IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received March 16, 2020, accepted March 29, 2020, date of publication April 7, 2020, date of current version April 29, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2986419

Deep Cost Adaptive Convolutional Network: A Classification Method for Imbalanced Mechanical Data

XUN DONG, HONGLI GAO, LIANG GUO, KESI LI, AND ANDONGZHE DUAN
Engineering Research Center of Advanced Driving Energy-Saving Technology, Ministry of Education, Southwest Jiaotong University, Chengdu 610031, China
School of Mechanical Engineering, Southwest Jiaotong University, Chengdu 610031, China

Corresponding authors: Hongli Gao (hongli_gao@swjtu.edu.cn) and Liang Guo (guoliang@swjtu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 51775452, Grant 51905452, in part by Fundamental Research Funds for Central Universities under Grant 2682017ZDPY09, Grant 2682019CX35, and in part by Planning Project of Science & Technology Department of Sichuan Province under Grant 2019YFG0353.

ABSTRACT Intelligent diagnosis is an important manner for mechanical fault diagnosis in the era of industrial big data, and deep network has received extensive attention in this field because of automatically learning features and classifying entered samples. As a classic deep learning model, Convolutional Neural Network has been applied in mechanical intelligent fault diagnosis. However, the limitation is that entered samples must be balanced to achieve satisfactory recognition rate. During the operation of machinery, the normal samples are abundant and the fault samples are rare. Therefore, the recognition rate of the minority category is minor when processing the imbalanced data with Convolutional Neural Network. To solve the above problem, an intelligent classification method for imbalanced mechanical data based on Deep Cost Adaptive Convolutional Network is proposed. According to this model, first, it learns intrinsic state characteristics in mechanical raw signals through multiple convolution and pooling operations. Second, it maps these characteristics to mechanical health condition by fully connected layers. Finally, the cost adaptive loss function adaptively assigns different misclassification costs for all categories and keeps updating them in training process to effectively classify the imbalanced mechanical data. The proposed method is verified by bearing data and milling cutter data with different imbalanced ratio, and compared with other methods. The experimental results show that the proposed method is robust and is able to effectively classify the imbalanced mechanical data.

INDEX TERMS Cost adaptive, convolutional network, imbalanced data, mechanical intelligent fault diagnosis.

I. INTRODUCTION

The structural complexity and functional coupling of machinery determine that any minor fault may trigger a chain reaction. So it is imperative to implement comprehensive and real time monitoring of it [1]–[3]. Mechanical intelligent fault diagnosis methods are used to extract the hidden fault characteristics from the monitoring signals [4], [5] and automatically identify the health condition of machinery through intelligent algorithm, which are current researched hotspot in the field of fault diagnosis. Since Hinton [6] first proposed the concept of ''deep learning'' in 2006, deep learning has become an emerging researched hotspot in academia and industry, and deep neural networks have also been successfully applied in different engineering fields, such as image recognition [7], text analysis [8], speech recognition [9], fault diagnosis [10]–[12] and remaining useful life prediction [13]–[15]. Jing *et al.* [10] proposed a fault diagnosis method based on convolutional neural network, which learns features from the frequency domain data of the original vibration signals. Guo *et al.* [11] proposed a deep convolution transfer learning network for fault recognition, and the domain adaptive module was used to learn the domain-invariant features to effectively classify unlabeled data. The above-mentioned deep learning methods that applied to the field of mechanical intelligent fault diagnosis have achieved great fault recognition results from balanced datasets, but none of the above research has involved the

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

imbalanced distribution of samples in different mechanical health condition. However, in industrial production, mechanical failures occur as the minority category, and require more attention than the majority category (machinery runs normally). If the minority category is misclassified as the majority category, the optimal maintenance time for machinery will be missed, and resulting in irreparable economic losses [16]. Therefore, it is of great significance to study how to accurately classify imbalanced mechanical data.

So far, the strategies for solving the imbalanced classification problem can be divided into two categories: data-level and algorithm-level [17], [18]. On the one hand, the data-level approaches are to increase or decrease the number of samples to make the imbalanced dataset more balanced. Fang and Li [19] reduced the error caused by single Random under-sampling through multiple Random under-sampling, effectively solving the problem of imbalanced data onto software defect detection. Random under-sampling is able to effectively reduce the training time of the model, but the dataset after under-sampling remains unchanged in training, which will result in that some useful information fail to participate in training and reduce the performance of the classifier. Based on the K-neighbor rule, Lin *et al.* [20] proposed two clustering under-sampling algorithms. Wu and Shen [21] extracted the support vectors that play a key role in classification according to the degree of class overlap, and proposed an under-sampling method. Despite these methods are able to effectively overcome the problem that Random under-sampling easily loses important samples' information, under-sampling methods may destroy the distribution information of the dataset through sample rejection. Huang *et al.* [22] divided the distance bands for the imbalanced datasets, and used the adaptive variable neighborhood SMOTE algorithm to generate samples for the samples within the distance bands, but the reasonable division of the distance bands has a greater impact on the classification accuracy of the datasets. Abdi and Hashemi [23] proposed an over-sampling method based on Mahalanobis distance, which synthesized samples only in the minority dense areas. This method can effectively overcome the problem of sample overlap, but cannot guarantee the boundary samples' information about the minority category. Although the over-sampling methods achieve a relatively balanced number of samples, the overall distribution of the data is not considered, so the data distribution of the new dataset after over-sampling cannot be guaranteed.

On the other hand, algorithm-level methods mainly include cost-sensitive learning methods and ensemble learning methods. The cost-sensitive learning methods adapt to the imbalanced dataset by directly modifying the existing classifiers. Jia *et al.* [24] proposed a mechanical intelligent fault diagnosis method based on deep normalized convolutional neural network, which provided an effective idea for solving the problem of imbalanced mechanical data classification. But the model's misclassification costs are preset according to the imbalanced ratio and are constant in training process. Dhar and Cherkassky [25] proposed a cost-sensitive support vector machine to assign different costs for different categories, so as to obtain the best classification results. The core of cost-sensitive learning methods is the setting of misclassified costs. Most current researches take imbalanced ratio as misclassification cost, which cannot reflect the true distribution characteristics of datasets, so the effect of cost-sensitive learning methods cannot be guaranteed. The ensemble learning methods integrate the classification results of multiple base classifiers in a certain way to improve classification performance. Chawla *et al.* [26] used SMOTE Boost to create the minority samples, indirectly changing the data scale and compensating for skewed distributions. Yan and Han [27] proposed a stack-integration method that performs cost-sensitive integration on the basis of sampling-integration, which can achieve better classification results than a single integration model. However, the training process of the ensemble algorithm is complex and takes a long time. At the same time, it is difficult to choose the type and the number of the base classifiers.

In view of the above problems, we propose an intelligent fault diagnosis method based on Deep Cost Adaptive Convolutional Network (DCACN) in this paper. First, the mechanical intrinsic state characteristics in raw data are learned by multiple convolution and pooling operations. Second, the fully connected layers map these features to the mechanical health condition. Finally, the cost adaptive loss function adaptively assigns different misclassification costs for all categories and keeps updating them in training process. It is expected that the recognition rate of minority samples will be improved while achieving a higher classification accuracy of overall samples. The difference between this method and the existing cost-sensitive learning methods is that it introduces two evaluation indexes in training process to adaptively set appropriate misclassification costs for different categories, thereby achieving effective classification of imbalanced datasets. The validity of DCACN is verified by the imbalanced milling cutter datasets and bearing datasets. Furthermore, by comparing with the conventional methods, the superiority of DCACN in classifying imbalanced mechanical data is verified.

II. PROPOSED METHOD

In this section, we describe the proposed DCACN in detail. As shown in Fig.1, the method includes health condition recognition module and cost adaptive module. The health condition recognition module is one-dimensional convolutional neural network, which learns the internal features from raw data of the monitored components. And the learned features are used to identify the health condition of the monitored components. The cost adaptive module is used to evaluate the model's parameters and modify the network to improve the recognition rate of imbalanced data. In cost adaptive module, v_p is the misclassification cost of minority category and v_n is that of majority category. BP means back propagation throughout the process.

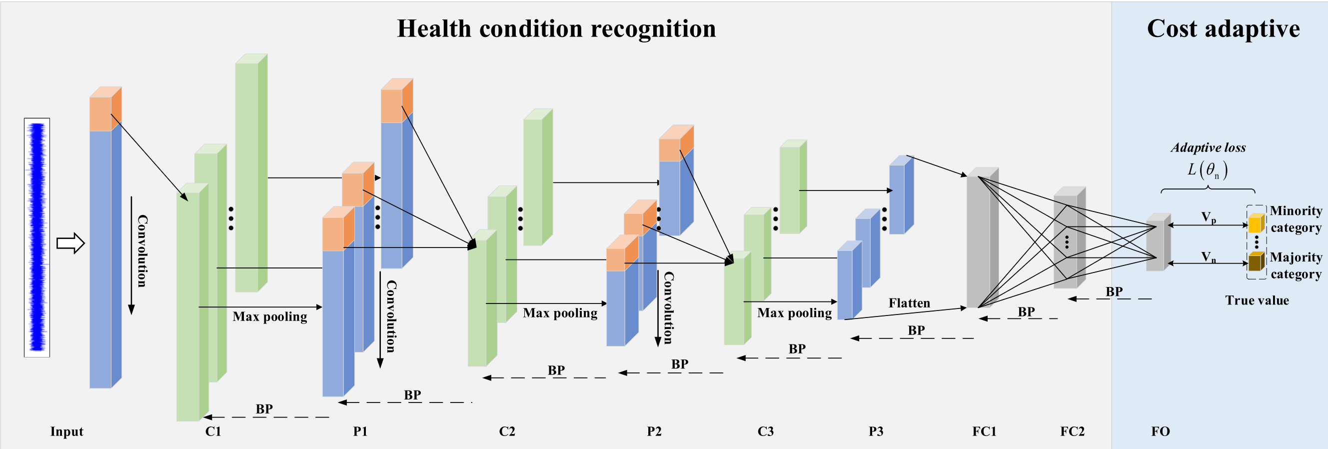


FIGURE 1. The structure illustration of the proposed method.

A. HEALTH CONDITION RECOGNITION

Because the input data are raw vibration signals, one-dimensional convolutional neural network is used in the health condition recognition module, which includes input layer, convolutional layers, pooling layers, fully connected layers, and output layer. The specific description of the network is shown in Table 1.

TABLE 1. The architecture of the DCACN.

Layer	Operator	Parameter size
Input	Input signals	1200*1
C1	Convolution	5*1*1
P1	Max pooling	4*1
C2	Convolution	5*1*20
P2	Max pooling	4*1
C3	Convolution	5*1*20
P3	Max pooling	4*1
FC1	Fully connected	1500*500
FC2	Fully connected	500*100
FO	Output	100*C

The input layer is built by raw vibration signals with a length of N. Next is the convolutional layer, in which the convolutional kernel is the most important part. Since the vibration signals are one-dimensional, one-dimensional convolution operation is used in this method. Let the input signals and the weight matrix of the convolution kernel be x and w_c, respectively. Then the convolution and activation operation can be expressed as follows,

$$X_j^l = Relu(\sum_{i\epsilon N_j} X_i^{l-1} * W_{ij}^l + b_j^l) \tag{1}$$

where X_i^{l-1} is the i-th feature of the $l-1$-th layer, W_{ij}^l is the convolutional kernel, b_j^l is the corresponding bias, $Relu(\cdot)$ is the activation function.

Generally, a pooling layer is used after the convolution layer, which can effectively compress the information, thereby reducing parameters and speeding up calculation. In this method, the max pooling operation is used. The pooling operation can be expressed as follows,

$$p_j^l = \max\left\{x_{j\times k:(j+1)\times k}^{l-1}\right\} \tag{2}$$

where p_j^l is the output value of the j-th neuron in the l-th layer of the pooling operation, max{} Is a pooling operation function that calculates the maximum value in the target area, $x_{j\times k:(j+1)\times k}^{l-1}$ is the j-th local area of the $l-1$-th layer, and the length of the neuron segment in this area is k.

After multiple convolution and pooling operations, the network learns the hidden features in the input signals, and then the fully connected layers map the learned features to the labeled space of the samples. We can calculate the output of each layer as follows,

$$x^l = f(\sum x^{l-1} * w_f^l + b_f^l) \tag{3}$$

where x^l is the output of the l-th layer, w_f^l is the weight matrix between the l-th layer and the previous layer, b_f^l is the corresponding bias, $f(\cdot)$ is the activation function, the health conditions of machine are predicted in the health condition output layer FO through the *softmax* regression.

B. COST ADAPTIVE

In section A, the specific classification process of one-dimensional CNN is introduced. In the previous classification methods, the distribution of samples with different health condition is balanced, so the model's misclassification costs are same in training process. However, when dealing with the imbalanced classification problem of mechanical fault data, this strategy of setting the same misclassification costs for different categories is difficult to meet the actual requirements. And the strategy is difficult to be applied in identifying the health condition with a small sample size. Therefore, we introduce a strategy of updating the loss function adaptively. During training process, a small misclassification cost is set for the majority category and a large misclassification cost is set for the minority category adaptively, so as to achieve accurate identification of the imbalanced mechanical data.

In the proposed method, the core part is how to adaptively update the misclassification costs in training process. Based on the imbalanced ratio of the samples, we consider the evaluation indexes after each mini-batch of training and combine the distance measurement results between the true and predict values of the training samples, so as to propose an updating method for the misclassification costs. We use the G_{mean} and Euclidean distance as parameters to achieve adaptively updating the misclassification costs. The misclassification costs of each type of samples can be expressed as follows,

$$v_c = \frac{max\{n_c\}_{c=1}^{C}}{n_c} \times exp(-\frac{G_{mean}}{2}) \times exp(-\frac{1}{2E_D}) \tag{4}$$

where, C is the number of neurons output by the last fully connected layer, that is, the total number of categories. c is the actual health condition, n_c is the number of samples in each c. G_{mean} is the geometric mean of each mini-batch, E_D is the Euclidean distance between the predict values and the actual values for each mini-batch. The two parameters are reversed and then exponentially transform to prevent gradient explosions when the imbalanced ratio is too large. The misclassification costs of each type of samples can be adaptively calculated through this formula.

n_c is expressed as follows,

$$n_c = \sum\nolimits_{i=1}^{Q} l(y_i = c) \tag{5}$$

where, y_i is the output of the model, Q is the number of samples in each mini-batch. If the equation in parentheses is true, $l(\cdot)$ is 1, otherwise 0.

In imbalanced classification problems, it is important to evaluate the classification model for the recognition performance. Referring to literature [28], G_{mean} combines the recognition accuracy of the majority and minority categories, and does not tend to a single category. In addition, it is often used as an effective index to evaluate model's performance. Therefore, it is used as an important parameter to adaptively update the misclassification costs, expressed as follows,

$$G_{mean} = \sqrt{Recall \times Specificity} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

$$Specificity = \frac{TN}{TN + FP} \tag{8}$$

where, TP is true positive, FN is false negative, TN is true negative, FP is false positive, *Recall* is the proportion of all positive samples that are correctly identified, *Specificity* is the proportion of negatives that are correctly identified. When dealing with multi-classification problems, it is treated as multiple binary classification problems.

In addition to G_{mean}, we also introduce Euclidean distance as an important indicator for updating v_c in this method. Euclidean distance measures the absolute distance between points in a multi-dimensional space. In this method, it is used to measure the distance between the true values and predict values of each category. The longer the distance, the greater the difference between the two sets. The formula as follows,

$$E_D = \frac{1}{P_c}\sqrt{\sum\nolimits_{i=1}^{P_c} \left(y_i - y_i'\right)^2} \tag{9}$$

where, y_i is the predictive value, y_i' is the real value, P_c is the number of samples corresponding to class c in each mini-batch, and the sum of P_c in a mini-batch is Q.

According to the above formulas, the cost adaptive loss function can be defined as,

$$L(\theta) = -\frac{1}{Q}\sum\nolimits_{c=1}^{C} v_c l(y_i = c) log(s_i) \tag{10}$$

$$s_i = \frac{e^{x_i}}{\sum_{k=1}^{C} e^{x_k}} \tag{11}$$

where, x_i is the i-th output value of the last fully connected layer, θ is the parameter in the networks, s_i is the softmax function.

Although the cross-entropy loss function minimizes the misclassification of overall samples, it treats the misclassification of different categories as equally. Therefore, we introduce the adaptive misclassification cost v_c, so that the minority category is valued in training process, and then the misclassification of the minority category is reduced.

After the loss function been established, in order to achieve the global optimization of the neural network, SGD is selected to optimize the parameter θ on each mini-batch. The parameter θ update as follows,

$$\theta \leftarrow \theta - \mu\frac{\partial L(\theta)}{\partial\theta} \tag{12}$$

$$x_k = \sum \theta \cdot x_{k-1} + b_k \tag{13}$$

$$\begin{aligned}
\frac{\partial L(\theta)}{\partial\theta} &= \frac{\partial L(\theta)}{\partial s_i}\frac{\partial s_i}{\partial x_k}\frac{\partial x_k}{\partial\theta} \\
&= \frac{\partial\left(-\frac{1}{Q}\sum_{c=1}^{C} v_c l\left(y_i = c\right) log\left(s_i\right)\right)}{\partial s_i}\frac{\partial s_i}{\partial x_k}\frac{\partial x_k}{\partial\theta} \\
&= -\frac{\sum_{c=1}^{C} l\left(y_i = c\right)}{Q}\left[\frac{\partial log\left(s_i\right)}{\partial s_i}v_c + \frac{\partial v_c}{\partial s_i}log\left(s_i\right)\right] \\
&\quad \times \frac{\partial s_i}{\partial x_k}\frac{\partial x_k}{\partial\theta} \\
&= -\frac{\sum_{c=1}^{C} l\left(y_i = c\right)}{Q}\left[\frac{v_c}{s_i} + \frac{v_c log\left(s_i\right)}{2E_D^2}\right]\frac{\partial s_i}{\partial x_k}\frac{\partial x_k}{\partial\theta} \\
&= \frac{v_c}{Q}\left[1 + \frac{s_i log\left(s_i\right)}{2E_D^2}\right]\left[s_i - l\left(y_i = c\right)\right]\frac{\partial x_k}{\partial\theta} \\
&= \frac{v_c x_{k-1}}{Q}\left[1 + \frac{s_i log\left(s_i\right)}{2E_D^2}\right]\left[s_i - l\left(y_i = c\right)\right]
\end{aligned} \tag{14}$$

where, μ is learning rate, Q is the number of samples in each mini-batch, x_k is the forward propagation function.

Once the network is built, DCACN learns the features from raw data, and identifies the health condition with learned features. Finally, the misclassification cost of each category is adjusted by the cost adaptive module, so as to achieve effective classification of imbalanced mechanical data.

III. CASE STUDY I: IMBALANCED FAULT DIAGNOSIS OF ROLLING BEARINGS

In this section, the imbalanced datasets constructed by the Paderborn University bearing data are used to verify the classification effect of the proposed method. In addition, in order to verify the superiority of the proposed method in fault diagnosis, the results are compared with that of several other methods.

A. PADERBORN BEARING DATASET

The vibration data used in this section is the public Paderborn bearing dataset, which is available at the Kat Data Center website of the Chair of Design and Drive Technology,

Paderborn University, Germany [29]. The dataset consists of artificial damage samples and generating real bearing damage samples by accelerated lifetime tests. A total of 32 bearings' samples including normal, outer ring fault, inner ring fault, and composite fault. These data are collected from 6203 ball bearings, with a sampling frequency of 64kHz under four operating conditions.

We select four bearings K001, KA15, KI04, KB23 to build imbalanced datasets for verification. The above bearings correspond to four health conditions: normal, outer ring fault, inner ring fault, and composite fault. The vibration waveforms of the four bearings are shown in Fig.2. B1, B2, B3 and B4 are datasets used to verify the proposed method constructed from the above-mentioned Paderborn bearing dataset. The number of samples for each health condition is shown in Table 2. The samples' number of dataset B1 is balanced. During the operation of the machine, the normal bearing data is easier to be obtained than the fault data, and the single fault data is easier to be obtained than the composite fault data. Therefore, we reduce the proportion of fault data in the imbalanced datasets, so as to simulate the actual imbalanced ratio. The ratio of N, OF, IF, and CF in the remaining three datasets are 10:1:1:1, 40:4:4:1, and 80:8:8:1, respectively.

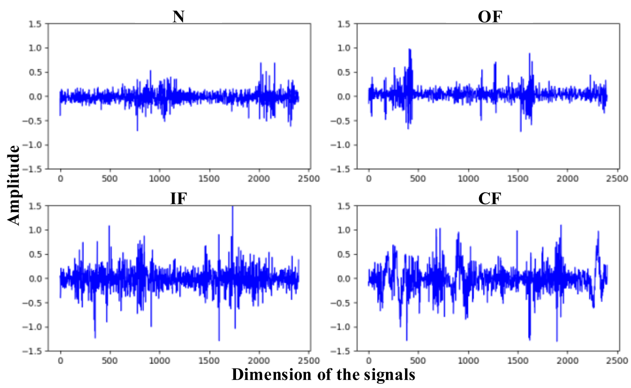


FIGURE 2. Vibration signals for each health condition of bearings.

TABLE 2. Imbalanced bearing datasets.

Datasets	N	OF	IF	CF
B1	4000	4000	4000	4000
B2	4000	400	400	400
B3	4000	400	400	100
B4	4000	400	400	50

B. EVALUATION METRICS

In order to accurately evaluate the classification effect of DCACN on imbalanced datasets, we introduce three evaluation metrics. The first metric is *accuracy*, it is a common metric for evaluating the performance of classification models, but it only plays an objective role in evaluating balanced datasets. When the minority category in the imbalanced dataset is misclassified, *accuracy* may still perform well. Therefore, when imbalanced datasets are used for verification, this metric cannot reflect the true performance of the model.

The other two metrics are *macro-F1* and *macro-Gmean*. In the binary classification of imbalanced data, *F1* and *Gmean* comprehensively evaluate the recognition rate of positive and negative categories, so they are usually used as effective evaluation metrics. In multi-classification problems, we can think of them as multiple binary-classification problems. Calculating the *F1* value of each category separately, and then calculating the average value, namely *macro-F1*. The larger the *macro-F1*, the better the classifier performance, its expression as follows,

$$macro-F1 = \frac{1}{q}\sum\nolimits_{i=1}^{q} \frac{2^{*}precision_i{}^{*}recall_i}{precision_i + recall_i} \tag{15}$$

$$precision_i = \frac{TP_i}{TP_i + FP_i} \tag{16}$$

where, q is the number of categories in the samples, the subscript i is the corresponding health condition, and *precision* is the accuracy of each type of sample.

Referring to the calculation steps of *macro-F1*, we propose *macro-Gmean* as another evaluation metric, so that the influence of the minority category on the model will not be ignored. The larger the *macro-Gmean*, the better the classifier performance, its formula as follows,

$$macro-Gmean = \frac{1}{q}\sum\nolimits_{i=1}^{q} \sqrt{precision_i{}^{*}recall_i} \tag{17}$$

C. IMBALANCED FAULT DIAGNOSIS THROUGH DCACN

When the four datasets introduced in Section A are used for verification, 60% of the samples used as the training set and the remaining 40% as the test set. The parameters of each round of verification are set as follows: each input sample contains 1200 data points, the convolution kernel of the convolution layer is 5, the dropout probability is 0.5, the pooling layer is 4, the mini-batch is 128, and the epochs is 200, SGD has a learning rate of 0.01 and a momentum of 0.9.

Training the model under the above parameters and explore the differences of diagnostic results in different imbalanced datasets. And each group of experiment is repeated ten times. The confusion matrices of the datasets B1, B2, B3, and B4 are shown in Fig.3, and the evaluation metrics are shown in Table 3. It can be seen from the results that the *macro-F1* and *macro-Gmean* of DCACN is greater than 93.84% and 94.76%, respectively. And the overall samples' classification accuracy is greater than 99.15%. The results show that DCACN can achieve higher accuracy of minority category while guaranteeing higher accuracy of overall samples, even if the proportion of majority category is large. In the field of mechanical fault diagnosis, it is particularly important to identify the minority category.

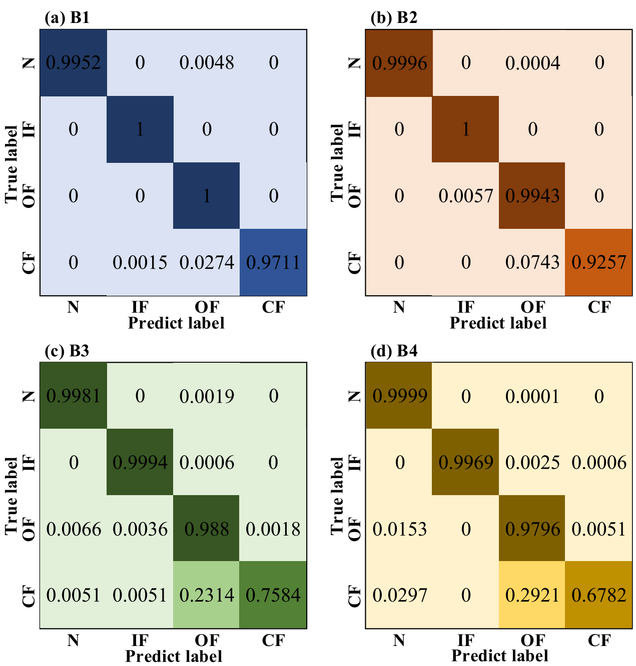


FIGURE 3. The confusion matrix of DCACN.

D. COMPARISON RESULTS

As shown in Table 4, we use other seven fault diagnosis methods to compare with DCACN in four aspects for verifying the effectiveness of the proposed DCACN. The seven methods are SVM, MLP, CNN, CSCNN [30], SMOTE, SMOTETomek, and BalancedBagging, and the comparative experimental design as follows.

1) Comparison to handcrafted features based fault diagnosis method. First, eight features such as variance, root mean square, maximum value, average absolute value, kurtosis index, margin index, waveform index, peak index, and skewness coefficient are extracted, and then the input feature vectors are learned and classified by SVM. This set of verification is to illustrate the effect of learning characteristics on the imbalanced fault diagnosis.

2) Comparison to the fault diagnosis methods that introduce the default misclassification cost in training process. MLP is a shallow artificial neural network that can learn features in training data and establish a mapping relationship between input data and labels. CNN is a deep learning algorithm with representational learning ability, which can perform translation-invariant classification of input information according to its hierarchical structure. The input data of both methods are raw vibration signals, and both are able to learn features from the raw data and then classify them, but there is no difference in the misclassification costs for different categories. So this verification aims to study the effect of setting different misclassification costs for majority and minority categories.

3) Comparison to the imbalanced fault diagnosis method which presetting the constant misclassification costs in training process. The difference between CSCNN and CNN is that CSCNN sets different misclassification costs for different types of samples according to the imbalanced ratio of the samples, and it does not change in training process. It is expected that the strategy of adaptively updating misclassification costs proposed in this paper is able to effectively identify the minority category in imbalanced mechanical data.

4) Comparison to several famous imbalanced data solutions. SMOTE is able to synthesize the minority samples until the data is balanced. Balanced Bagging Classifier (BBC) is an ensemble learning method, and it works by building multiple estimators on different randomly selected subsets of the data. SMOTE-TL is a comprehensive sampling method for imbalanced data, which combines SMOTE and Tomek Links to achieve a balanced dataset. In this part, when SMOTE or SMOTE-TL is used to resample imbalanced datasets, we still use CNN to train and test the resampled balanced data. The imbalanced datasets B2, B3 and B4 will be used for verification. The superiority of the proposed method is expected to be demonstrated through this set of validations.

The recognition *accuracy*, *macro-F1*, and *macro-Gmean* of the five methods are shown in Table 3, and the metrics with the best recognition effect are shown in bold. It can be seen that the proposed DCACN is superior to other comparison methods. Since DCACN, CSCNN and CNN are all deep networks, here we use paired t-test to verify the difference between DCACN and the other two methods, and further illustrate the superiority of the proposed method. Paired t-test is used to analyze the difference between the paired quantitative data. Compared to the independent sample t-test, the paired t-test requires the samples to be paired. During validating, each method is run ten times with any datasets and three metrics (*accuracy, macro-F1, and macro-Gmean*) are recorded under each dataset. Therefore, we use these metrics for paired t-test. In this section, four bearing datasets with different imbalanced ratio are used for validating.

TABLE 3. The results of imbalanced bearing datasets.

Method	B1			B2			B3			B4		
	acc	macro-F1	macro-Gmean	acc	macro-F1	macro-Gmean	acc	macro-F1	macro-Gmean	acc	macro-F1	macro-Gmean
DCACN	**0.9915**	**0.9915**	**0.9915**	**0.9936**	**0.9793**	**0.9844**	**0.9926**	**0.9518**	**0.9587**	**0.9947**	**0.9384**	**0.9476**
SVM	0.8381	0.8388	0.8345	0.9264	0.7608	0.8064	0.9423	0.6661	0.7197	0.9541	0.6626	0.7222
MLP	0.9267	0.9267	0.9262	0.8525	0.6975	0.7104	0.9137	0.6781	0.7132	0.924	0.6649	0.7065
CNN	0.9688	0.9681	0.9672	0.9435	0.8087	0.8448	0.9561	0.7033	0.7511	0.9747	0.7244	0.7782
CSCNN	0.9873	0.9872	0.9871	0.9852	0.9606	0.9628	0.9886	0.9344	0.9421	0.9816	0.8432	0.8734
SMOTE	-	-	-	0.9676	0.9224	0.9427	0.9270	0.8338	0.8677	0.9566	0.8111	0.8406
BBC	-	-	-	0.6809	0.5120	0.5445	0.6009	0.4003	0.5131	0.5768	0.3667	0.4963
SMOTE-TL	-	-	-	0.9121	0.8446	0.8763	0.9195	0.7811	0.8207	0.9008	0.7680	0.8158

TABLE 4. Various fault diagnosis methods.

Methods	Features	Misclassification cost
DCACN	Learned features	Adaptive
SVM	Handcrafted features	Default
MLP	Learned features	Default
CNN	Learned features	Default
CSCNN	Learned features	Constant
SMOTE	Over sampling	
BBC	Ensemble learning	
SMOTE-TL	Comprehensive sampling	

When verifying the effectiveness of DCACN, each metric mentioned above is recorded ten times in each dataset, and the results of other methods are recorded in the same way. Table 5 shows the paired t-test results of DCACN and CNN on bearing datasets, in which pair 1 is the accuracy-pair including ten times of DCACN's *accuracy* and ten times of CNN's *accuracy* on dataset B1. Pair 2 and 3 are the Gmean-pair and F1-pair on dataset B1, respectively. The remaining nine pairs in Table 5 are the accuracy-pair, Gmean-pair, and F1-pair on the other three datasets. Table 6 shows the paired t-test results of DCACN and CSCNN on bearing datasets. The S.E. Mean in these tables is the standard error mean, and the Sig. means significance level. In this part, we use 95% confidence, in other words, when the Sig. is less than 0.05, the two methods involved in the paired t-test are significantly different. Since the number of samples in dataset B1 is balanced, the Sig. is greater than 0.05, indicating that DCACN is not significantly different from the other two methods. However, in other imbalanced datasets, the Sig. is basically less than 0.05, indicating that DCACN is significantly different from the other two methods.

Through these comparative experiments, four results can be obtained: 1) In the imbalanced classification of mechanical data, the method of learning features performs better than the method with handcrafted features. In comparison, SVM identifies the health condition with 8 types of features extracted from raw vibration signals, and other methods directly learn features from raw data. The experimental results show that the four methods of learning features are better than the method with handcrafted features. The reason is that handcrafted features may lose some key information in raw data. 2) When dealing with imbalanced mechanical data, setting a small misclassification cost for the majority category and a large misclassification cost for the minority category can effectively improve the recognition rate of the minority category. MLP and CNN use the default misclassification costs, and all categories have the same misclassification costs. In addition, the results of these two methods show that the feature learning ability of deep networks is superior than that of shallow networks. CSCNN and DCACN assign different misclassification costs to each category. The latter two methods are better at identifying imbalanced data. The difference between

TABLE 5. Paired t-test results of DCACN and CNN on bearing datasets.

Pair	Dataset	Average	Standard deviation	S.E. Mean	Sig.
1	B1	-0.023	0.046	0.015	0.155
2		-0.023	0.048	0.015	0.159
3		-0.024	0.050	0.016	0.159
4	B2	-0.050	0.027	0.009	0.000
5		-0.171	0.094	0.030	0.000
6		-0.140	0.080	0.025	0.000
7	B3	-0.037	0.018	0.006	0.000
8		-0.248	0.108	0.034	0.000
9		-0.208	0.094	0.030	0.000
10	B4	-0.020	0.016	0.005	0.003
11		-0.214	0.076	0.024	0.000
12		-0.169	0.061	0.019	0.000

TABLE 6. Paired t-test results of DCACN and CSCNN on bearing datasets.

Pair	Dataset	Average	Standard deviation	S.E. Mean	Sig.
1	B1	-0.004	0.007	0.002	0.082
2		-0.004	0.007	0.002	0.079
3		-0.004	0.007	0.002	0.080
4	B2	-0.008	0.007	0.002	0.006
5		-0.019	0.018	0.006	0.009
6		-0.022	0.018	0.006	0.005
7	B3	-0.004	0.007	0.002	0.085
8		-0.017	0.034	0.011	0.136
9		-0.017	0.024	0.008	0.060
10	B4	-0.013	0.006	0.002	0.000
11		-0.095	0.044	0.014	0.000
12		-0.074	0.031	0.010	0.000

CNN, CSCNN, and DCACN is only that the misclassification costs are different, so there is not much difference in model performance when processing balanced dataset B1. But when dealing with imbalanced datasets B2, B3, and B4, the performance of CNN is obviously inferior to that of CSCNN and DCACN. It can be seen that when identifying imbalanced mechanical data, setting different misclassification costs for each category can improve the effectiveness of the model. 3) Compared with CSCNN, DCACN achieved better recognition effect on imbalanced data. The confusion matrices of DCACN and CSCNN on the four datasets are shown in Fig. 3 and 4 respectively. It can be seen that in the three imbalanced datasets B2, B3, and B4, the recognition accuracy of the CSCNN for the category with the fewest samples (that is, the composite fault CF) is 83.56%, 73.32%, and 40.93%, and the recognition accuracy of DCACN is 93%, 75.84%, 67.82%, CSCNN will misclassify some of the majority samples into the minority samples when processing highly imbalanced datasets. This shows that it is more effective to constantly adjust the misclassification costs based on the results of each mini-batch in training process than to preset constant misclassification costs. 4) Using SMOTE and SMOTE-TL to resample the original mechanical signals and then using CNN is better than directly using CNN. However, these two methods are not as effective as DCACN, especially on the dataset B4 with the highest imbalanced ratio, the proposed method is significantly better than these two methods. As for the BBC, it performs poorly on all three imbalanced datasets, and it may not be suitable for mechanical vibration signals. It can be seen that DCACN is superior than some famous imbalanced data solutions.

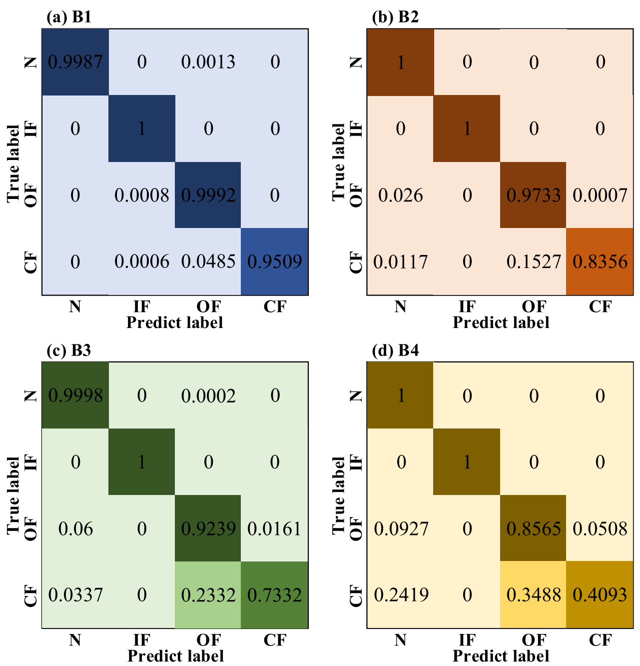


FIGURE 4. The confusion matrix of CSCNN.

IV. CASE STUDY II: IMBALANCED FAULT DIAGNOSIS OF MILLING CUTTERS

In order to further verify the effectiveness and superiority of the proposed method for processing imbalanced mechanical fault data, we use DCACN in this section to classify the health condition of milling cutters and compare it with other four classification methods.

A. MILLING CUTTER EXPERIMENT

The milling cutter experiment is performed on a DAHENG VMC850 CNC machining center. The material of the workpiece cut on the machining center is 45 steels, the size is 100mm*60mm*60mm(length*width*height), and the milling cutter is APMT 1135 Duracarb. Cutting conditions as follows: the spindle speed is 2500 rpm, the feed speed is 200 mm/min, and the cutting depth is 2 mm. During the cutting process, the table feeds the workpiece from left to right along the X axis. After each pass, the blade is taken off and the abrasion value is measured with a GP-300C industrial microscope. The PCB 356A15 acceleration sensor is used to collect the vibration signals of the spindle, and the sampling frequency is 10kHz. The arrangement of the experimental equipment is shown in Fig.5, and the acceleration sensor is arranged on the spindle. According to the milling cutter life criterion recommended in GBT16460-1996, once the average width value (VB) of the wear surface of the flank face reaches 0.3mm, it is considered that the milling cutter is blunt, and the experiment is stopped concurrently. The collected raw signals are the life cycle data of the milling cutter, and the two forms of cutter failure during the experiment: flank wear and brittle fracture, as shown in Fig.6 (a) (b).

B. THE DIVISION OF DATASETS

According to literature [31], the milling cutter's health condition is divided into three stages by the abrasion value during the whole life cycle. The three stages are the running-in period (VB<0.1mm), normal operation (0.1mm<VB< 0.3mm), and wear (VB>0.3mm), which are represented by

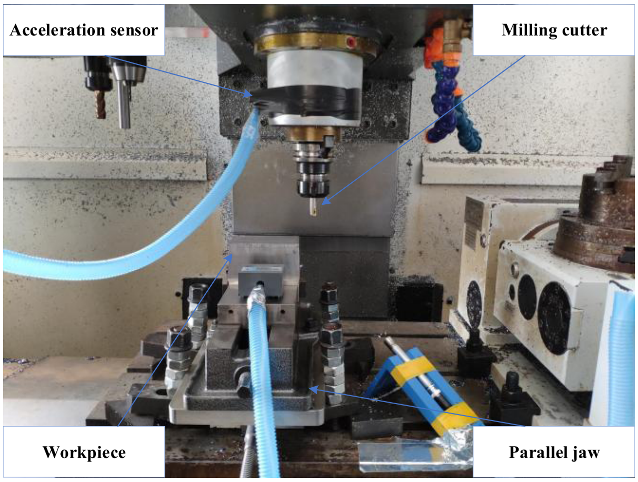
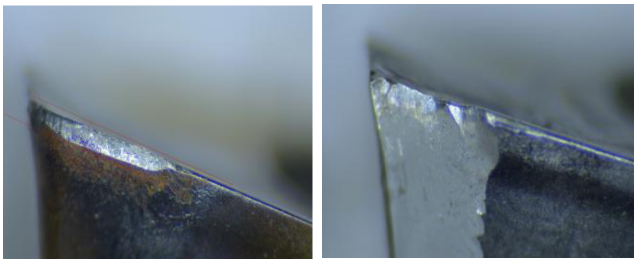


FIGURE 5. CNC milling machine and sensor placement.



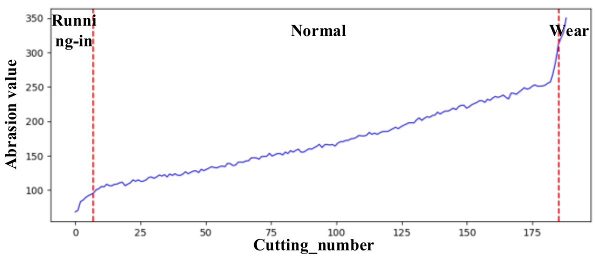
(a) Flank wear (b) Brittle fracture

FIGURE 6. Cutter failure form. (a) Flank wear. (b) Brittle fracture.

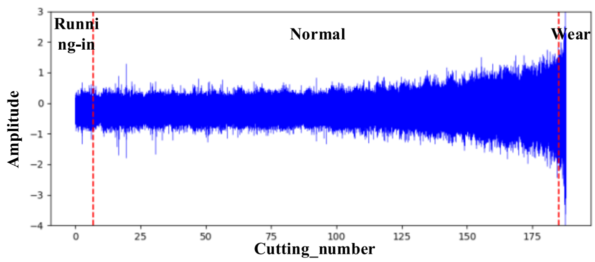
TABLE 7. The results of milling cutters' datasets.

Method	M1			M2		
	acc	macro-F1	macro-Gmean	acc	macro-F1	macro-Gmean
DCACN	**0.9521**	**0.9517**	**0.9505**	0.9893	**0.8854**	0.9859
SVM	0.7977	0.7927	0.7916	0.7742	0.7698	0.7683
MLP	0.8596	0.8566	0.8547	**0.9906**	0.7809	0.7607
CNN	0.9504	0.9499	0.9501	0.9898	0.7981	0.8352
CSCNN	0.9517	0.9511	0.9505	0.9808	0.8154	**0.9867**
SMOTE	-	-	-	0.9749	0.7513	0.9743
BBC	-	-	-	0.8350	0.4104	0.7854
SMOTE-TL	-	-	-	0.9771	0.7714	0.9675

RI, N, and W respectively. We select a cutter's data to build the verification datasets, the abrasion value over the entire life cycle is shown in Fig.7(a), and the raw vibration signals are shown in Fig.7(b). M1 is a balanced dataset, and M2 is an imbalanced dataset. The samples' number for these two datasets is 1200 and 4080, respectively. During the cutting process, the data of the normal operating state is easier to be obtained, and the data of the other two states are more difficult to be obtained. Therefore, the proportion of the normal state samples is increased in M2, and the ratio of the running-in period, normal operation and wear data is 1:100:1. At the time of verification, 60% of the data is used for training, and the remaining 40% for testing.



(a) Abrasion value



(b) Raw vibration signals

FIGURE 7. The division of milling cutter data.

C. VERIFICATION RESULTS AND COMPARISON

When verifying the proposed method with milling cutter datasets, the same parameters as that in the previous chapter are used. At the same time, *accuracy*, *macro-F1* and *macro-Gmean* are used for quantitative evaluation of classification results, and the verification results of the eight methods on two datasets are shown in Table 7. From the table, we can see that when processing the balanced dataset M1, the *accuracy* of SVM is only 79.77%, which is far from the other four methods. It can be seen the effect of using SVM classify the handcrafted features is significantly lower than other methods by learning features from raw data. When using MLP process imbalanced dataset, the *accuracy* reached 99.06%, but macro-F1 and *macro-Gmean* are lower than the other three types of deep neural network. The main reason is that the data in the normal state of the milling cutter are very lot. The recognition rate for this state is very high and for the other two states is very low. This shows that the shallow network has a weak ability in learning and characterizing raw data. The three methods CNN, CSCNN, and DCACN have significant differences only when dealing with the imbalanced dataset M2, their *macro-F1* is 0.7981, 0.8154, 0.8854, and the *macro-Gmean* is 0.8352, 0.9867, and 0.9859 respectively. The *macro-F1* of SMOTE and SMOTE-TL is 0.7513 and 0.7714 respectively, which is less than that of CNN, but the *macro-Gmean* is greater than CNN. It can be seen that the three methods have little difference in the recognition effect of imbalanced milling cutter data. The recognition effect of BBC is still lower than other methods.

We still use paired t-test to verify the differences between CNN, CSCNN, and DCACN. The paired t-test results are shown in Table 8 and 9. In this part, two milling cutter datasets are used, so pair 1 to 3 in Table 8 and 9 are respectively accuracy-pair, Gmean-pair and F1-pair on dataset M1, and pair 4 to 6 are that on dataset M2. Since the dataset M1 is balanced, the Sig. associated with this dataset in both tables is greater than 0.05. This means that when the dataset is balanced, there is almost no difference between the three

TABLE 8. Paired t-test results of DCACN and CNN on milling cutter datasets.

Pair	Dataset	Average	Standard deviation	S.E. Mean	Sig.
1		-0.002	0.037	0.012	0.890
2	M1	-0.002	0.038	0.012	0.886
3		0.000	0.040	0.013	0.978
4		0.000	0.008	0.002	0.842
5	M2	-0.087	0.125	0.039	0.054
6		-0.151	0.111	0.035	0.002

TABLE 9. Paired t-test results of DCACN and CSCNN on milling cutter datasets.

Pair	Dataset	Average	Standard deviation	S.E. Mean	Sig.
1		0.000	0.044	0.014	0.977
2	M1	-0.001	0.046	0.014	0.968
3		0.000	0.049	0.015	0.996
4		-0.009	0.008	0.003	0.009
5	M2	-0.070	0.047	0.015	0.001
6		0.001	0.014	0.004	0.863

methods of DCACN, CSCNN, and CNN. For the imbalanced dataset M2, the Sig. of some pairs is less than 0.05, which also indicates that DCACN is different from CSCNN and CNN. And this is also consistent with the paired t-test results in case study I. These comparisons further validate the advantages of DCACN in dealing with imbalanced mechanical data.

V. CONCLUSION

Aiming at the problem that the existing intelligent fault diagnosis methods have low accuracy in identifying the minority category in imbalanced mechanical data, we propose an intelligent mechanical fault diagnosis method with adaptive misclassification costs. During training, the model adaptively assigns appropriate misclassification costs for all samples, and continuously updates to achieve effective classification of imbalanced mechanical data. The proposed method is verified by the bearing and milling cutter datasets, and compared with other four methods, the following conclusions can be drawn. 1) Learning features from vibration signals can make full use of hidden information in raw data. 2) Setting different misclassification costs for each category can effectively improve the recognition rate of the minority category. 3) When the imbalanced ratio is large, the method with constant misclassification costs will misclassify some majority samples into minority samples, and the method with adaptive misclassification costs proposed in this paper shows robustness. The above experiments and conclusions show that the proposed method can solve the problem of imbalanced mechanical fault classification to a certain extent under the condition of lack of fault samples, and provide a potentially feasible solution for the industrial application of intelligent fault diagnosis methods.

REFERENCES

[1] X. F. Chen and Y. Y. Zi, *Intelligent Maintenance and Health Management*. Beijing, China: China Machine Press, 2018.
[2] Y. Si, Y. Q. Wang, and D. Zhou, "Key-performance-indicator-related process monitoring based on improved kernel partial least squares," *IEEE Trans. Ind. Electron.*, early access, Feb. 13, 2020, doi: 10.1109/TIE.2020.2972472.
[3] Y. G. Lei, F. Jia, D. Kong, J. Lin, and S. Xing, "Opportunities and challenges of machinery intelligent fault diagnosis in big data era," *J. Mech. Eng.*, vol. 54, no. 2, pp. 94–104, 2008.
[4] L. Guo, H. Gao, H. Huang, X. He, and S. Li, "Multifeatures fusion and nonlinear dimension reduction for intelligent bearing condition monitoring," *Shock Vibrat.*, vol. 2016, pp. 1–10, Feb. 2016.
[5] Y. Wang, Y. Si, B. Huang, and Z. Lou, "Survey on the theoretical research and engineering applications of multivariate statistics process monitoring algorithms: 2008-2017," *Can. J. Chem. Eng.*, vol. 96, no. 10, pp. 2073–2085, Oct. 2018.
[6] G. E. Hinton, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006.
[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[8] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, "Natural language processing (almost) from scratch," *J. Mach. Learn. Res.*, vol. 12, no. 76, pp. 2493–2537, 2011.
[9] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-R. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, and B. Kingsbury, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, Nov. 2012.
[10] L. Jing, M. Zhao, P. Li, and X. Xu, "A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox," *Measurement*, vol. 111, pp. 1–10, Dec. 2017.
[11] L. Guo, Y. Lei, S. Xing, T. Yan, and N. Li, "Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data," *IEEE Trans. Ind. Electron.*, vol. 66, no. 9, pp. 7316–7325, Sep. 2019.
[12] X. Zhang, X. Feng, Z. Mu, and Y. Q. Wang, "State and fault estimation for nonlinear recurrent neural network systems: Experimental testing on a three-tank system," *Can. J. Chem. Eng.*, early access, Jan. 30, 2020, doi: 10.1002/cjce.23714.
[13] L. Zhang, H. Gao, D. Dong, G. Fu, and Q. Liu, "Wear calculation-based degradation analysis and modeling for remaining useful life prediction of ball screw," *Math. Problems Eng.*, vol. 2018, pp. 1–18, Nov. 2018.
[14] L. Guo, N. Li, F. Jia, Y. Lei, and J. Lin, "A recurrent neural network based health indicator for remaining useful life prediction of bearings," *Neurocomputing*, vol. 240, pp. 98–109, May 2017.
[15] L. Guo, Y. Lei, N. Li, T. Yan, and N. Li, "Machinery health indicator construction based on convolutional neural networks considering trend burr," *Neurocomputing*, vol. 292, pp. 142–150, May 2018.
[16] V. López, A. Fernández, S. García, V. Palade, and F. Herrera, "An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics," *Inf. Sci.*, vol. 250, pp. 113–141, Nov. 2013.
[17] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009.
[18] T. Alam, C. F. Ahmed, S. A. Zahin, M. A. H. Khan, and M. T. Islam, "An effective recursive technique for multi-class classification and regression for imbalanced data," *IEEE Access*, vol. 7, pp. 127615–127630, 2019.
[19] H. Fang and Y. Li, "Random undersampling and POSS method for software defect prediction," *J. Shandong Univ., Eng. Sci.*, vol. 47, no. 1, pp. 15–21, 2017.
[20] W.-C. Lin, C.-F. Tsai, Y.-H. Hu, and J.-S. Jhang, "Clustering-based undersampling in class-imbalanced data," *Inf. Sci.*, vols. 409–410, pp. 17–26, Oct. 2017.
[21] Y. Y. Wu and L. Y. Shen, "Imbalanced fuzzy multiclass support686 vector machine algorithm based on class-overlap degree undersampling," *J Univ. Chin. Acad. Sci.*, vol. 35, no. 4, pp. 536–543, 2018.
[22] H. S. Huang, J. A. Wei, and P. D. Kang, "A new over-sampling SVM classification algorithm based on unbalanced data sample characteristics," *Control Decis.*, vol. 2018, vol. 33, no. 9, pp. 1549–1558.
[23] L. Abdi and S. Hashemi, "To combat multi-class imbalanced problems by means of over-sampling techniques," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 1, pp. 238–251, Jan. 2016.
[24] F. Jia, Y. Lei, N. Lu, and S. Xing, "Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization," *Mech. Syst. Signal Process.*, vol. 110, pp. 349–367, Sep. 2018.
[25] S. Dhar and V. Cherkassky, "Development and evaluation of cost-sensitive universum-SVM," *IEEE Trans. Cybern.*, vol. 45, no. 4, pp. 806–818, Apr. 2015.
[26] N. V. Chawla, A. Lazarevic, L. O. Hall, and K. W. Bowyer, "SMOTEBoost: Improving prediction of the minority class in boosting," in *Proc. PKDD Eur. Conf. Princ. Pract. Knowl. Discovery Databases (DBLP)*, Cavtat, Croatia, Sep. 2003, pp. 107–119.
[27] J. Yan and S. Han, "Classifying imbalanced data sets by a novel RE-sample and cost-sensitive stacked generalization method," *Math. Problems Eng.*, vol. 2018, Jan. 2018, Art. no. 5036710, doi: 10.1155/2018/5036710.
[28] V. Raj, S. Magg, and S. Wermter, "Towards effective classification of imbalanced data with convolutional neural networks," in *Artificial Neural Networks in Pattern Recognition*. Cham, Switzerland: Springer, 2016.

[29] C. Lessmeier, J. K. Kimotho, and D. Zimmer, "Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: A benchmark data set for data-driven classification," in *Proc. Eur. Conf. Prognostics Health Manage. Soc.*, 2016, pp. 1–17.
[30] B. Krawczyk and M. Woźniak, "Cost-sensitive neural network with ROC-based moving threshold for imbalanced classification," in *Proc. Int. Conf. Intell. Data Eng. Automated Learn.* Cham, Switzerland: Springer, 2015, pp. 45–52.
[31] R. Liu, *Research on the Recognition and Prediction of Tool Wear State*. Chengdu, China: Southwest Jiaotong Univ., 2014.

XUN DONG was born in Xuzhou, China. He received the B.S. degree in measuring and control technology and instrumentations from Nantong University, Nantong, China, in 2018. He is currently pursuing the master's degree with Southwest Jiaotong University, Chengdu, China.

His research interest includes machinery intelligent fault diagnosis.

HONGLI GAO received the Ph.D. degree in mechanical engineering from Southwest Jiaotong University, Chengdu, in 2005.

He is currently a Professor with Southwest Jiaotong University. His research interests focus on designing and reliability analysis of complex electromechanical equipment systems, intelligent machinery state monitoring, and fault diagnosis.

LIANG GUO received the B.S. and Ph.D. degrees in mechanical engineering from Southwest Jiaotong University, Chengdu, China, in 2011 and 2016, respectively.

He is currently an Assistant Professor with Southwest Jiaotong University. Prior to joining Southwest Jiaotong University, in 2018, he was a Postdoctoral Research Fellow with Xi'an Jiaotong University, Xi'an, China. He was also a Visiting Ph.D. Student with Clemson University, Clemson, SC, USA. His research interests focus on machinery intelligent fault diagnostics and remaining useful life prediction.

KESI LI received the B.S. degree in measurement and control technology and instruments from Southwest Jiaotong University, Chengdu, China, where she is currently pursuing the Ph.D. degree in mechanical engineering. She was a Visiting Ph.D. Student with Northwestern University, Evanston, IL, USA. Her research interests include the condition monitoring, prognostics health management, and nondestructive test and evaluation.

ANDONGZHE DUAN received the B.S. degree in mechanical engineering from the Anhui University of Technology, Ma'anshan, China. He is currently pursuing the master's degree with Southwest Jiaotong University, Chengdu, China.

His current research interests include machinery condition monitoring and data-based intelligent fault diagnostics.