

## Research Article

# A New Transfer Learning Ensemble Model with New Training Methods for Gear Wear Particle Recognition

Chunhua Zhao,<sup>1,2,3</sup> zhangwen Lin,<sup>3</sup> Jinling Tan ,<sup>2,3</sup> Hengxing Hu,<sup>3</sup> and Qian Li<sup>3</sup>

<sup>1</sup>Hubei Key Laboratory of Hydroelectric Machinery Design & Maintenance, China Three Gorges University, Yichang, Hubei 443002, China

<sup>2</sup>Quality Education Center for College Students, China Three Gorges University, Yichang, Hubei 443002, China

<sup>3</sup>College of Mechanical and Power Engineering, China Three Gorges University, Yichang, Hubei 443002, China

Correspondence should be addressed to Jinling Tan; [tanjinling@ctgu.edu.cn](mailto:tanjinling@ctgu.edu.cn)

Received 19 August 2021; Revised 11 October 2021; Accepted 4 December 2021; Published 5 January 2022

Academic Editor: Yi Qin

Copyright © 2022 Chunhua Zhao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Aiming at solving the acquisition problems of wear particle data of large-modulus gear teeth and few training datasets, an integrated model of LCNNE based on transfer learning is proposed in this paper. Firstly, the wear particles are diagnosed and classified by connecting a new joint loss function and two pretrained models VGG19 and GoogLeNet. Subsequently, the wear particles in gearbox lubricating oil are chosen as the experimental object to make a comparison. Compared with the other four models' experimental results, the model superiority in wear particle identification and classification is verified. Taking five models as feature extractors and support vector machines as classifiers, the experimental results and comparative analysis reveal that the LCNNE model is better than the other four models because its feature expression ability is stronger than that of the other four models.

## 1. Introduction

As a key component of special equipment such as ship lift equipment and lifting platform equipment, large-modulus gear racks and pinions will cause huge losses if they fail. Because they are running at low speed and heavy load, it is difficult to diagnose them with conventional fault diagnosis methods, and the lubrication system of large-modulus rack and pinion transmission contains a lot of wear fault information [1, 2]. Therefore, ferrographic analysis is a method of monitoring the health of the machine by observing the material, size, special diagnosis, and quantity of the wear particles, which can be used for the fault diagnosis of large-modulus rack and pinion. However, this technology relies too much on personal experience and technology, and the diagnosis time is too long to meet the actual needs of mechanical fault diagnosis under the background of big data [3].

In recent years, with computer vision [4, 5], speech recognition [6, 7], motion capture [8], and physiology [9],

the use of deep learning technology has achieved great success. Some deep learning techniques have also begun to be applied to machinery health monitoring and have made certain achievements. For example, Lu et al. proposed a stacked denoising autoencoder (SDA) model for fault diagnosis of rotating machinery components [10]. In [11], a convolutional neural network (CNN) model is used to diagnose the health status of bearings. However, the existing deep learning model cannot be directly used for the automatic identification of wear particle images, mainly because the existing deep learning model requires sufficient data, and the training dataset and the test dataset have the same feature distribution. For a small sample of wear particle datasets, the performance degradation of the intelligent fault diagnosis method is very serious. To this end, a deep learning model based on migration learning is proposed. The model trained on other large datasets is replaced with a wear particle image dataset and fine-tuned to meet the requirements of automatic wear particle images of small samples [12]. The image recognition and

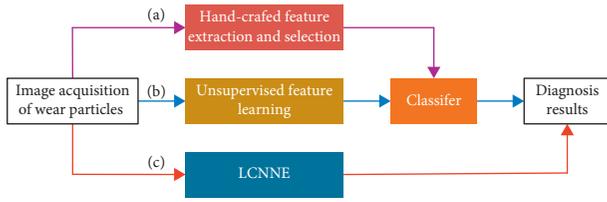


FIGURE 1: Three intelligent fault diagnosis frameworks: (a) traditional method; (b) features extracted by unsupervised learning; (c) the proposed method: LCNNE.

classification of wear particles can promote the popularization of ferrography analysis

There were two main options for using pretrained model for transfer learning [12, 13]. One was fine-tuning the model: short-term additional training was applied to the original model to add a specific training set to the knowledge base of the model. The other was to use pretrained CNN as a feature extractor to transform images into feature vectors for classification.

For transfer learning of convolutional neural network, it is very popular to use classic pretrained networks such as AlexNet [14], VGG [15], and GoogLeNet [16] for solving the image classification task for a particular application. In this paper, a method using transfer learning of pretrained VGG19 and GoogLeNet combined into an ensemble with increasing the accuracy of wear particles in diagnosis is proposed. Fine-tuning method based on optimization algorithm AdamW [17] was used to implement transfer learning.

The main advantage of this method is that the optimization algorithm AdamW was used to fine-tune the loss function convolutional neural network ensemble (LCNNE) when the general features of different pretrained networks are shared and applied to specific applications. Our method includes the following steps: (1) the pretrained VGG19 and GoogLeNet were fine-tuned for the gear abrasive images; (2) LCNNE model was initialized with the weight and biases from fine-tuned networks; (3) the optimization algorithm AdamW was used to fine-tune the LCNNE model for the gear abrasive images; (4) the fine-tuned LCNNE model can be used as a classifier by itself or as a feature extractor for an external classifier. The contributions of this paper are summarized below:

- (1) We proposed a novel and simple learning framework that processes raw wear particle images directly. A comparison of three different learning frameworks is shown in Figure 1.
- (2) The performance can be easily improved by applying the optimization algorithm AdamW to fine-tune the LCNNE model.
- (3) Four different comparison models are investigated: DCNN which needs pretraining, VGG19 and Google which are single-network transfer learning models, and VGNet which is optimized by Adam and is a combination of VGG19 and GoogLeNet. LCNNE can achieve the highest accuracy of 99.5% in the fault diagnosis of wear particles.

- (4) Fine-tuned LCNNE is used as a feature extractor to transform wear particles into input feature vectors of support vector machine (SVM) classifier for wear particle fault identification, and t-SNE is used to visualize the features of model learning to study the internal mechanism.

Classifier integration is a well-known technique for improving classification accuracy. There are many different ways to combine classifiers into ensembles. In terms of the image classification task, the ensemble approach is most commonly used for solving multiclassification task. For example, Harangi proposed the aggregation of multiple robust convolutional neural networks into a network framework, and the final classification result is the weight output of each convolutional neural network, which improves the classification accuracy of skin lesion images [18]. Dietterich combines the weighted average or unweighted voting classifiers for independent prediction to form an ensemble, which explains the reason why such sets are usually higher than the single classifiers that makes up them from the aspects of statistics, calculation, and features [19]. Gao et al. proposed a locally weighted ensemble learning framework, which is used to combine multiple models for transfer learning, in which the weights are dynamically assigned according to the predictive power of each test example in the target domain [20]. Nanni et al. proposed that the system combined multiple neural networks into a whole and combined scores by summation rules, providing a simple and effective method to improve the performance of the neural network after training. Because the system combines different feature types learned and manually extracted, it has not only strong recognition ability, but also powerful generalization ability [21]. Some of the most advanced models were carefully optimized and enhanced based on a series of rigorous analyses and evaluations. Then, a more powerful ECNN model based on multiscale feature learning method is formed by combining feature extractors of these models. A pixel-level recognition of village buildings is achieved [22]. Li et al. combined the improved D-S evidence theory with the modified Gini index and the deep convolutional neural network to form a novel bearing fault diagnosis model, IDSCNN. Using the open bearing dataset of Case Western Reserve University to verify the model, the conclusion that the model performance is better than the existing machine learning method is obtained [23]. Pre-trained convolutional neural network fine-tuning technique is successfully used in different fields. Recent research shows that pretraining on general data followed by application-specific fine-tuning yields significant performance improvement in the image classification task. A fine-tuning convolutional neural network for medical image classification is proposed. It describes a method for classifying the modalities of medical images using ensemble of different CNN architectures [24]. The various CNNs in the ensemble allow image features to be extracted at different semantic levels, thereby being able to characterize different, distinct, and subtle differences between different modes. The ensemble model of fine-tuned CNNs allows adapting to the

common features learned from natural images, making it more suitable for different medical imaging modes. Ding et al. use a trunk-branch ensemble convolutional neural network (TBE-CNN) for video-based face recognition. TBE-CNN consists of one trunk network that learns representations of the overall face image and two branch networks that learn the feature representation of image block cropped around facial components. The output feature maps of the trunk network and branch networks are concatenated in series, and finally the fully connected layer is used for classification [25].

The main contribution of diagnosis method proposed in this paper in comparison with existing approaches is a fine-tuning procedure for pretrained LCNNE model by using the optimization algorithm AdamW, which consists of pretrained VGG19 and GoogLeNet and uses the new joint loss function as the objective function. For wear particle images, each network model is fine-tuned independently before integration. Fine-tuned LCNNE can be used as image classifier or as feature extractor for further image processing.

## 2. Proposed LCCNE Intelligent Diagnosis Method

**2.1. Structure of the Proposed LCCNE Model.** As mentioned in Section 2, there are many studies on mechanical intelligent diagnosis methods based on deep learning, and most of them use the time series of vibration signals or the frequency domain signal as the diagnostic object after FFT conversion. However, they seldom identify wear particles directly. At the same time, high performance requires a larger and deeper model structure. A large CNN model means that a large amount of data is needed. However, in actual engineering applications, the information in the actual monitored data is highly repeatable and lacks typical fault information. During the long-term operation of equipment, a large amount of monitoring data is accumulated, but only a small amount of it corresponds to the health status of the equipment. This results in lack of data that can be used to train intelligent diagnosis model.

To address these problems, a transfer learning procedure for ensemble model based on the new joint loss function is proposed here. A fine-tuning method based on optimization algorithm was used to implement transfer learning. This model directly uses the wear particle images as input and can work with a small sample dataset, while ensuring a high recognition rate and generalization ability. This ensemble model is composed of the classical CNN model, VGG19, and GoogLeNet, hence the name LCNNE, where  $L$  denotes the new joint loss function and  $E$  denotes the ensemble of pretrained models VGG19 and GoogLeNet. Specifically, the last layers of VGG19 and GoogLeNet, each, were removed and connected, and the two fully connected layers were then added after the connection. The two new fully connected layers were composed of 100 fully connected neurons and Softmax layers, whose output equals the number of wear fault categories. The structure framework of the proposed LCNNE model is shown in Figure 2.

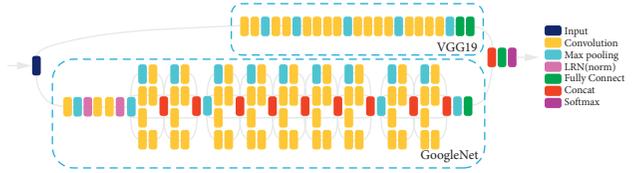


FIGURE 2: Architecture of the proposed LCNNE model.

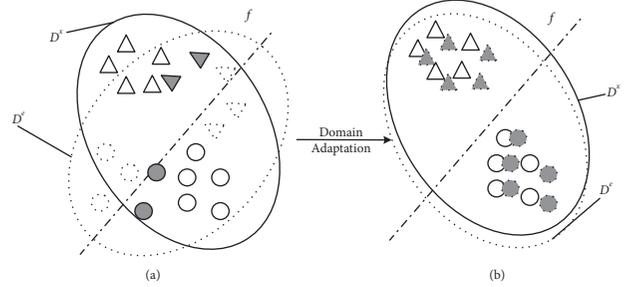


FIGURE 3: Diagram of failover diagnosis.

**2.2. Training of the LCNNE Model.** Firstly, the concepts and terminology of transfer learning in [26] are introduced before introducing the model training method.

**Definition 1 (domain).** A domain can be represented as follows:

$$\mathcal{D} = \{\mathcal{X}, P(X)\}, \quad (1)$$

where  $\mathcal{X}$  denotes the feature space and  $P(X)$  denotes the marginal probability distribution of  $X = \{x_1, \dots, x_n\} \in \mathcal{X}$ ,  $x_i \in \mathbb{R}^d$ ,  $i = 1, 2, \dots, n$ .

**Definition 2 (task).** Given a specific domain,  $\mathcal{D} = \{\mathcal{X}, P(X)\}$ , a task can be represented as follows:

$$\mathcal{T} = \{\mathcal{Y}, f(\cdot)\}, \quad (2)$$

where  $\mathcal{Y}$  represents the class label space and the function  $f(\cdot)$  represents a target function that can predict the class label of an arbitrary instance. A solution to a task can be learned from the training data,  $\{(x_i, y_i)\}_{i=1}^m$ .

Pan et al. define transfer learning as follows: Given a source domain  $\mathcal{D}^s$ , a learning task  $\mathcal{T}^s$ , a target domain  $\mathcal{D}^t$ , and a learning task  $\mathcal{T}^t$ , transfer learning aims to help improve the learning of the target predictive function,  $f(\cdot)$ , in  $\mathcal{D}^t$ , using the knowledge in  $\mathcal{D}^s$  and  $\mathcal{D}^t$ , where  $\mathcal{D}^s \neq \mathcal{D}^t$  or  $\mathcal{T}^s \neq \mathcal{T}^t$  [25]. Therefore, the nonlinear mapping relationship  $\mathcal{Y}^s = f(\mathcal{X}^s)$  from the sample space of source domain data,  $\mathcal{X}^s$ , to the label space,  $\mathcal{Y}^s$ , is established through the training of source domain data samples, that is, the acquired fault diagnosis knowledge. As shown in Figure 3(a), the source domain fault diagnosis knowledge cannot accurately identify the target domain fault marker sample category, due to the large difference in distribution between the source domain and the target domain data, resulting in misjudgment of fault diagnosis. As shown in Figure 3(b), this study aims to develop an ensemble model for transfer learning, to adapt the monitoring data distribution of source domain ImageNet dataset [27], and target domain wear

particle images, for diagnosis knowledge of source domain ImageNet dataset, to identify the health state of target domain wear particle images.

Figure 1 shows the ensemble model structure of LCNNE which is composed of four parts, namely, the input layer, multilevel convolution layer, fully connected layer, and output layer. The feature vector of the model can be obtained by calculating the convolution sum of the convolution kernel,  $k \in R$ , and the input sample  $x_i^{\mathcal{D}} = \{x_i^s, x_i^t\}$ , where  $x_i^s$  and  $x_i^t$  are the  $i$ th samples of the source domain ImageNet dataset and the target domain wear particle images, respectively, as follows: where  $\theta^{\mathcal{D}} = \{w, b\}$  is the set of parameters to be trained for the ensemble model.

$$\begin{aligned} F &= f(x_i^{\mathcal{D}}, \theta^{\mathcal{D}}) \\ &= \sigma[x_i^{\mathcal{D}} * w + b], \end{aligned} \quad (3)$$

The convolution kernel parameters,  $w$ , are divided into common feature and domain feature parameters:

$$w = w_c + v_s + v_t, \quad (4)$$

where  $w_c$  is a common parameter, while  $v_s$  and  $v_t$  are specific parameters for the source task and the target task, respectively. Equation (3) can be rewritten as follows:

$$\begin{aligned} F &= f(x_i^{\mathcal{D}}, \theta^{\mathcal{D}}) \\ &= \sigma[x_i^t * (w_c + v_s + v_t) + b]. \end{aligned} \quad (5)$$

In the ensemble model training stage, in samples corresponding to the source domain,  $\mathcal{D}^s$ , the  $v_t$  corresponding feature values are all 0; in the samples corresponding to the target domain,  $\mathcal{D}^t$ , the  $v_s$  corresponding feature values are all 0. Thus, the two domains learn from each other by sharing  $w_c$ .

In practical application, there is usually a big difference in the amount of data corresponding to  $\mathcal{D}^s$  and  $\mathcal{D}^t$ . During model iteration training, the number of iterations corresponding to different domains can be adjusted, the data can be sampled, or the cost sensitive loss function can be defined for adjustment. At the same time, it should be noted that transfer learning only focuses on the learning effect of the target domain,  $\mathcal{D}^t$ . The loss function is modified to prioritize the corresponding learning task,  $\mathcal{S}^t$ . The new joint loss function is defined as follows:

$$\text{Loss} = -\frac{1}{n} \sum_{i=1}^n (y_i^t)^T \cdot \log(P_i^t) + \frac{\lambda_1}{2} \sum_{i=1}^n \|v_t\|^2 + \lambda_2 \|w_0\|^2, \quad (6)$$

where  $n$  is the number of training samples in minibatch,  $\lambda_1$  and  $\lambda_2$  are the penalty factors for the marginal probability distribution, and  $P_i^t$  is the probability distribution of the final prediction mark after the target domain sample passes through Softmax layer. After the fine-tuning process of the ensemble model is finalized, it can be used as a classifier, or features from its different layers can be extracted and used as an input of external classifiers.

The proposed training method includes four main steps. Firstly, the pretrained VGG19 and GoogLeNet are fine-tuned for the gear abrasive images. Secondly, the ensemble

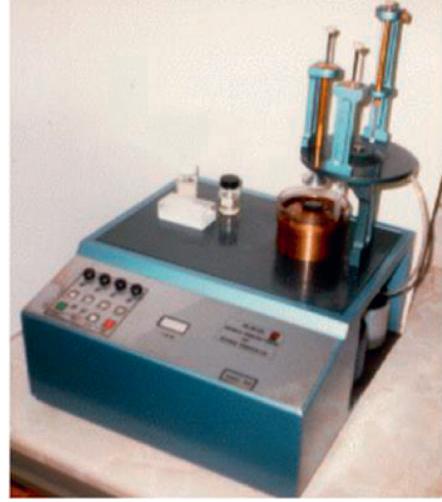


FIGURE 4: The rotary particle depositor.

model is initialized by using weights and biases from single fine-tuned networks. In the third step, the optimization algorithm AdamW [17] is used to fine-tune the LCNNE model for the gear wear particle images. In the fourth step, final wear particle image recognition takes place. The proposed fine-tuned LCNNE model can be used as a classifier or as a feature extractor for an external classifier. For each pretrained CNN fine-tuning process, the last fully connected layer of VGG19 or GoogLeNet is replaced by two new fully connected layers, consisting of a fully connected layer with 100 neurons and Softmax layers equal to the number of wear fault categories. The new layer parameters are then randomly initialized. After fine-tuning each CNN model, the last hidden layer can acquire a feature vector whose feature dimension is equal to 100. The feature vector can be uploaded to the Softmax layer, to obtain the probabilities of the corresponding fault wear classes in the test image, or can be also used as the input of additional classifier to obtain test pattern class. The weights and biases of pretrained VGG19 and GoogLeNet, after removing the last layer, are used for initializing the LCNNE ensemble model. The two fully connected new layers are then added after concatenation. The new layer is randomly initialized before starting the fine-tuning process. Details of AdamW for LCNNE are described in Algorithm 1.

where  $\lambda$  is the learning rate,  $\lambda_1$  and  $\lambda_2$  are the regular term penalty factors,  $\rho_1$  and  $\rho_2$  are the moment estimation exponential attenuation rates,  $\epsilon$  is the numerical stability constant,  $g$  calculates the gradient,  $s_i$  is the updating biased first moment estimation,  $v_i$  is the updating biased second moment estimation,  $\hat{s}_i$  is the fixed first-order moment deviation,  $\hat{v}_i$  is the correction of second-order moment deviation, and  $\theta_i$  is the weight updates.

### 3. Experiments

**3.1. Data Description.** In the current study, the wear particle information in lubricating oil of gearbox is collected by the rotary particle depositor, as shown in the test bench. The

rotary particle depositor extracted wear debris from a carrier fluid by the action of magnetic, centrifugal, and gravitational forces on the debris. The debris is deposited onto a substrate, in the form of three concentric rings, namely, inner, middle, and outer rings, as shown in Figure 4. During the deposition process, the wear debris also undergoes a sizing process, such that the inner ring has full particle size range, the middle ring has intermediate- and small-sized particles, and the outer ring has small-sized particles. The particles were viewed by using an optical microscope to obtain information regarding the shape, size, concentration, and type of particles present, alongside some information on their composition, as shown in Figure 5. A metallurgical microscope was used for imaging because it has facilities for both reflected and transmitted light, polarizers, and magnifications of up to  $\times 400$ .

The images of the wear particle datasets were captured by using a Sony 3CCD color video camera attached to the ferroscope. This enabled the display of live images on the monitor, running SYNOPTICS GRABBER software in a PC running on Windows operating system. Images were then captured and saved in the computer hard drive. During the study, a total of 64 original image datasets were adopted. Since the original size of the images did not conform to the input rules of the model, they were scaled to  $240 \times 240$  pixel images, and the long side was used as the reference. According to morphology of wear particles [28], they can be divided into 10 categories: normal, spherical, sliding, fatigue chunk, laminar, oxidized, cutting, rubbing, nonferrous, and nonmetal, with the corresponding category labels represented by 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9, respectively.

**3.2. Data Augmentation.** As mentioned in Section 4.1, the experiment obtained 64 original wear particle image datasets, which were insufficient to train the network model having numerous parameters. To improve classification accuracy, data augmentation technology was used to increase the datasets. Data augmentation technology uses random cropping, affine transformation, perspective transformation, color jitter, contrast enhancement, superimposed noise, and other methods to introduce slight interference into an image sample for data augmentation, which can reduce the overfitting of the training phase and improve the generalization performance of the network. The affine transformation includes basic transformations such as translation, rotation, scaling, flipping, and miscutting. It has straightness and parallelism; that is, the straight lines remain straight after transformation, and the parallel lines remain parallel, as shown in Figure 6. Perspective transformation can be used to simulate different perspectives of the actual scenario.

During the study, the original wear particle image samples were flipped, randomly clipped, and perspective-transformed randomly to expand the experimental dataset.

For the training sample, the images of  $240 \times 240$  pixels were randomly cropped into input sample of  $224 \times 224$  pixels through overlapping sampling techniques, which is equivalent to expansion of the dataset by  $16 \times 16 = 256$  times. For the test

samples, random rotation and random perspective transformation are used, instead of overlapping technology, to expand the test dataset, thus greatly expanding the diversity of the dataset. To avoid severe distortion of the transformed image, the displacement of the corresponding point in the perspective transformation is limited to around 10% of the image edge length, and the canvas size of the target image is consistent with the original image.

Using data augmentation technology, the original image of wear particles is expanded, and 10,800 samples are selected as the training set and 800 samples as the test set. The details of the wear particle image dataset are provided in Table 1.

**3.3. Experiment Settings.** The model is initialized using the weights and biases of the pretrained VGG19 and GoogLeNet models fine-tuned on wear particle images, and the last layer of the model is replaced by the fully connected layer with a unit number of 100 and the corresponding Softmax classifier. Finally, the new joint loss function and the optimized algorithm AdamW are used to train and test LCNNE in the wear particle image dataset.

The model learning rate is 0.001, the number of iterations is 1500, and the minibatch size is 128. To verify the advantages of the integrated model, comparative experiments are carried out. The comparative experimental objects include a self-built deep convolutional neural network (DCNN) model using dataset pretraining, a single pre-trained model combining VGG19 and GoogLeNet, and convolutional neural network ensemble model based on the commonly used cross-entropy loss function (CNNE). The model structure is the same as LCNNE. The three-model structure of DCNN, VGG19, and GoogLeNet is shown in Figure 7. The experiments were carried out using the TensorFlow toolbox of Google [29]. By setting the control model DCNN, using the image dataset of wear particles to conduct pretraining directly without knowledge transfer learning, to compare the performance of the proposed LCNNE transfer model with that of the conventional DCNN, the four control models were trained by the optimization algorithm Adam [30].

In fault diagnosis, the collected dataset can be balanced [31] or unbalanced [32], and the balanced training dataset leads to differences in evaluation methods. As shown in Table 1, the dataset in this study is completely balanced, which indicates that accuracy is still an appropriate method to evaluate the algorithm in the subsequent experiments. Since the parameters of the neural network are randomly initialized, to verify the stability of the model, each experiment was repeated 10 times, and the results are shown in Figure 8.

## 4. Results and Discussion

**4.1. Model as Classifier.** Here, the proposed LCNNE model and four sets of comparison models are directly used as classifiers, to compare the diagnostic performance of wear particle image datasets. Meanwhile, to prove the superiority and generalization ability of the proposed fine-tuned LCNNE model for small samples, the diagnostic

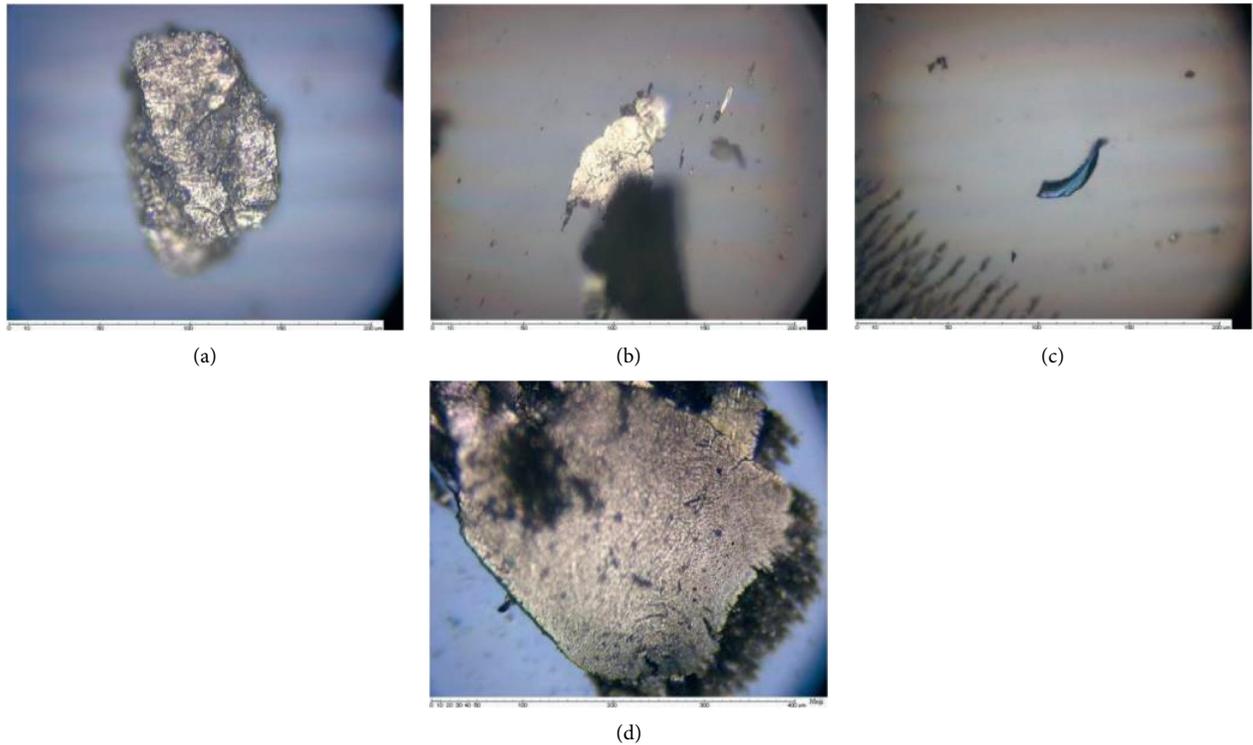


FIGURE 5: Image examples of wear particles.

performance under different training wear particle samples is tested. The total number of training samples is set to 100, 150, 300, 1000, 3000, 5000, 8000, 10,000, and 10,800, respectively. Since the parameters of the neural network are randomly initialized, to verify the stability of the model, each experiment is repeated 10 times. The experimental results are shown in Figure 8. In terms of accuracy and generalization ability, the LCNNE model is superior to other models. Moreover, the performance of ensemble model CNNE and LCNNE is better than the fine-tuned single pretrained VGG19 and GoogLeNet model, indicating the effectiveness of model ensemble. It also shows that fine-tuned single pretrained model performs better than traditional DCNN, which could be caused by the stronger feature expression ability of pretrained VGG19 and GoogLeNet model than that of the DCNN model.

Figure 8 shows that the recognition rate of models DCNN, VGG19, GoogLeNet, CNNE, and LCNNE increased with the increase of training samples, and the standard deviation of 10 trials gradually decreased. When the number of training samples is 10,800, the recognition rate of wear particle image diagnosis by LCNNE is higher than that of the other four models. When the number of training samples is 100, the recognition rates of DCNN, VGG19, GoogLeNet, CNNE, and LCNNE on the test set are, respectively, 79.25%, 82.5%, 83.5%, 84.13%, and 85.63%, with an error of 5.25%, 4.25%, 3.75%, 3.0%, and 2.62%, respectively, showing that the integration model of LCNNE is superior to that of CNNE. This shows that the new joint loss function and optimization algorithm AdamW can improve the network classification.

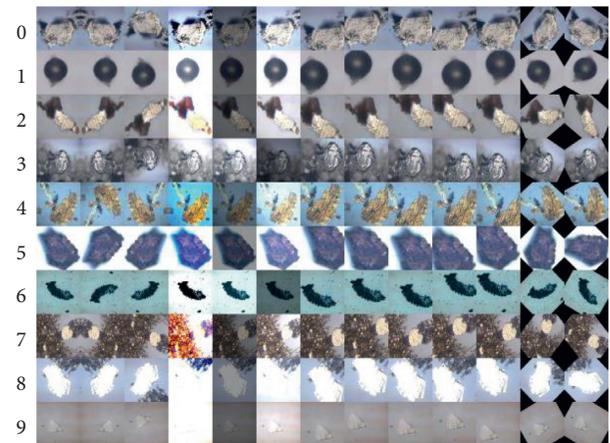


FIGURE 6: Expansion diagram of wear particle image.

In order to further analyze the performance of the model, the precision, recall, recognition accuracy, standard deviation, and detection time of the five models are compared, as shown in Table 2. The ensemble model proposed in this paper is higher than the other four models in terms of precision, recall, recognition accuracy, and stability, and the recognition speed also meets the needs.

**4.2. Fine-Tuned CNN as a Feature Extractor.** Here, LCNNE is used as a feature extractor, and the feature vector with feature dimension of 100 extracted from the last hidden layer is uploaded to SVM classifier based on sklearn library of classification. In terms of accuracy and generalization ability,

```

Input:       $\lambda = 0.001, \lambda_1 = 0.1, \lambda_2 = 0.1, \rho_1 = 0.9, \rho_2 = 0.999, \varepsilon = 10^{-8}$ 
Initialize Training times  $i \leftarrow 0$ , first moment vector  $s_0 \leftarrow 0$ , second moment vector  $v_0 \leftarrow 0$ 
Output:    Optimized parameters  $\theta_t$ 
For        Stopping criterion is met
do
   $i \leftarrow i + 1$ 
   $g \leftarrow 1/n \nabla_{\theta} \sum_{j=1}^n L(f(x^j; \theta))$ 
   $s_i \leftarrow \rho_1 s_{i-1} + (1 - \rho_1)g$ 
   $v_i \leftarrow \rho_2 v_{i-1} + (1 - \rho_2)g^2$ 
   $\hat{s}_i \leftarrow s_i / (1 - \rho_1^i)$ 
   $\hat{v}_i \leftarrow v_i / (1 - \rho_2^i)$ 
   $\theta_i \leftarrow \theta_{i-1} - \lambda \hat{s}_i / (\sqrt{\hat{v}_i} + \varepsilon) - \lambda (\lambda_1 v_i^{i-1} + \lambda_2 w_0^{i-1})$ 
End for

```

ALGORITHM 1: AdamW for LCNNE.



FIGURE 7: Diagram of the transfer model structure of DCNN, VGG19, and GoogLeNet.

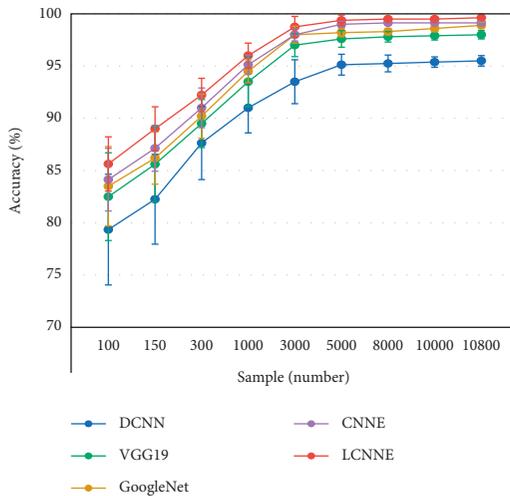


FIGURE 8: Fault diagnosis results of abrasive grains by different models under different numbers of training samples.

the LCNNE model is superior to other models. The last hidden layer feature with a feature dimension of 100 is extracted and input into SVM for classification. To validate the stability of the model, each test is repeated 10 times, and the results are shown in Table 2.

According to Table 3, the proposed LCNNE model as feature extractor, classified by SVM, has a higher recognition rate than DCNN, VGG19, GoogLeNet, and VGG + GoogLeNet model feature extractors, and it has the lowest standard deviation,  $99.63\% \pm 0.26\%$ . This shows that the LCNNE model is the most stable and has the best generalization ability. In addition, the feature extraction ability of the LCNNE model is better than that of the VGG19 and GoogLeNet combined model, which is about 2% to 3%

**4.3. Feature Visualizations.** To better understand the effect of feature extraction for diagnostic results, t-SNE [33] is used to investigate the feature distribution learned by the last

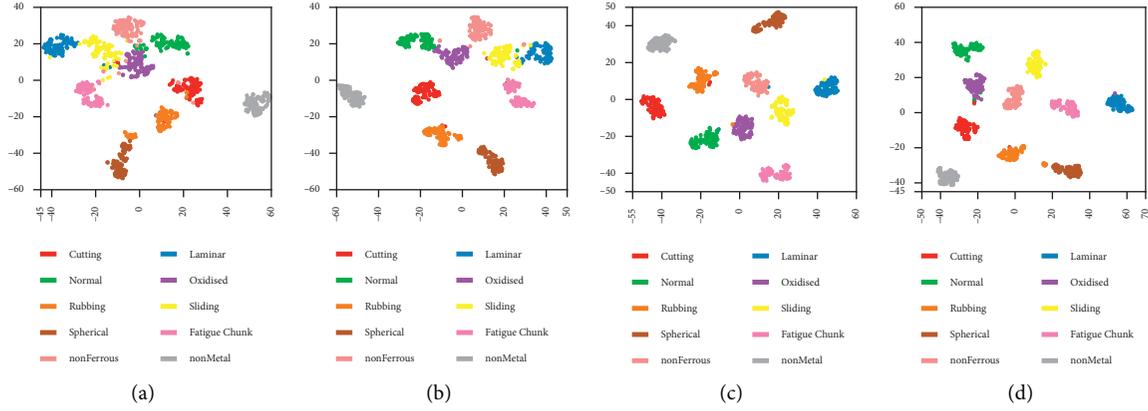


FIGURE 9: Feature visualization via t-SNE. Last hidden fully connected layer representation for the test samples in four models: (a) DCNN model; (b) fine-tuned pretrained VGG19 model; (c) fine-tuned pretrained GoogLeNet model; (d) LCNNE model.

TABLE 1: Description of wear particle image datasets.

Wear particle fault category	Normal	Spherical	Sliding	Fatigue chunk	Laminar	Oxidized	Cutting	Rubbing	Nonferrous	Nonmetal
Category labels	0	1	2	3	4	5	6	7	8	9
Training	1080	1080	1080	1080	1080	1080	1080	1080	1080	1080
Test	80	80	80	80	80	80	80	80	80	80

TABLE 2: Performance comparison of five models.

	DCNN	VGG19	GoogLeNet	CNNE	LCNNE
Precision	0.860	0.873	0.886	0.897	0.913
Recall	0.845	0.863	0.877	0.891	0.910
Accuracy	95.63%	98.25%	98.94%	99.13%	99.50%
Standard deviation	0.5%	0.38%	0.35%	0.25%	0.13%
Time (s)	0.127	0.112	0.106	0.124	0.115

TABLE 3: Classification accuracy when fine-tuned CNN is used for feature extraction for further input to support vector machine (SVM) classification.

	DCNN + SVM	VGG19 + SVM	GoogLeNet + SVM	(VGG19 + GoogLeNet) + SVM	LCNNE + SVM
1 Accuracy	94.63%	97.35%	99.56%	96.30%	99.26%
2 Accuracy	96.57%	98.95%	99.68%	96.80%	99.76%
3 Accuracy	97.18%	99.13%	97.45%	96.77%	99.42%
4 Accuracy	93.35%	97.62%	97.99%	95.61%	99.74%
5 Accuracy	93.69%	98.79%	97.70%	94.66%	99.20%
6 Accuracy	96.32%	99.70%	99.39%	97.63%	99.82%
7 Accuracy	95.51%	99.81%	99.94%	96.97%	99.96%
8 Accuracy	95.39%	99.16%	99.34%	96.41%	99.66%
9 Accuracy	98.14%	96.34%	97.66%	96.78%	99.54%
10 Accuracy	96.76%	99.49%	99.14%	95.83%	99.90%
Mean accuracy	95.75%	98.63%	99%	96.38%	99.63%
Standard deviation	1.54%	1.15%	0.97%	0.83%	0.26%

hidden fully connected layer representation for the test samples in four models trained by 10,800 samples. Figure 9 presents some interesting phenomena worth noting. Firstly, it can be noticed that the nonmetal wear particles can be divided into four categories, suggesting that nonmetal particle features are easy to distinguish. This is consistent

with the subjective impression that nonmetallic abrasive particles are better differentiated than other abrasive particles. Secondly, the graphs in Figures 9(a)–9(d), respectively, correspond to the visualization of features of the last hidden layer of DCNN, VGG19, GoogLeNet, and LCNNE models, and their separability gets better. This is also directly

reflected in the final accuracy of testing the wear particle dataset, which makes the diagnostic recognition rate of the LCNNE model proposed in this work higher than that of the other three models.

## 5. Conclusions

This paper proposes an ensemble transfer learning model based on the new joint loss function, to address the problem of abrasive image recognition. The model directly diagnoses fault wear particles and thus avoids excessive time-consuming preprocessing. The LCNNE model presents two main advantages: First, it does not require too much data in the migration domain to achieve better diagnostic results, making it ideal for widespread application in fault diagnosis. Second, the model avoids sacrificing diagnostics. Combining multiple models improves the diagnostic accuracy. On the wear particle dataset, the model achieves the accuracy rate of 99.63%.

Results shown in Section 5 indicate that, compared to the single-model migration learning and the latest popular CNN structure, which yielded better diagnostic accuracy on the wear particle dataset, training LCNNE with the optimization algorithm AdamW achieved nearly 100% diagnosis accuracy rate under limited small-scale datasets, making it an effective approach to address the problem of insufficient data samples in engineering applications. It also shows good application prospects in the machine learning technology. Compared with a single-network transfer learning model, it has higher computational efficiency for wear particle image recognition and better performance than the training model of DCNN from scratch. Its extracted features are also more separable.

As mentioned in Section 3.1, the wear particle dataset used in this paper is completely balanced. In contrast, it is possible to encounter unbalanced datasets in actual practice. Therefore, in future work, we would investigate the performance of ensemble transfer model VGNet on an unbalanced dataset, to expand the range of applications for the algorithm.

## Data Availability

After the distance matrix and the modified Gini index, the improved D-S evidence theory is combined with the deep convolution neural network to form a novel bearing fault diagnosis model IDSCNN. The open bearing data set of Case Western Reserve University is used to verify the model, and it is concluded that the performance of the model is better than the existing machine learning methods.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Acknowledgments

The authors thank the National Natural Science Foundation of China (no. 51975324) and Hubei Key Laboratory Open Fund (nos. 2018KJX10 and 2018KJX03) for supporting this research.

## References

- [1] Z. Liu, Q. Li, and J. Li, "Force analysis of the gear rack of the Three Gorges shiplift," *Mechanical Transmission*, vol. 38, no. 10, pp. 54–57+66, 2014.
- [2] D. Fan, X. Cui, and Y. Cao, "The force analysis of the rack and pinion lifting system of the jack-up platform," *Petroleum Field Machinery*, vol. 33, no. 12, pp. 27–30, 2010.
- [3] V. Macián, R. Payri, B. Tormos, and L. Montoro, "Applying analytical ferrography as a technique to detect failures in diesel engine fuel injection systems," *Wear*, vol. 260, pp. 562–566, 2006.
- [4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [5] A. Brunetti, D. Buongiorno, G. F. Trotta, and V. Bevilacqua, "Computer vision and deep learning techniques for pedestrian detection and tracking: a survey," *Neurocomputing*, vol. 300, pp. 17–33, 2018.
- [6] L. Deng, J. Li, J. T. Huang et al., "Recent advances in deep learning for speech research at Microsoft," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8604–8608, Vancouver, BC, Canada, May 2013.
- [7] G. Hinton, L. Deng, D. Yu et al., "Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, 2012.
- [8] W. Pan and L. Torresani, "Unsupervised hierarchical modeling of locomotion style," in *Proceedings of the International Conference on Machine Learning (ICML)*, pp. 785–792, Montreal, QC, Canada, June 2009.
- [9] R. Boostani, F. Karimzadeh, and M. Nami, "A comparative review on sleep stage classification methods in patients and healthy individuals," *Computer Methods and Programs in Biomedicine*, vol. 140, pp. 77–91, 2017.
- [10] C. Lu, Z.-Y. Wang, W.-L. Qin, and J. Ma, "Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification," *Signal Processing*, vol. 130, pp. 377–388, 2017.
- [11] O. Janssens, V. Slavkovikj, B. Vervisch et al., "Convolutional neural network based fault detection for rotating machinery," *Journal of Sound and Vibration*, vol. 377, pp. 331–345, 2016.
- [12] O. Korzh, M. Joaristi, and E. Serra, "Convolutional neural network ensemble fine-tuning for extended transfer learning," *International Conference on Big Data*, Springer, Berlin, Germany, 2018.
- [13] H. Fan, G. S. Xia, H. Jingwen, and L. Zhang, "Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery," *Remote Sensing*, vol. 7, Article ID 14680, 2015.
- [14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proceedings of the NIPS 2012: Neural Information Processing Systems*, Lake Tahoe, NV, USA, December 2012.
- [15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, <https://arxiv.org/abs/1409.1556>.
- [16] C. Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, June 2014.
- [17] I. Loshchilov and F. Hutter, "Fixing weight decay regularization in Adam," 2017, [https://www.researchgate.net/publication/339370196\\_Multimodal\\_Encoder-Decoder\\_Attention\\_Networks\\_for\\_Visual\\_Question\\_Answering](https://www.researchgate.net/publication/339370196_Multimodal_Encoder-Decoder_Attention_Networks_for_Visual_Question_Answering).

- [18] B. Harangi, "Skin lesion classification with ensembles of deep convolutional neural networks," *Journal of Biomedical Informatics*, vol. 86, pp. 25–32, 2018.
- [19] T. G. Dietterich, "Ensemble methods in machine learning," in *International Workshop on Multiple Classifier Systems*, Springer, Berlin, Germany, 2000.
- [20] J. Gao, W. Fan, J. Jiang, and J. Han, "Knowledge transfer via multiple model local structure mapping," in *Proceedings of the 14th ACM SIGKDD Int'l conf. Knowledge Discovery and Data Mining*, pp. 283–293, Las Vegas, NV, USA, August 2008.
- [21] L. Nannia, S. Ghidoni, and S. Brahnham, "Ensemble of convolutional neural networks for bioimage classification," *Applied Computing and Informatics*, vol. 17, no. 1, pp. 19–35, 2021.
- [22] Z. Guo, Q. Chen, G. Wu, Y. Xu, R. Shibasaki, and X. Shao, "Village building identification based on ensemble convolutional neural networks," *Sensors*, vol. 17, no. 11, 2017.
- [23] S. Li, G. Liu, X. Tang, J. Lu, and J. Hu, "An ensemble deep convolutional neural network model with improved D-S evidence fusion for bearing fault diagnosis," *Sensors*, vol. 17, no. 8, 2017.
- [24] A. Kumar, J. Kim, D. Lyndon, M. Fulham, and D. Feng, "An ensemble of fine-tuned convolutional neural networks for medical image classification," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 1, pp. 31–40, 2017.
- [25] C. Ding and D. Tao, "Trunk-branch ensemble convolutional neural networks for video-based face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 1002–1014, 2016.
- [26] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.
- [27] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and F. F. Li, "ImageNet: a large-scale hierarchical image database," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 248–255, Providence, RI, USA, June 2009.
- [28] Z. F. Wu, H. F. Zuo, and Z. Yang, "Quantitative description system of micromorphological characteristics of wear particles," *Journal of Transportation Engineering*, vol. 1, no. 1, pp. 115–119, 2001.
- [29] TensorFlow, <http://www.tensorflow.org>, 2018.
- [30] D. Kingma and J. Ba, "Adam: a method for stochastic optimization," 2014, <https://arxiv.org/abs/1412.6980>.
- [31] L. F. Villa, A. Reñones, J. R. Perán, and L. J. D. Miguel, "Angular resampling for vibration analysis in wind turbines under non-linear speed fluctuation," *Mechanical Systems and Signal Processing*, vol. 25, no. 6, pp. 2157–2168, 2011.
- [32] X. Zhang, Y. Zhuang, W. Wang, and W. Pedrycz, "Transfer boosting with synthetic instances for class imbalanced object recognition," *IEEE Transactions on Cybernetics*, vol. 48, no. 1, pp. 357–370, 2018.
- [33] L. Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.