the values of the model parameters, and therefore, they require relevant expertise for parameter setting.

The aforementioned methods of dealing with imbalanced data have improved the diagnostic performance of models to a certain extent. However, their efficiency becomes limited when the data quantity is small and the degree of imbalance is severe. If the model can dig more information in the minority class samples, the sensitivity of the model to the minority class will improve. Therefore, the key to dealing with the imbalanced data problem is to enhance the feature extraction ability of the model. Some researches were found on improving the feature extraction ability of the diagnostic models, and among them, gated convolutional neural network (GCNN) has gained attention due to its powerful performance. Dauphin et al. (2017) introduced the gating mechanism into convolutional neural network (CNN) and firstly proposed the GCNN. It facilitated the propagation of the feature and alleviated the gradient disappearance. With the gated convolutional layer, the network can limit the flow of information between the layers, which makes the useless information filtered and the concentration of the model improved. Zhang et al. (2022) implemented the group-gating module in the CNN to improve the performance of the network. Guo et al. (2022) proposed a novel gated convolutional residual unit to solve the difficult problem of identifying the initial position of translation. Li et al. (2022a) proposed a method that combines gated convolution and pyramidal loss to improve the learning ability of the model and the image edge restoration. Despite its powerful feature extraction capability, GCNN is still influenced by imbalanced data. To further enhance the performance of GCNN in imbalanced datasets, we employ the batch normalisation (BN) technique to improve the model's generalisation capability and introduce the label-distribution-aware margin (LDAM) loss function to improve the model's ability to handle unbalanced data.

In this paper, a fault diagnosis method based on an improved gated convolutional neural network (IGCNN) for rolling bearing with imbalanced data is proposed. Firstly, the gated convolutional layer is improved by adding a BN layer to enhance the feature extraction and generalisation capability of the model. Secondly, the LDAM loss function is employed to reduce the difficulty in recognising a minority class and the influence of imbalanced data on the model performance. The effectiveness of the proposed method is verified with the Case Western Reserve University (CWRU) bearing data and an experimental cylindrical roller bearing data. The superiority of the proposed method is illustrated by comparison experiments.

The remainder of this paper is as follows: in Section 2, the relevant theoretical background is introduced. The methodology, model structure and fault diagnosis process of the proposed method are explained in Section 3. Experimental validation is carried out in Section 4. Finally, Section 5 draws the conclusions.

2 Theoretical background

2.1 Convolutional neural network

As one of the most representational algorithms of deep learning, CNN is a feedforward neural network that contains convolutional operations. The CNN has strong feature extraction capability and is widely used in mechanical fault diagnosis. It mainly composed of the convolutional layer, pooling layer and the fully connected layer (Chen et al., 2020).

Considering the bearing vibration signal is one-dimensional time series, the one-dimensional convolutional layer is applied in this paper. Its specific mathematical equation is given by:

$$y_{i,m}(n) = w_{i,m} x_{i-1}(n) + b_{i,m} \tag{1}$$

where $w_{i,m}$ and $b_{i,m}$ respectively represent the weight matrix and bias matrix of the m^{th} convolution kernel at the i^{th} convolution layer, $x_{i-1}(n)$ is the output value of the n^{th} channel in the $(i-1)^{\text{th}}$ convolution layer, and $y_{i,m}(n)$ represents the output value of the m^{th} convolution kernel in the n^{th} channel of the i^{th} layer.

After the convolutional layer is the pooling layer, which is used to compress the dimensionality of the features extracted by the convolutional layer. In this paper, the max-pooling layer is being taken, and its formula is as follows:

$$Y_{i,m}(n) = \max\left\{0,\, y_{i,m}(n)\right\} \tag{2}$$

where $y_{i,m}(n)$ is the output value of the data after the m^{th} convolution kernel in the n^{th} channel of the i^{th} convolution layer. $Y_{i,m}(n)$ denotes the output value after the maximum pooling operation.

After multiple layers of convolution and pooling operations, the obtained feature is input to the fully connected layer to realise the mapping of the feature vector to the sample label space. The output of the i^{th} fully connected layer can be expressed as:

$$y_i = \left(w_i\right)^T x_{i-1} + b_i \tag{3}$$

where x_{i-1} represents the $(i-1)^{\text{th}}$ layer output value, w_i is the weight matrix in the i^{th} layer, and b_i indicates the bias matrix in the i^{th} layer.

For classification tasks, the model commonly uses softmax classifier to calculate the probability that the sample belongs to which class to achieve classification. The softmax classifier can be described by:

$$P_i = soft\max\left(y_i\right) = \frac{e^{y_i}}{\sum_n e^{y_i}} \tag{4}$$

where P_i denotes the probability that the predicted outcome of the network belongs to the i^{th} class, n represents the total number of classes in the classification task, and y_i denotes the predicted value of the network.

2.2 Batch normalisation

With the increasing of the deep learning network depth, the number of its parameters increases significantly. During the model training phase, the parameters of each layer change as the convolutional operations. The feature distributions of parameters of each layer are significantly different. It increases model complexity and results in difficulties in convergence. At present, the BN technique has been gradually introduced into the domain of deep learning to solve the abovementioned problems. The BN can readjust the distribution of the input data to a standard normal distribution by normalisation, which reduces covariance and distribution differences between batches, accelerates the model's convergence speed and alleviates the overfitting phenomenon, enhancing the model's robustness and generalisation ability (Ioffe and Szegedy, 2015).

The BN layer performs the normalisation operation on the data, which can be described by:

$$\hat{x}_i = \frac{x_i - E[x_i]}{\sqrt{Var[x_i]}} \tag{5}$$

where $E[x_i]$ and $\sqrt{Var[x_i]}$ denote the mean and standard deviation of the input data of the i^{th} neuron, respectively.

To prevent the network's performance from decreasing after the transformation, two adaptive modulation parameters γ_i and β_i are added to each neuron, and the transformation equation can be written as:

$$y_i = \gamma_i \hat{x}_i + \beta_i. \tag{6}$$

2.3 Dropout

Due to the considerable complexity of the deep learning model, various numbers of parameters are required to be determined and fitted during the training process. Therefore, if the training samples are insufficient, the model will inevitably suffer from overfitting, which can affect the model's performance significantly. To effectively alleviate the overfitting problem, dropout is introduced in the proposed model. Dropout randomly selects some neurons to deactivate during the forward propagation of the neural network, ensuring that the model does not rely too heavily on specific features over the training process, ultimately alleviating the overfitting phenomenon (Srivastava et al., 2014).

3 IGCNN-based imbalanced fault diagnosis method

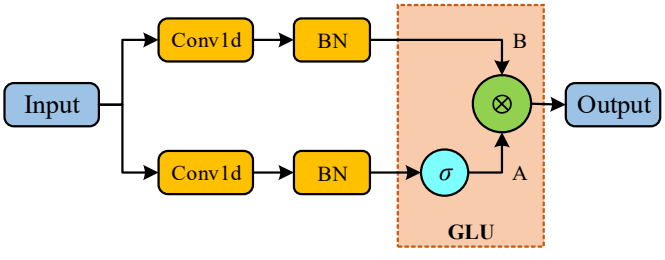
In this section, we give a description of the methodology and structural components of the proposed method, and then introduce the fault diagnosis procedure based on IGCNN.

3.1 Improved gated convolutional layer

The gating mechanism has a wide range of applications in deep learning models. Through the gating mechanism, the network can limit the flow of information between levels, select valuable information and filter out useless information, so that the data information is fully utilised. Dauphin et al. (2017) proposed the gated linear unit (GLU), which was combined with CNN to construct the gated convolutional network. For time series, GCNN can concentrate on valuable information (Dauphin et al., 2017). The GLU is the core of gated convolution, which is used to add a gating switch on the convolution layer to determine the probability of features that are passed to the next layer. In comparison with the long short-term memory (LSTM) network, GCNN enables parallel computation and reduces the number of nonlinear computations, thus alleviating the problem of gradient vanishing and accelerating the convergence of the model.

In order to handle imbalanced data better, an improved gated convolutional layer (denoted as IGCL) is proposed. It consists of three parts, namely the convolutional layer, the BN layer and the GLU, which are shown in Figure 1. The one-dimensional convolutional layer is used to process the one-dimensional vibration signal of the bearing. The BN layer is connected behind the convolutional layer to adjust the data distribution and reduce the difference caused by the imbalance between the number of samples in each class. It enables the minority class of samples to converge quickly, which improves the generalisation ability of the network. The GLU boosts the use of features by the model and also plays the role of the activation function.

Figure 1 The IGCL (see online version for colours)



After the data input to the IGCL, it first goes through the convolutional layer for feature extraction, then the data distribution is adjusted by the BN layer. Finally, the output result is obtained by GLU processing. The result of the operation contains two parts. One called the gate value, denoted as A, is calculated by the results obtained from the convolution and BN layer, multiplied by the sigmoid function. The other is the convolution value, which is the outcome of the convolution and BN process and is denoted as B. The gate value represents the weight corresponding to the feature. The stronger the feature is, the larger the gate value is. It is multiplied with the convolution value to achieve the function of information filtering. The final output is obtained by multiplying A and B, using the following equation:

$$h(X) = BN(X * W + b) \otimes \sigma\left[BN(X * V + c)\right] \tag{7}$$

where $*$ is the convolution operation, W and V are the convolution kernels, b and c represent the bias terms, $BN(\cdot)$ denotes the batch normalisation transform, σ represents

the sigmoid activation function, and ⊗ denotes the matrix corresponding to the element multiplication.

3.2 LDAM loss

For the fault diagnosis method based on deep learning, the cross-entropy (CE) loss is one of the most commonly utilised loss functions. However, the CE loss cannot weigh each class when the training set is an unbalanced dataset. This results in the loss value leaning toward the majority class samples and the feature margin of the minority class becoming smaller, which then affects the performance of the model. Cao et al. (2019) proposed a LDAM loss that can expand the classification margin of the minority class and thus make it easier for the model to recognise the minority-class. It is simple and easy to implement, and its parameters can be automatically adjusted by the sample size of each class, which decreases the difficulty of parameter selection and improves the performance of the model for classification in unbalanced datasets. LDAM loss has a wide range of applications in imbalance classification. Yang et al. (2022) have enhanced the ability of the model to learn the minority class using LDAM loss, which further boosts the performance of the network. Wang et al. (2021) mitigated the problem of label ambiguity in the test phase by re-weighting the top labels through label distribution learning. Zhu et al. (2021) proposed a distribution-aware local metric that more efficiently made full use of the limitation of training samples.

The essence of LDAM loss is to give a larger classification margin to the minority class, so that its actual margin is shifted towards the majority-class, improving the model's sensitivity to the minority-class. The LDAM loss principle is described below.

In model f, for a sample (x, y), whose true label $y = j$, the margin can be defined as:

$$\gamma(x, y) = f(x)_y - \max_{j \neq y} f(x)_j \tag{8}$$

Then, for all samples in the dataset with $y = j$ (denoted as S_j), each class has a margin:

$$\gamma_j = f(x)_y - \min_{i \in S_j} \gamma(x_i, y_i) \tag{9}$$

Therefore, the optimal margin can be derived and expressed in the following equation:

$$\gamma_j = \frac{C}{n_j^{1/4}} \tag{10}$$

where n_j represents the number of class j and C denotes the hyperparameter.

Based on the hinge loss (Wang et al., 2018), the final loss can be given by:

$$L_{LDAM}((x, y); f) = -\log \frac{e^{z_y - \Delta_y}}{e^{z_y - \Delta_y} + \sum_{j \neq y} e^{z_j}} \tag{11}$$

where $\Delta_j = \frac{C}{n_j^{1/4}}$, $j \in \{1, \cdots, k\}$.

3.3 The framework of IGCNN

To improve the feature extraction capability of the model in imbalanced datasets, an IGCNN is proposed. The framework of IGCNN is illustrated in Figure 2.

Figure 2 Framework of the proposed IGCNN (see online version for colours)

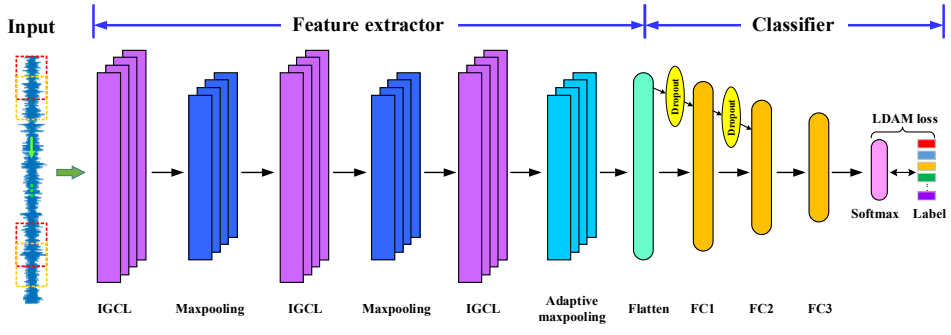


Table 1 The parameters of IGCNN

No.	*Layer type*	*Kernel size*	*Number of kernels*	*Stride*	*Other parameters*	*Output size*
0	Input	\	\	\	\	1 × 1,200
1	IGCL	16	16	1	\	16 × 1,185
2	Maxpooling	2	\	2	\	16 × 592
3	IGCL	3	32	1	\	32 × 590
4	Maxpooling	2	\	2	\	32 × 295
5	IGCL	3	64	1	\	64 × 293
6	Adaptive maxpooling	\	\	\	\	64 × 4
7	Dropout	\	\	\	Dropout rate = 0.5	\
8	FC	\	128	\	\	128
9	Dropout	\	\	\	Dropout rate = 0.5	\
10	FC	\	64	\	\	64
11	FC	\	10	\	\	10
12	Softmax	\	\	\	\	10

This model takes the end-to-end processing approach, where the raw bearing vibration signal is applied as input. The IGCNN has two parts: the feature extractor and the classifier. The feature extractor consists of three IGCLs and pooling layers alternately. The IGCL is used to achieve deep feature extraction and filtering. GLU implements the nonlinear transformation and effectively avoids gradient vanishing. Then, the feature is dimensionalised through the pooling layer. The last pooling layer is selected as the adaptive max-pooling layer to adaptively control the output dimension to meet the dimensionality requirements of the network. After the layer-by-layer feature extraction, the obtained feature vector is flattened to one-dimensional vectors and fed into the classifier. The classifier consists of three fully connected layers and the softmax classification layer. Furthermore, because of the large number of neurons in the

one-dimensional feature vector after flattening and the first fully connected layer, redundant parameters can be generated for the minority classes of samples, causing overfitting problems. Therefore, dropout layers are added after each of them. The parameters of IGCNN are listed in Table 1. In addition, as the LDAM loss can alleviate the effect of imbalanced data on the model, it is taken as the loss function of the model.

3.4 The fault diagnosis procedure based on IGCNN

The IGCNN-based fault diagnosis procedure mainly contains the following steps:

Step 1 Signal collection and preprocessing: The vibration signals of rolling bearings under different health states are collected by sensors and data collection devices. The samples are segmented using sliding windows and then split into training sets and a test set.

Step 2 Initialise the model: Initialise the model and set hyperparameters.

Step 3 Model training: Import the training data into the model for training, after the maximum number of iterations is reached, stop training and save model parameters for testing.

Step 4 Model validation: The test set is used to test the trained model to evaluate the performance of the model.

4 Experimental validation and analysis

To verify the effectiveness of the proposed method in imbalanced data fault diagnosis, two experimentally rolling bearing datasets are applied in this paper, which are the CWRU bearing dataset and the cylindrical roller bearing dataset. The experimental environment is Python3.7, Pytorch1.9 and cuda10.2. The computer is equipped with Intel(R) Core(TM) i7-10700F CPU @ 2.90 GHz and NVIDIA GeForce GT 1030.

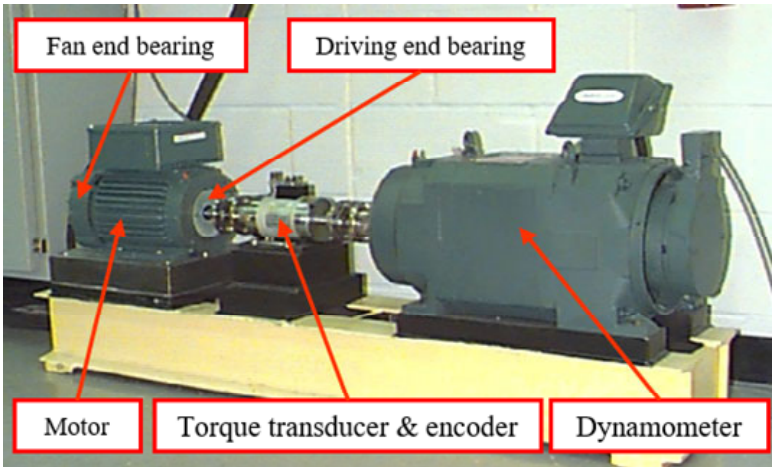
4.1 Experimental verification on the CWRU dataset

The CWRU dataset from the Case Western Reserve University Bearing Data Center (http://csegroups.case.edu/bearing-datacenter/home) is an open access dataset, the fault experimental bench of this dataset is illustrated in Figure 3. This experimental bench includes an AC motor, a dynamometer, a torque transducer, an encoder and rolling bearings. The rolling bearings were deep groove ball bearing (type SKF6205-2RSJEM), and they were respectively located on the drive end and the fan end of the motor. The experiments used electrical discharge machining (EDM) technique to setup three degrees of single point defects on the inner ring, outer ring and roller of the bearing with the fault diameters of 0.18 mm, 0.36 mm and 0.54 mm, respectively. The vibration signal of the rolling bearing was measured by the acceleration sensor, which was installed on the AC motor casing.

In this paper, the drive end bearing data are used for the experiment. This data contains normal data and three types of fault data. Each fault data includes three kinds of fault degree; consequently, there are ten health states in total. Among them, the normal

samples are labelled as 0 and the faulty samples are labelled as 1–9 in order. The experiment used a sliding window to segment the samples. Considering the influence of the sample length and the number of samples on the model, the length of the sliding window and sliding step are set to 1,200 and 400, respectively, to segment the samples and construct the dataset.

Figure 3 CWRU bearing fault experimental bench (see online version for colours)



Source: http://csegroups.case.edu/bearing-datacenter/home

Table 2 The information of the CWRU imbalanced datasets

Dataset		*Sample size*			*Imbalance rate R*
		Normal	*Each type of fault*	*Total*	
Training set	A	400	200	2,200	2:1
	B	400	100	1,300	4:1
	C	400	40	760	10:1
	D	400	20	580	20:1
Test set	E	150	150	1,500	1:1

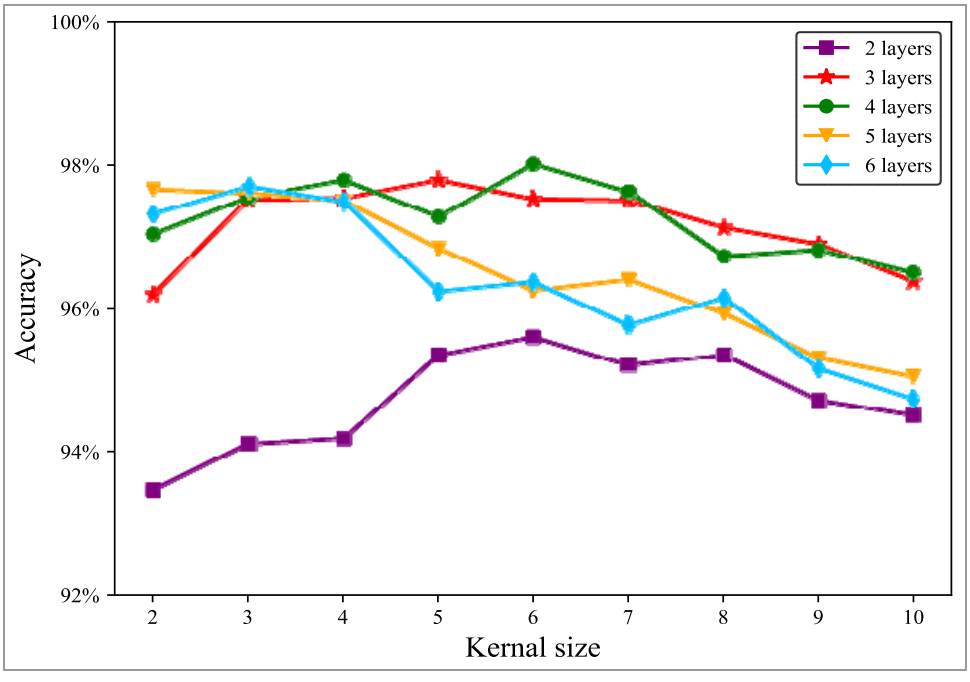
In order to demonstrate the proposed method's performance, the imbalance rate is designed to generate data samples. According to the actual operation of the machine, most of the time, the machine is working under normal condition; therefore, in the collected data, there are many normal samples and limited fault samples, and the probability of each class of faults is almost the same. Based on this description, the imbalance ratio is defined as:

$$R = N_n : N_f \tag{12}$$

where N_n represents the number of samples under the normal condition, N_f denotes the number of samples of each type of fault, and the number of all kinds of fault samples is equal in the experiment. The imbalance ratio R is separately set to 2:1, 4:1, 10:1 and 20:1. Four imbalanced training sets are constructed and respectively denoted as training sets A, B, C and D. Moreover, each dataset has 400 normal samples and the number of various

fault samples is respectively calculated according to the imbalance rate. The test set is set as a balanced dataset with 150 samples of each health state, which is denoted as test set E. The number of samples of each dataset as shown in Table 2. The model is separately trained on the above four imbalanced training sets and then tested on the test set E to evaluate the performance of the model under different levels of imbalanced data. It is worth noticing that the training and testing of the model are performed independently, and there is no crossover between each dataset.

Figure 4 Test accuracy of IGCNN with different network depth and convolutional kernel size (see online version for colours)



4.1.1 Optimisation of the model structure

To design the optimised structure of the feature extractor, the effects of network depth and convolutional kernel size for model performance are considered in this paper. Increasing the depth of the network can enhance the performance of the model, but too many layers can be back-productive and prone to over-fitting. Therefore, the number of IGCLs is increased from 2 to 6 layers. For the convolutional kernel size, according to Zhang et al. (2017), the large convolutional kernel in the first layer can extract the short-time features of the signal and remove useless features. Therefore, refer to this paper, the convolutional kernel in the first layer is set as a wide kernel with a size of 16. The other layers are set as small kernel with the size increased from 2 to 10, which is beneficial to increase the depth of the network. The experiments are trained in training set C and tested in test set E. Before the model training, the model is firstly initialised, and the hyperparameters of the model are set. We set the number of iterations to 100 and the batch size to 64. The Adam optimiser is selected to optimise the model parameters with

an initial learning rate of 0.001. The experiment is repeated ten times, and the results are illustrated in Figure 4.

From Figure 4, it can be seen that the increase in network depth and convolutional kernel size has no significant improvement on the accuracy, and the increase has brought more computational effort instead. In addition, the accuracy decreases after five layers, which is due to the overfitting of the model when the network depth is too large and the amount of data is limited. The oversized convolutional kernel makes the model filter out important information during feature extraction, which causes a decrease in accuracy. Therefore, to balance the diagnostic accuracy and computation of the model, the number of layers of the IGCLs is set to 3 and the convolutional kernel size is set to 3, which constitute the basic structure of the model.

4.1.2 Validation of the loss function

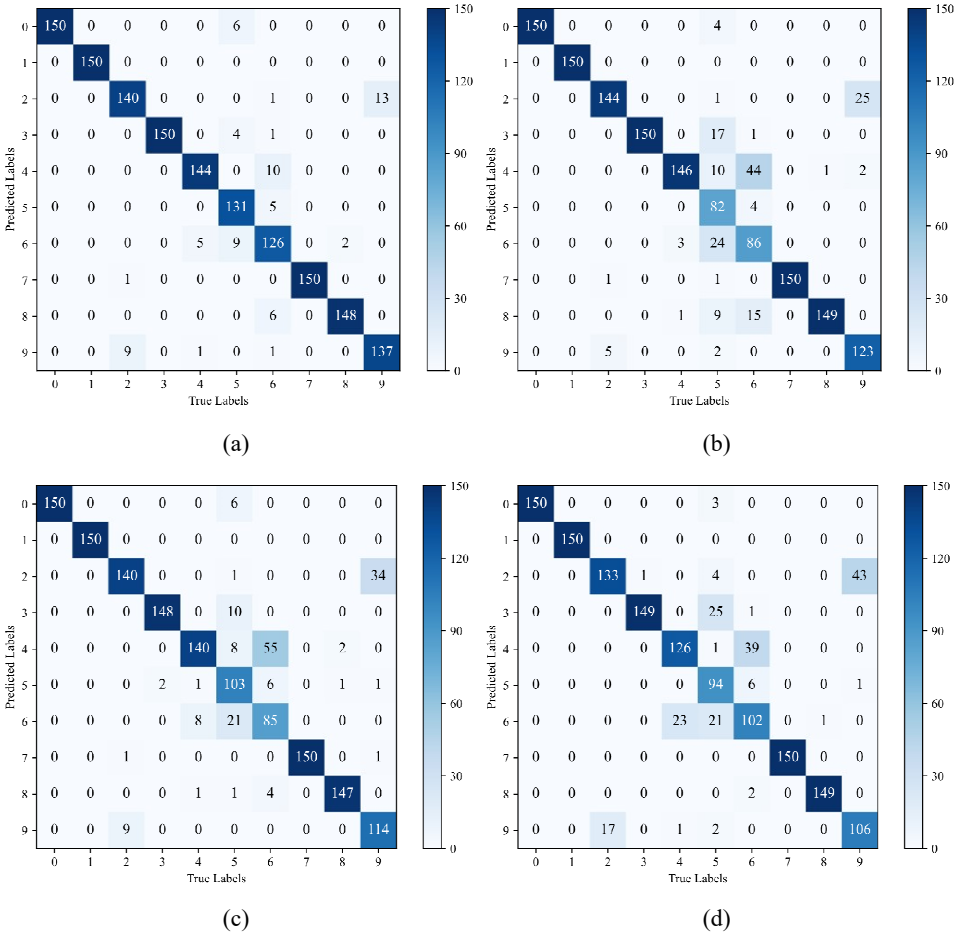
The LDAM loss function can expand the classification margin of the minority class, making it easier for the model to identify the minority class. In this paper, the LDAM loss function is compared with the commonly used CE loss, focal loss (Romdhane et al., 2020), and class-balance (CB) loss (Cui et al., 2019) function to demonstrate its advantage in dealing with data imbalance problems. In the experiment, the loss function of IGCNN is set to the above four functions in turn, while the other hyperparameters remain constant. Both are trained on different training sets, and experiments are repeated ten times and the average value as the final result. The test results are shown in Table 3. Among them, the confusion matrix of each loss function under the severely imbalanced test set E is shown in Figure 5.

Table 3 Test accuracy of each loss function under different training sets

Method	*Training set A*	*Training set B*	*Training set C*	*Training set D*
IGCNN-LDAM loss	99.87%	99.02%	97.53%	95.07%
IGCNN-CE loss	97.34%	94.71%	91.60%	88.67%
IGCNN-focal loss	98.26%	96.22%	93.27%	88.47%
IGCNN-CB loss	99.05%	96.84%	94.45%	87.27%

From the test results, all four loss functions can achieve high accuracy under the structural framework of IGCNN, but there are also differences in each training set, and the accuracy of LDAM loss is slightly higher than other losses. As the imbalance ratio increases, the accuracy of each loss function begins to decrease. LDAM loss is gradually better than other loss functions, and when the imbalance ratio between normal and faulty samples reaches 20:1, the accuracy of IGCNN-LDAM loss still reaches 95.07%, while all other losses have dropped to below 90%. As seen in the confusion matrix, the recognition accuracy of LDAM loss for minority classes (such as 4, 5, 6, 9) is significantly higher than the other losses. In summary, the LDAM loss is more effective and robust than the commonly used loss function in handling datasets with a high imbalance rate.

Figure 5 Confusion matrix of test results on the training set D for each loss function, (a) IGCNN-LDAM loss (b) IGCNN-CE loss (c) IGCNN-focal loss (d) IGCNN-CB loss (see online version for colours)

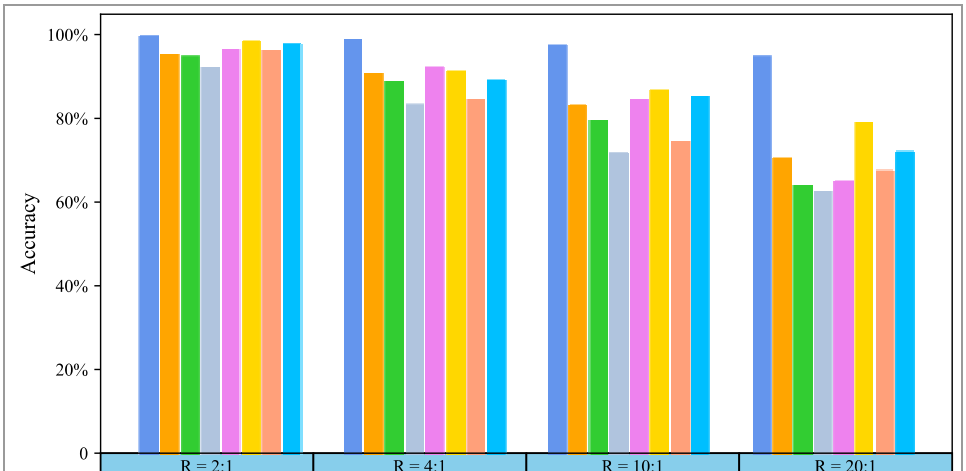


4.1.3 Comparison with other methods

To verify the superiority of the proposed method, comparison experiments are setup. We compare with the original GCNN, one-dimensional convolutional neural network (1D-CNN) and the bi-directional long short-term memory (Bi-LSTM) in Zhao et al. (2020), the classical CNN model AlexNet and some CNN architectures with strong feature extraction capability, such as MC-CNN (Huang et al., 2019), SECNN (Tang et al., 2021) and MDCNN (Fu et al., 2021). Among them, the GCNN uses the original gated convolutional layer and the other structures and parameters are the same as the proposed model to demonstrate the advantages of the IGCL compared to the original gated convolutional layer.

To guarantee that the comparison experiment is fair, the hyperparameters of each model are the same as Section 4.1.1. The models are trained in four datasets (A to D in Table 2) set by different imbalance rates, separately, and the parameters of the trained models are saved after the training process. Then applying these saved parameters, the model is tested using test set E. The average of the results of ten times of experiments result is taken for the final result. Figure 6 presents the test results for each model. As can be seen, the IGCNN model trained by the four imbalanced datasets can achieve high results in the test set. Compared with the other seven methods, our method has obvious superiority. As the increase in the degree of imbalanced data, IGCNN can still maintain a high recognition accuracy with strong robustness. When the data imbalance ratio is up to 20:1, it still achieves a recognition rate of more than 95%. However, each comparison method is greatly affected by imbalanced data, and the recognition accuracy is significantly reduced. For example, MC-CNN, although it has a strong feature extraction capability, it is prone to overfitting problems in the case of imbalanced and small sample size, resulting in lower accuracy. In summary, the experimental results demonstrate the effectiveness and superiority of the proposed method. In addition, the results of comparison with GCNN show that the proposed IGCL has a stronger performance than the original gated convolutional layer, which improves the feature extraction ability and generalisation ability of the model. After using the BN layer in the gated convolutional layer, the model's ability to handle imbalanced data is effectively improved.

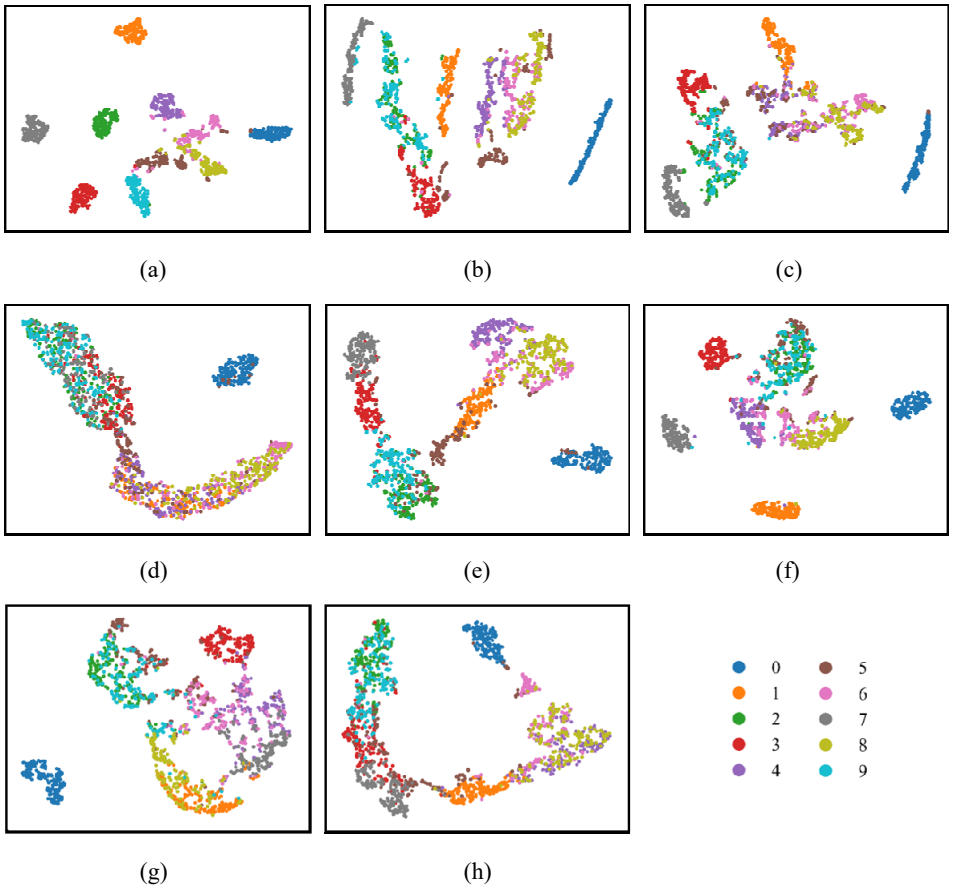
Figure 6 Test results of each model on test set E (see online version for colours)



	R = 2:1	R = 4:1	R = 10:1	R = 20:1
IGCNN	99.87%	99.02%	97.53%	95.07%
GCNN	95.53%	90.73%	83.20%	70.67%
1D-CNN	95.06%	89.07%	79.86%	64.25%
Bi-LSTM	92.33%	83.47%	72.04%	62.53%
AlexNet	96.73%	92.40%	84.67%	65.27%
MC-CNN	98.49%	91.44%	87.01%	79.22%
SECNN	96.53%	84.59%	74.73%	67.82%
MDCNN	97.91%	89.57%	85.28%	72.29%

To further demonstrate the superiority of the proposed method in feature extraction, the t-SNE technique is applied for visualising the fault classification results of IGCNN and each comparison method using a severe imbalanced dataset (training set D, with the imbalance ratio of 20:1), which is illustrated in Figure 7. As can be seen from Figure 7, the features extracted by the IGCNN achieve effective separation, while the other seven methods are more confusing, with various classes of features overlapping together. The results show that the IGCNN's diagnostic performance in the case of imbalanced data is significantly stronger than the other methods.

Figure 7 Visualisation of t-SNE features for each model, (a) IGCNN (b) GCNN (c) 1D-CNN (d) Bi-LSTM (e) AlexNet (f) MC-CNN (g) SECNN (h) MDCNN (see online version for colours)



With the training set D as an example, Figures 8 and 9 show the loss and accuracy change process of the proposed method and each model trained under severe imbalance, and Table 4 shows the time taken by each model to complete the training. It can be seen that the proposed method can converge quickly and has the lowest value of the loss function

and takes less time to complete the training than other methods, which indicates that the proposed method can eliminate the influence of the imbalance between the data and complete the convergence of the model quickly. For other models, they are more influenced by the imbalanced data, and it is difficult to complete the convergence of the model. For example, MC-CNN, MDCNN, which have high structural complexity, are highly susceptible to model overfitting with large imbalance rate and small sample size, and take longer time for training, resulting in poor performance of these methods. Therefore, the superiority of the proposed method is demonstrated by comparison.

Figure 8 The loss curves of each model trained in the training set D (see online version for colours)

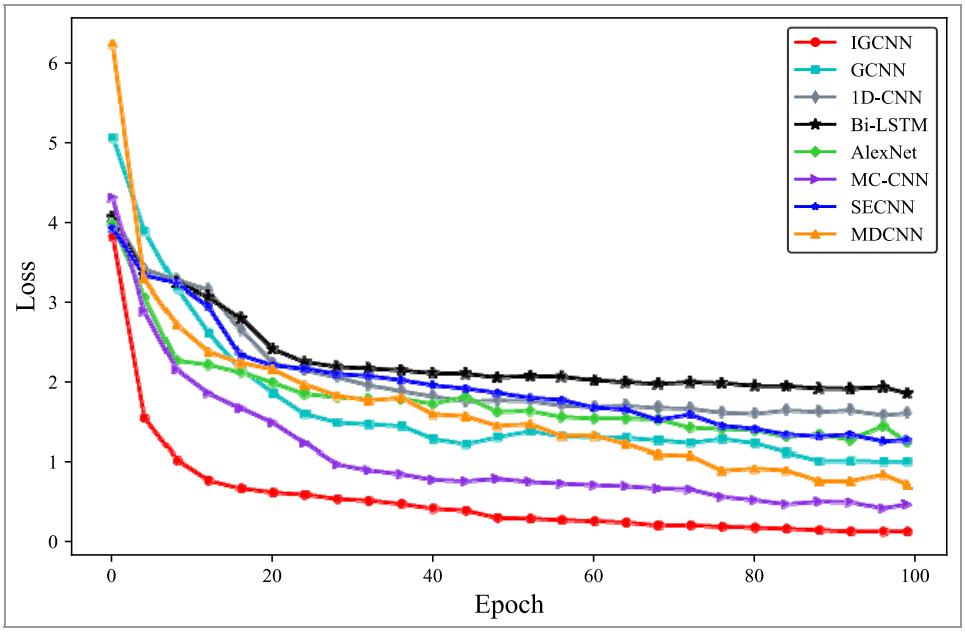
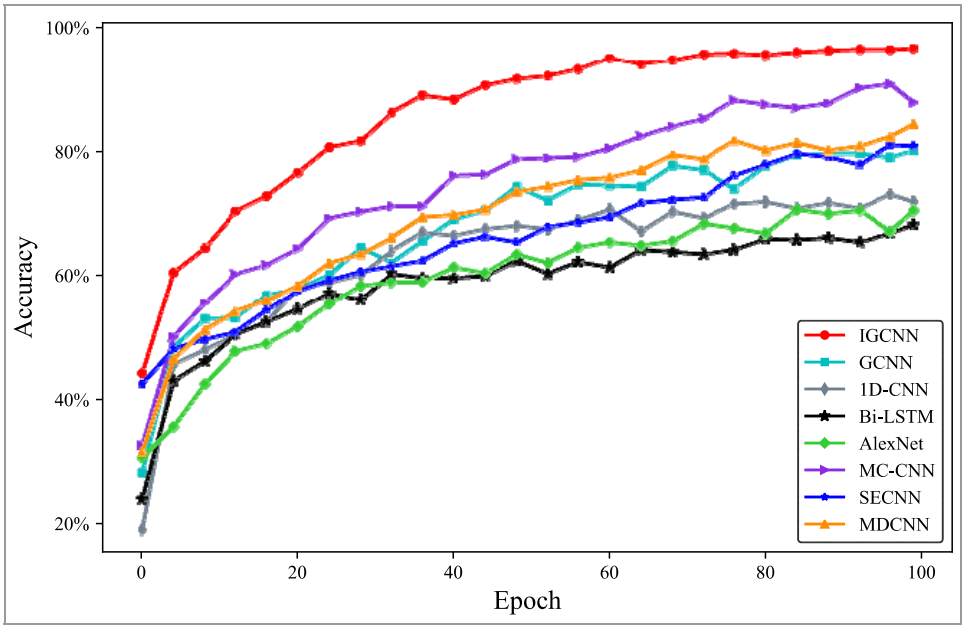


Table 4 Training time of each model with training set D

Model	*Time (s)*	*Model*	*Time (s)*
IGCNN	97.67	AlexNet	162.04
GCNN	115.42	MC-CNN	198.29
1D-CNN	141.66	SECNN	231.81
Bi-LSTM	134.53	MDCNN	214.83

Figure 9 The accuracy curves of each model trained in the training set D (see online version for colours)



4.2 Experimental verification of the cylindrical roller bearing dataset

4.2.1 Data description and experimental setup

To further verify the proposed fault diagnosis method, a cylindrical roller bearing fault test bench was setup in our lab. The structure of this cylindrical roller bearing fault test bench is shown in Figure 10, which consists of an electric generator, couplings, an AC motor, an intermediate shaft, bearing supports and cylindrical roller bearings. The model type of the cylindrical roller bearing was N406. In the bearing failure experiment, the healthy state of the cylindrical roller was divided into ten classes, including the normal status and nine types of fault. Different degrees of scratch faults on the inner ring, outer ring and roller were simulated by EDM of 1 mm in depth, 0.18 mm in width and 30%, 60% and 100% in length of the bearing components, respectively. The location and severity of each fault can be seen in Figure 11. The speed of the AC motor is 1,468 rpm and vibration signals from the bearings are collected by acceleration sensors installed on the upper surface of the bearing support with sampling frequency of 96 kHz and each signal is collected for 10 s.

Figure 10 The cylindrical roller bearing fault test bench (see online version for colours)

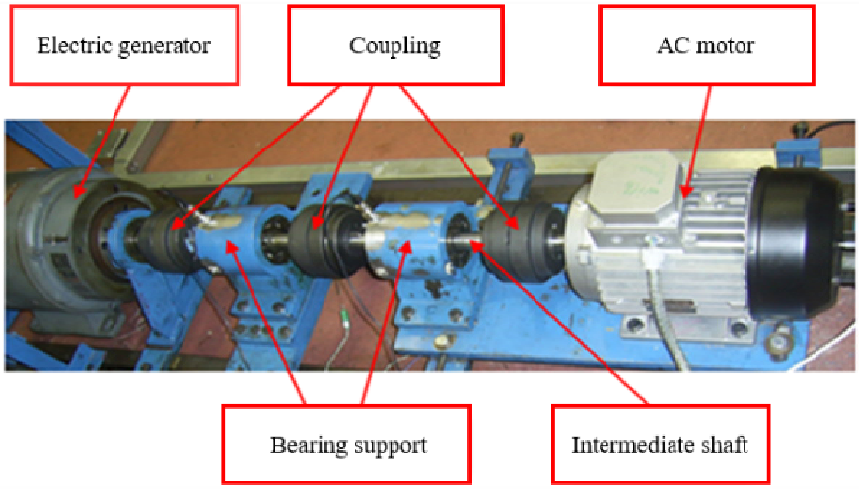
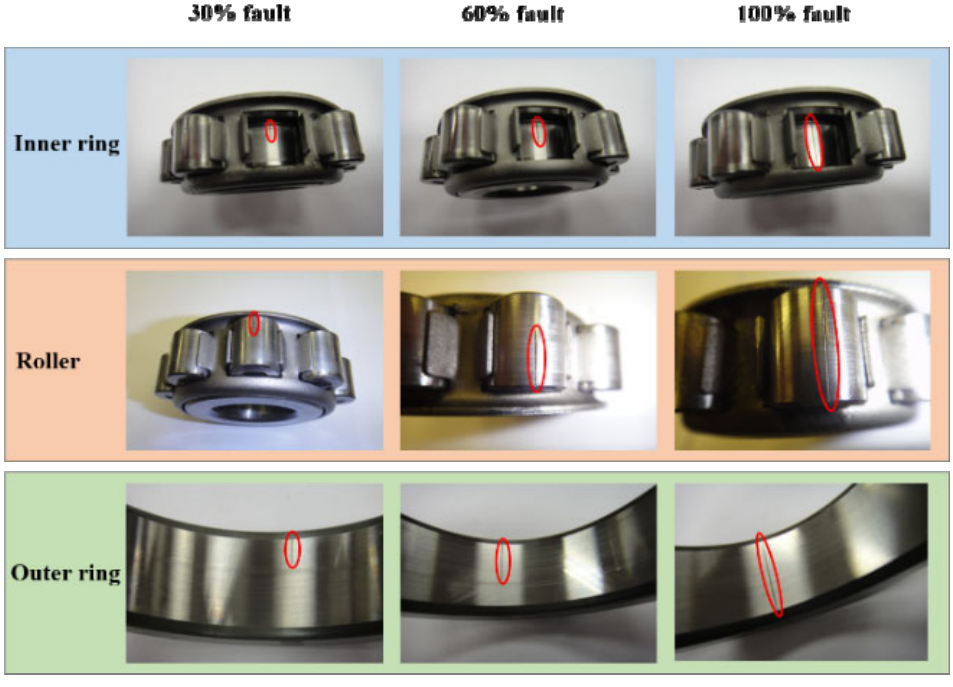


Figure 11 Fault types of cylindrical roller bearings (see online version for colours)



In this paper, we use the signal of cylindrical roller bearing for the experiment. To simulate the imbalanced data, the dataset is setup in the same way as those in Section 4.1, as shown in Table 5. The four imbalanced training sets are labelled as F, G, H, and I, and the test set is J. The model is trained in each of the four training sets and tested in the test set J to evaluate its performance.

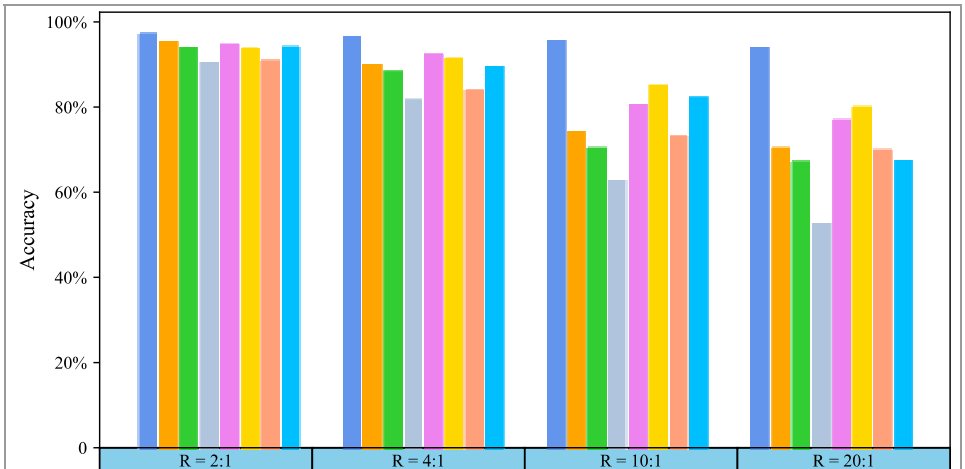
Table 5 Information of the roller bearing imbalanced dataset

Dataset		*Sample size*			*Imbalance rate R*
		Normal	*Each type of fault*	*Total*	
Training set	F	400	200	2,200	2:1
	G	400	100	1,300	4:1
	H	400	40	760	10:1
	I	400	20	580	20:1
Test set	J	150	150	1,500	1:1

4.2.2 Comparison with other methods

The procedure of the experiment is the same as Experiment 1. Four datasets F to I (Table 5) are applied as training data of the IGCNN and seven comparison models, and then tested using test set J. The final test results are shown in Figure 12. When increasing the imbalance rate, the IGCNN model maintains high accuracy and stability, while the accuracy of the four comparison models decreases. When the imbalance rate reaches a serious imbalance of 20:1, the classification accuracy of the IGCNN reaches 94.47%, while the accuracy of comparison methods decreases significantly. Meanwhile, IGCNN still has a better performance than GCNN, demonstrating the proposed method can solve the impact of imbalanced data on model accuracy and has better stability than other methods.

Figure 12 Test results of each model on test set J (see online version for colours)



	R = 2:1	R = 4:1	R = 10:1	R = 20:1
IGCNN	97.42%	96.62%	96.03%	94.47%
GCNN	95.71%	90.04%	74.45%	70.56%
1D-CNN	94.47%	88.68%	70.53%	67.53%
Bi-LSTM	90.73%	82.03%	62.93%	52.93%
AlexNet	95.11%	92.70%	80.67%	77.41%
MC-CNN	94.19%	91.68%	85.46%	80.17%
SECNN	91.14%	84.30%	73.41%	70.23%
MDCNN	94.52%	89.77%	82.42%	67.62%

Similarly, the fault features extracted by each model are observed using the t-SNE technique. The results can be found in Figure 13. It is clear from Figure 13 that the IGCNN has a better extraction effect, and the different fault types can be separated from each other and easily distinguished; while the extraction effect of other models is poor, and the aggregation of each class is low and difficult to distinguish. The comparison proves that IGCNN has stronger feature extraction ability than other models.

Table 6 Training time of each model with training set I

Model	*Time (s)*	*Model*	*Time (s)*
IGCNN	109.24	AlexNet	192.30
GCNN	124.39	MC-CNN	213.27
1D-CNN	187.61	SECNN	261.77
Bi-LSTM	162.84	MDCNN	204.25

Figure 13 Visualisation of t-SNE features for each model, (a) IGCNN (b) GCNN (c) 1D-CNN (d) Bi-LSTM (e) AlexNet (f) MC-CNN (g) SECNN (h) MDCNN (see online version for colours)

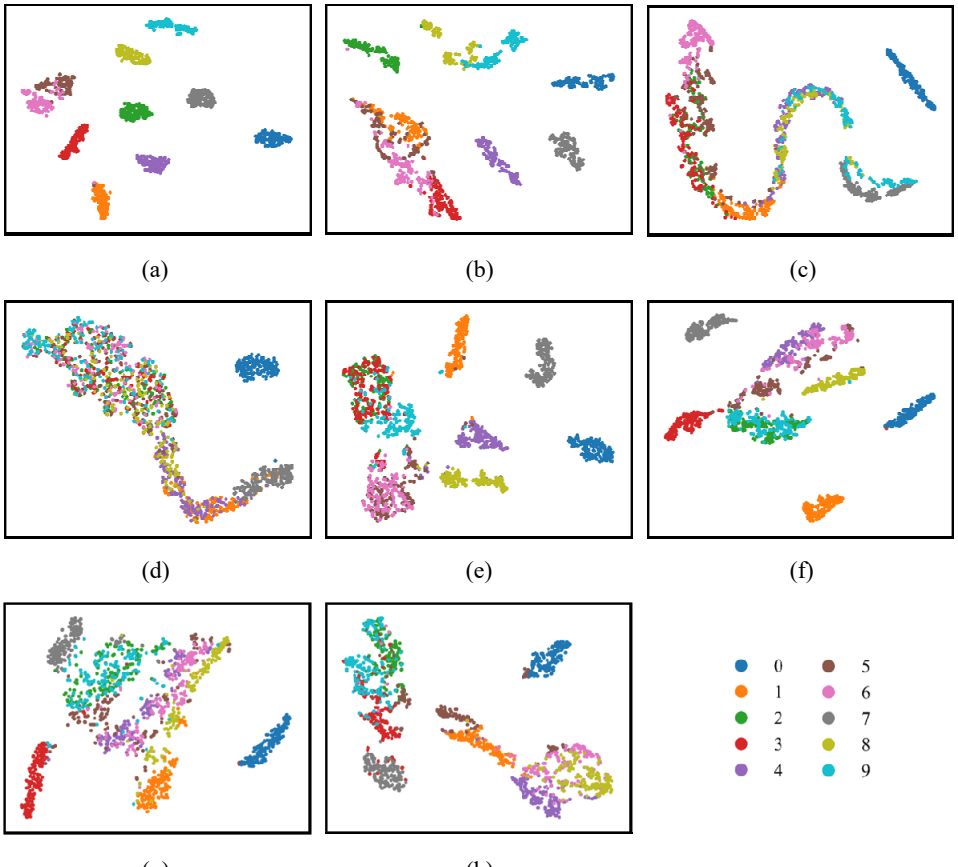


Figure 14 The loss curves of each model trained in the training set I (see online version for colours)

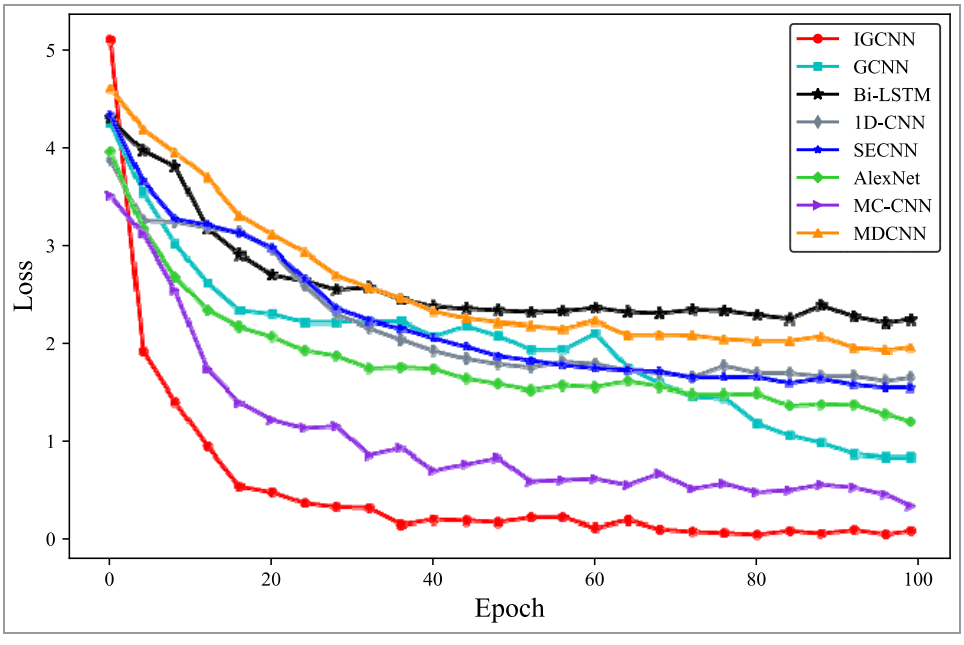
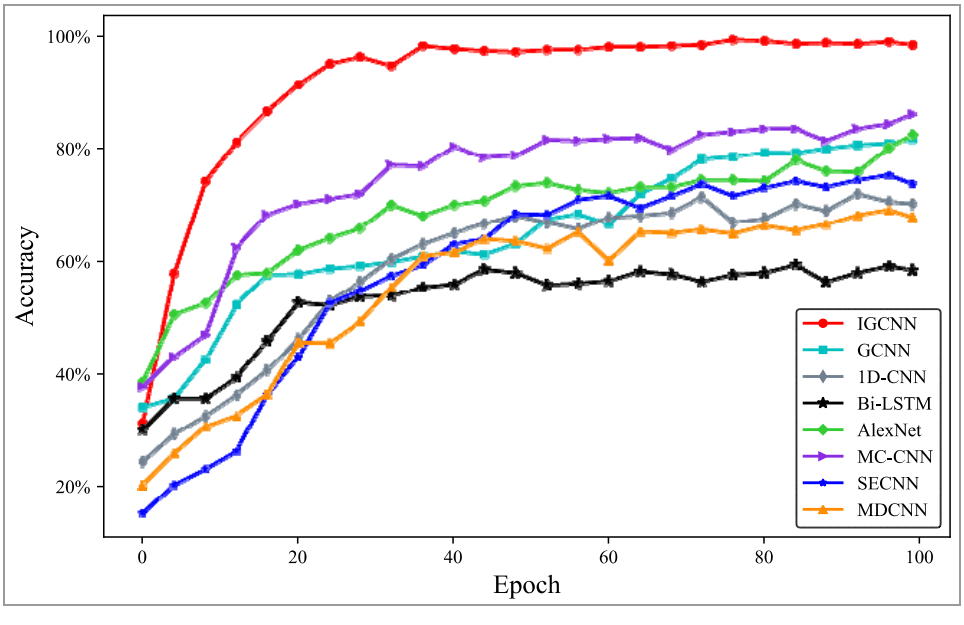


Figure 15 The accuracy curves of each model trained in the training set I (see online version for colours)



The variation and time taken for each model in the training set I with a large imbalance rate are shown in Figures 14 and 15 and Table 6. It can be obtained that the proposed method completes convergence after 30 iterations, still performs well in this dataset, and it is better than the compared methods. The results show that the proposed method has the advantage of stable training and convergence quickly.

5 Conclusions

This paper proposes a fault diagnosis method for imbalanced data based on an IGCNN, which has shown strong diagnostic performance and robustness in imbalanced datasets. The superiority of the proposed method is verified through experiments and comparisons. The conclusions based on the experimental results are listed as follows:

1 IGCNN has achieved high diagnostic accuracy in training sets with different degrees of imbalance. It can effectively overcome the imbalanced data problem and has strong robustness. Compared with other methods, IGCNN has obvious superiority.

2 The proposed IGCL in this paper improves the feature extraction and generalisation ability of the model. Its performance is stronger than the original gated convolutional layer. Comparison results show that the BN layer can adjust the data distribution and facilitate the model to handle imbalanced data.

3 The label distribution-aware margin loss function enhances the ability of the model to identify the minority class, improving the model's performance on the imbalanced dataset.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 52275101), the Natural Science Foundation of Tianjin (Grant No. 21JCZDJC00720), and the Natural Science Foundation of Hebei (Grant No. E2022202101). The authors would like to express their sincere thanks to the editor and anonymous referees for their valuable comments and suggestions.

References

Cao, K., Wei, C., Gaidon, A., Arechiga, N. and Ma, T. (2019) 'Learning imbalanced datasets with label-distribution-aware margin loss', *Advances in Neural Information Processing Systems (NeurIPS 2019)*, Vol. 32, pp.1–12.

Case Western Reserve University Bearing Data Center Website [online] http://csegroups.case.edu/bearing-datacenter/home (accessed 11 July 2021).

Chen, Z., Mauricio, A., Li, W. and Gryllias, K. (2020) 'A deep learning method for bearing fault diagnosis based on cyclic spectral coherence and convolutional neural networks', *Mechanical Systems and Signal Processing*, Vol. 140, pp.1–16.

Cui, Y., Jia, M., Lin, T.Y., Song, Y. and Belongie, S. (2019) 'Class-balanced loss based on effective number of samples', *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2019)*, pp.9268–9277.

Dauphin, Y.N., Fan, A., Auli, M. and Grangier, D. (2017) 'Language modeling with gated convolutional networks', *International Conference on Machine Learning (PMLR)*, pp.933–941.

Dixit, S. and Verma, N.K. (2020) 'Intelligent condition-based monitoring of rotary machines with few samples', *IEEE Sensors Journal*, Vol. 20, No. 23, pp.14337–14346.

Dong, X., Gao, H., Guo, L., Li, K. and Duan, A. (2020) 'Deep cost adaptive convolutional network: a classification method for imbalanced mechanical data', *IEEE Access*, Vol. 8, pp.71486–71496.

Fan, Y., Cui, X., Han, H. and Lu, H. (2019) 'Chiller fault diagnosis with field sensors using the technology of imbalanced data', *Applied Thermal Engineering*, Vol. 159, pp.1–12.

Fu, L., Zhang, L. and Tao, J. (2021) 'An improved deep convolutional neural network with multiscale convolution kernels for fault diagnosis of rolling bearing', *Materials Science and Engineering*, Vol. 1043, No. 5, pp.1–10.

Glowacz, A., Glowacz, W., Glowacz, Z. and Kozik, J. (2018) 'Early fault diagnosis of bearing and stator faults of the single-phase induction motor using acoustic signals', *Measurement*, Vol. 113, pp.1–9.

Guo, Y., Zhou, D., Cao, J., Nie, R., Ruan, X. and Liu, Y. (2022) 'Gated residual neural networks with self-normalization for translation initiation site recognition', *Knowledge-Based Systems*, Vol. 237, pp.1–12.

Hao, S., Ge, F., Li, Y. and Jiang, J. (2020) 'Multisensor bearing fault diagnosis based on one-dimensional convolutional long short-term memory networks', *Measurement*, Vol. 159, pp.1–8.

Hoang, D.T. and Kang, H.J. (2019) 'A survey on deep learning based bearing fault diagnosis', *Neurocomputing*, Vol. 335, pp.327–335.

Huang, W., Cheng, J., Yang, Y. and Guo, G. (2019) 'An improved deep convolutional neural network with multi.scale information for bearing fault diagnosis', *Neurocomputing*, Vol. 359, pp.77–92.

Ioffe, S. and Szegedy, C. (2015) 'Batch normalization: accelerating deep network training by reducing internal covariate shift', *International Conference on Machine Learning (PMLR)*, pp.448–456.

Jia, F., Lei, Y., Lu, N. and Xing, S. (2018) 'Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization', *Mechanical Systems and Signal Processing*, Vol. 110, pp.349–367.

Lei, Y., Jia, F., Lin, J., Xing, S. and Ding, S.X. (2016) 'An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data', *IEEE Transactions on Industrial Electronics*, Vol. 63, No. 5, pp.3137–3147.

Li, H., Wang, G., Gao, K. and Li, H. (2022a) 'A gated convolution and self-attention-based pyramid image inpainting network', *Journal of Circuits, Systems and Computers*, Vol. 31, No. 12, pp.1–18.

Li, J., Liu, Y. and Li, Q. (2022b) 'Generative adversarial network and transfer-learning-based fault detection for rotating machinery with imbalanced data condition', *Meas. Sci. Technol.*, Vol. 33, No. 4, pp.1–16.

Li, J., Liu, Y. and Li, Q. (2022c) 'Intelligent fault diagnosis of rolling bearings under imbalanced data conditions using attention-based deep learning method', *Measurement*, Vol. 189, pp.1–15.

Radford, A., Metz, L. and Chintala, S. (2015) *Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks* [online] https://arxiv.org/abs/1511.06434 (accessed 20 February 2022).

Romdhane, T.F., Alhichri, H., Ouni, R. and Atri, M. (2020) 'Electrocardiogram heartbeat classification based on a deep convolutional neural network and focal loss', *Computers in Biology and Medicine*, Vol. 123, pp.1–13.

Srivastava, N., Hinton, G. and Krizhevsky, A. (2014) 'Dropout: a simple way to prevent neural networks from overfitting', *The Journal of Machine Learning Research*, Vol. 15, No. 1, pp.1929–1958.

Sun, J., Gu, X., He, J., Yang, S., Tu, Y. and Wu, C. (2022) 'A robust approach of multi-sensor fusion for fault diagnosis using convolution neural network', *Journal of Dynamics Monitoring and Diagnostics*, Vol. 1, No. 2, pp.103–110.

Tang, H., Gao, S., Wang, L., Li, X., Li, B. and Pang, S. (2021) 'A novel intelligent fault diagnosis method for rolling bearings based on Wasserstein generative adversarial network and convolutional neural network under unbalanced dataset', *Sensors*, Vol. 21, No. 20, pp.1–24.

Wang, F., Cheng, J., Liu, W. and Liu, H. (2018) 'Additive margin softmax for face verification', *IEEE Signal Processing Letters*, Vol. 25, No. 7, pp.926–930.

Wang, J., Geng, X. and Xue, H. (2021) 'Re-weighting large margin label distribution learning for classification', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 44, No. 9, pp.5445–5459.

Xu, J., Li, Y., Meng, F., Zhang, D., Ye, Y. and Lu, L. (2021) 'Fault diagnosis on imbalanced data using an adaptive cost-sensitive multiscale attention network', *2021 International Conference on Intelligent Technology and Embedded Systems (ICITES)*, pp.77–82.

Yang, W., Hu, Z., Zhou, L. and Jin, Y. (2022) 'Protein secondary structure prediction using a lightweight convolutional network and label distribution aware margin loss', *Knowledge-Based Systems*, Vol. 237, pp.1–12.

Zhang, C., Tan, K.C., Li, H. and Hong, G.S. (2019) 'A cost-sensitive deep belief network for imbalanced classification', *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 30, No. 30, pp.109–122.

Zhang, W., Li, X., Jia, X., Ma, H., Luo, Z. and Li, X. (2020) 'Machinery fault diagnosis with imbalanced data using deep generative adversarial networks', *Measurement*, Vol. 152, pp.1–12.

Zhang, W., Peng, G., Li, C., Chen, Y. and Zhang, Z. (2017) 'A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals', *Sensors*, Vol. 17, No. 2, pp.1–21.

Zhang, X., Yang, F., Hu, Y., Tian, Z., Liu, W., Li, Y. and She, W. (2022) 'RANet: network intrusion detection with group-gating convolutional neural network', *Journal of Network and Computer Applications*, Vol. 198, pp.1–12.

Zhao, B. and Yuan, Q. (2021) 'Improved generative adversarial network for vibration-based fault diagnosis with imbalanced data', *Measurement*, Vol. 169, pp.1–11.

Zhao, Z., Li, T., Wu, J., Sun, C., Wang, S., Yan, R. and Chen, X. (2020) 'Deep learning algorithms for rotating machinery intelligent diagnosis: an open source benchmark study', *ISA Transactions*, Vol. 107, pp.224–255.

Zhu, W., Li, W., Liao, H. and Luo, J. (2021) 'Temperature network for few-shot learning with distribution-aware large-margin metric', *Pattern Recognition*, Vol. 112, pp.1–10.