A Generic Intelligent Bearing Fault Diagnosis System Using Convolutional Neural Networks With Transfer Learning

TAO LU\textsuperscript{1}, FANQIANHUI YU\textsuperscript{2}, BAOKUN HAN\textsuperscript{1}, AND JINRUI WANG\textsuperscript{1,1}

\textsuperscript{1}College of Mechanical and Electronic Engineering, Shandong University of Science and Technology, Qingdao 266000, China
\textsuperscript{2}College of Food Science and Engineering, Ocean University of China, Qingdao 266003, China

Corresponding authors: Fanqianhui Yu (yufanqianhui@163.com) and Baokun Han (bk_han@163.com)

This work was supported in part by the China Scholarship Council under Grant 201806330047 and Grant 201808370191, and in part by the Acoustical Society of America International Student Grant (2018).

\textbf{ABSTRACT} It is very important and necessary to diagnose bearing faults timely, quickly, and accurately in practical applications, because the operation status of the bearings is directly related to the performance and reliability of the rotating machinery. Therefore, a generic intelligent bearing fault diagnosis system based on AlexNet with transfer learning was proposed to automatically identify and classify different bearing faults. Transfer learning was used to avoid overfitting problem of deep network. Five bearing faults at two different motor loads and speeds were selected from the Case Western Reserve University (CWRU) bearing dataset to validate the performance of the proposed method. Results showed that compared to previous methods, the proposed method achieved excellent performance, with overall classification accuracy over 99.7\%, and fast training and testing times. Feature visualization displayed the common and high-level features of spectrograms of vibration signals learned by the trained classification model. And strongest activations demonstrated the classification model had learned the correct features of each bearing fault. Importantly, there were two testing datasets employed in this study, where the training dataset and testing dataset (2) were completely independent and the number of sub-spectrograms used for testing was 3-5 times greater than those used for training. Thus, all the results suggest that the proposed method is stable, reliable, suitable for diagnosing different bearing faults and has a great potential in practical applications.

\textbf{INDEX TERMS} AlexNet, bearing fault, feature visualization, strongest activations, transfer learning.

\section*{I. INTRODUCTION}
Rotating machinery is one of the widely used types of machines in modern industry and its reliability and safe operation are critical to the mechanical systems [1]. Rolling element bearings are the most important components of rotating machinery, and their operation status (e.g., loss of surface integrity, cracking, or peeling) is directly related to the performance and reliability of the machinery in which they are installed [2]–[6]. This is because bearings often work in extreme or harsh environments, such as high speed, high temperature, and high pressure, which leads to various unexpected faults and damages to the bearing systems, and even serious accidents and additional economic losses [7], [8]. Thus, it is really necessary to diagnose bearing faults of the rotating machinery timely, quickly, and accurately.

The status monitoring of rolling element bearings can be seen as a pattern recognition task that has been successfully addressed by bearing fault diagnostic methods [9]. Generally, a diagnostic approach consists of the following four steps: data acquisition, feature extraction, feature selection, and feature classification [10]. Feature extraction is a critical step in the past pattern recognition task, as it maps the original signals to statistical parameters that convey the information about the condition of the machine [11]. As a result, feature extractors are often designed in conventional diagnostic methods based on machine learning to obtain highly accurate recognition results [12]. However, manual design of feature extractors not only requires experts with signal processing techniques, but also is a time-consuming process since feature extraction does not have a common procedure for every
task [13]. Moreover, the shallow structure of conventional machine learning algorithms has very limited ability to learn non-linear relation for extracting features [14]. Therefore, based on these drawbacks, deep learning algorithms have been introduced in bearing fault diagnosis systems in recent years to make them more intelligent.

Deep learning is a branch of machine learning where multiple layers of data processing units are assembled into a deep architecture to extract multiple levels of data abstraction, and each layer automatically learns higher-level of data representations from the output of the previous layer [9]. With the significant advantage of automatically learning data representations, it is considered as an advanced technique in big data analysis [15]. Therefore, deep learning-based intelligent diagnostic methods are more flexible and capable to deal with difficulties in real-world applications than traditional machine learning-based methods [16]. Convolutional Neural Network (CNN), one of the popular deep learning models, are specially designed to handle variable and complex signals [3]. CNN started being used to identify bearing fault in 2016 and it has been demonstrated to be an efficient method that outperforms other traditional methods [17]. For instance, it can identify some early-stage faulty conditions without explicit characteristics frequencies (e.g. lubrication degradation) and can help to remove noise from the vibration signals [3], [16].

In general, CNN as a deep network requires large dataset to train the network, such as AlexNet is exploited on ImageNet, which contains approximately 50 million images [18]. However, it is too difficult to obtain such a large amount of training data in practice, especially for bearing fault data, since the amount of data from faulty bearings is much less than that from normal bearings. Besides, the length of each bearing fault data in CWRU dataset applied in this study is only about 10 seconds. The small dataset makes it unsuitable for training deep neural networks, as it can easily cause overfitting problem of the network [19]. Therefore, previous studies have used many different data augmentation methods to solve this problem, such as using a synthetic over-sampling method to generate new samples by interpolating real samples, transforming the available data to generate additional data variants, or using different signal processing techniques to add Gaussian noise, signal translation, amplitude shifting, and time stretching [20], [21]. While the results suggest that deep learning-based fault diagnosis methods can largely benefit from an expanded dataset with more generated labeled instances, but the new samples are generally similar to the real samples and lack sufficient diversity of data, so improvement in model generalization are limited [21]–[23]. In addition, generative adversarial networks (GANs) have been recently used to learn the distributions of the machinery vibration data and generate additional realistic fake samples to expand the training dataset, despite the improved testing performance, the main drawback of GANs-based bearing fault diagnostics is its reliance on a relatively large model consisting of multiple GANs for the minority classes [24]. Instead of using data augmentation methods described above, transfer learning was employed in this study. It is a very effective and convenient method to train deep neural networks, especially when there are not enough labeled images [25]–[27].

Based on the above, in order to diagnose bearing faults in rotating machinery timely, quickly, and accurately, a new CNN-based generic real-time intelligent bearing fault diagnosis system was developed in this study. Specifically, a method of AlexNet with transfer learning was proposed to automatically identify and classify 5 different bearing faults from the Case Western Reserve University (CWRU) dataset. The purpose of using transfer learning was to avoid the overfitting problem of the deep network since the bearing fault data was limited. Moreover, confusion matrix, training and testing time, comparison with previous methods, feature visualization, and strongest activations were investigated to validate the performance of the proposed method. The obtained results show that the proposed method achieves excellent performance and has great potential to be applied in the fault diagnosis of many other mechanical and electrical components, such as gears, pumps, compressors, and electric motors.

II. THEORETICAL BACKGROUND OF THE PROPOSED NETWORK

A. AlexNet

AlexNet, one of the leading CNN architectures, is designed by Krizhevsky in 2012 [28] and has proven to achieve superior performance over previous approaches in image classification tasks [29]. It is a complex and large neural network with 60 million parameters and 650,000 neurons, consisting of 5 convolutional layers, some of which are followed by max-pooling layers, and 3 fully connected layers (FCL) with final 1000-way softmax [30]. The network structure of AlexNet is shown in Table 1, and details of Krizhevsky’s improvements in the training of these parameters are as follows [25].

Firstly, in order to solve the problem that traditional activation function (including logistic function, tanh function, and arctan function) often trapped in gradient vanishing of deep networks, Krizhevsky used a new activation function as a Rectified Linear Unit (ReLU). The definition of ReLU is

\[
ReLU(x) = \max(x, 0)
\]

Secondly, the use of dropout in fully connected layers was another improvement to avoid overfitting, reduce the joint adaptation between neurons, and improve the generalization and robustness. Meanwhile, the convolution can automatically learn features from the training images and reduce the complexity of the network by parameter sharing. The definition of convolution is:

\[
C(m, n) = (M \ast w)(m, n) = \sum_k \sum_l M(m-k, n-l)w(k, l)
\]
Moreover, pooling was employed to feature reduction. Pooling considers a group of neighboring pixels in the feature map and generates a value for representation by some strategy. Last, cross channel as local normalization method is inspired by real neurons and it generates sum from adjacent maps at the same position, which means the feature maps are normalized before feeding into the next layers of the network. The neurons in adjacent fully connected layers are directly linked, and the softmax activation function activates neurons by constraining the output in the range of (0,1). The softmax activation function can be defined as:

$$\text{soft} \max (X) _i = \frac{\exp(x_i)}{\sum_{j=1}^{n} \exp(x_j)}$$  \hspace{1cm} (3)\\

**TABLE 1.** The network structure of AlexNet.

| Layers | Name | Layers | Name |
|--------|------|--------|------|
| 1      | Image input layer | 14 | Grouped convolution layer |
| 2      | Convolution layer | 15 | ReLU layer |
| 3      | ReLU layer | 16 | Max pooling layer |
| 4      | Cross channel normalization layer | 17 | Fully connected layer |
| 5      | Max pooling layer | 18 | ReLU layer |
| 6      | Grouped convolution layer | 19 | Dropout layer |
| 7      | ReLU layer | 20 | Fully connected layer |
| 8      | Cross channel normalization layer | 21 | ReLU layer |
| 9      | Max pooling layer | 22 | Dropout layer |
| 10     | Convolution layer | 23 | Fully connected layer |
| 11     | ReLU layer | 24 | Softmax layer |
| 12     | Grouped convolution layer | 25 | classification output layer |

**B. TRANSFER LEARNING**

Transfer learning is a deep learning approach in which a model that has been trained for one task is used as a starting point to train a model for another task. With its help, the entire structure of the proposed network is divided into the pre-trained network and the transferred network. The parameters in the pre-trained network have been trained with more readily available images from different domains, such as large-scale image databases like ImageNet with millions of images [18]. Meanwhile, the layers and parameters of the pre-trained network are used as feature extractors to extract features and make the training converge. Since the parameters in the transferred network are a small fraction of all the parameters in the entire network, even small training data can meet the requirements of deep learning with transfer learning. Moreover, transfer learning can help reduce the dependence of the deep networks on computer hardware and training time. Therefore, the whole training process can be completed from a few minutes to a few hours by transfer learning, which effectively avoids the time-consuming training and computation required by CNN-based methods. The schematic of AlexNet with transfer learning used in this study is shown in Fig. 1. The cross-domain transfer learning was first pre-trained on source domain (ImageNet) and then applied on target domain (CWRU bearing dataset), the two domains do not have any similar properties.

**FIGURE 1.** The schematic of AlexNet with transfer learning.

**III. METHODS**

**A. DATA PRE-PROCESSING**

Perform the following processing steps for the vibration signals in the training and testing dataset.

1. Generate spectrogram (RGB image) by the Non-Uniform Fast Fourier Transform (NFFT) with Hamming window (a segment length of $t_d = fs/100$, where $fs$ is sampling rate of the vibration signal, and the percent overlap is 80% between adjacent frames).

2. Cut the spectrograms into several consecutive sub-spectrograms (cut in the order of the columns). Each obtained sub-spectrogram is 5 pixels in the length and corresponds to a vibration signal duration of 10 ms, and no overlapping between adjacent sub-spectrograms.

3. The size of each sub-spectrogram becomes $227 \times 227 \times 3$ pixels by resizing to meet the image input size of AlexNet.

**B. CLASSIFICATION MODEL**

A multiclass task included five different bearing faults, the classification unit was the obtained sub-spectrogram and the classification model was based on AlexNet with transfer learning. As shown in Table 2, we replaced the last nine layers of the pre-trained AlexNet, where 5 nodes were used to replace the 1000 nodes in the last FCL of the pre-trained AlexNet, and the rest parameters of the pre-trained AlexNet were preserved and served as the initialization. In the classification model, the transferred part of AlexNet was trained by Adaptive Moment Estimation (ADAM). The initial learning rate was 0.0003, the learn rate drop factor and period was 0.1 and 10, respectively, the minibatch size was set as 64, and the max epochs was set as 15.

**C. FEATURE VISUALIZATION AND STRONGEST ACTIVATIONS**

Since convolutional neural networks automatically learn features from training dataset during the training process, feature visualization is used for interpreting to humans how
TABLE 2. Revision of the AlexNet in the classification model.

| Layers | Original                                | Replaced                                |
|--------|-----------------------------------------|-----------------------------------------|
| 17 FCL (4096) with pre-trained weights and biases | FCL (4096) with random initialization    |
| 18 ReLu layer                                    | ReLu layer                              |
| 19 dropout layer (Probability = 0.5)            | dropout layer (Probability = 0.5)       |
| 20 FCL (4096) with pre-trained weights and biases | FCL (4096) with random initialization    |
| 21 ReLu layer                                    | ReLu layer                              |
| 22 dropout layer (Probability = 0.5)            | dropout layer (Probability = 0.5)       |
| 23 FCL (1000) with pre-trained weights and biases | FCL (5) with pre-trained weights and biases |
| 24 Softmax layer                                 | Softmax layer                           |
| 25 Classification layer (1000)                  | Classification layer (5 classes)        |

convolutional neural networks build an understanding of images or answering the questions about what a network looking for by generating examples [31]. Strongest activations is used to examine whether the trained network has learned the correct features. Detailed methods for feature visualization and strongest activations were described in our previous published paper [30].

IV. A CASE OF BEARING FAULTS CLASSIFICATION

This section first introduces the experimental setup of CWRU.

Then, the performance of AlexNet with transfer learning-based intelligent bearing fault diagnosis system is extensively evaluated by using CWRU bearing dataset.

Since the natural degradation of bearings is a gradual process that can take many years, most people conduct experiments and collect data by either artificially inducing bearing faults or by using accelerated life testing methods, but data collection is still time-consuming [32]. Fortunately, there are a number of organizations that have shared their bearing fault datasets for study, such as CWRU dataset [33], Intelligent Maintenance Systems (IMS) dataset [34], Paderborn university bearing dataset [35], and Pronostia bearings accelerated life test dataset [36]. Especially since the CWRU dataset has been widely used as a benchmark dataset [37], this study used motor bearing vibration signals of the CWRU dataset to test and verify the performance of the proposed diagnosis system. The test rig of CWRU consists of a 2-hp Reliance Electric motor, a torque transducer/encoder, and accelerometers (Fig. 2). Drive end bearing type of the motor is SKF deep groove ball bearing (6205-2RS JEM). Three kinds of single point defects are manufactured to the testing bearings and reinstalled into the test rig. Motor run at a constant speed (approximately 1720–1797 rpm) with different loads (0 to 3 hp) provided by the dynamometer. The accelerometers vertically attached to the housing are employed for collect vibration data and the torque transducer/encoder is employed for collect speed and horsepower data. Further details can be found at the CWRU Bearing Data Center website [33], [37].

FIGURE 2. Apparatus for the bearing vibration signal collection of the CWRU bearing dataset [30].

The drive end bearing fault data of CWRU bearing data center includes three fault types: inner race (IR), outer race (OR), and rolling element (ball). The drive end bearing specifications are listed in Table 3. The sampling rate of these data is 12 kHz. The bearing fault frequency includes the ball pass frequency of the outer race (BPFO), ball pass frequency of the inner race (BPFI), fundamental train frequency (FTF), and ball spin frequency (BSF). Among them, the fault characteristic frequency in the different bearing fault signals is the product of these frequencies and the main frequency, which is proportional to the running speed. The fault characteristic frequency of the IR, OR, cage train, and rolling element are listed in Table 4.

TABLE 3. Drive end bearing specifications (the bearing geometry is deep groove ball bearing and the size is inches).

| Inside diameter | Outside diameter | thickness | Ball diameter | Pitch diameter |
|-----------------|------------------|----------|---------------|---------------|
| 0.9843          | 2.0472           | 0.5906   | 0.3126        | 1.537         |

As shown in Table 5, five kinds of bearing faults (fault diameter of 0.007 inches) at two different motor loads and speeds were selected from the CWRU dataset for training and testing. Specifically, 1/4 of each fault type was first (first row: IR007_0, B007_0, OR007@6_0, OR007@3_0, and OR007@12_0) was used for training and validation, where 200 sub-spectrograms were used as the training dataset and 49 sub-spectrograms were used as the validation dataset. Then, the remaining (3/4) of each fault type (749 sub-spectrograms) in the first row was employed as the testing dataset (1). Besides, all five fault types in the second row (IR007_1, B007_1, OR007@6_1, OR007@3_1 and OR007@12_1), 999 sub-spectrograms of each fault type, were employed as the testing dataset (2). Both testing dataset (1) and (2) were employed to test the performance of the proposed intelligent bearing fault diagnosis system.
TABLE 5. Information on the data selected from the CRWU dataset for training, validation, and testing.

| Motor load (HP) | Approx. inner motor speed (rpm) | Inner Ball race | Outer race-Position relative to load zone (load zone centered at 6:00) |
|----------------|---------------------------------|----------------|-------------------------------------------------|
|                |                                 |                | Centered @6:00 | Orthogonal @3:00 | Opposite @12:00 |
| 0              | 1797                            | B007           | OR007@         | OR007@           | OR007@          |
| 1              | 1772                            | B007           | OR007@         | OR007@           | OR007@          |
| 2              | 1750                            | B007           | OR007@         | OR007@           | OR007@          |
| 3              | 1730                            | B007           | OR007@         | OR007@           | OR007@          |

According to the method of **A. Data pre-processing**, their corresponding spectrograms are shown in Fig. 3.

The proposed diagnosis system was developed on a workstation with Intel Xeon Gold 5120 CPU*2 and NVIDIA Quadro P2000 GPU. The code was written by using Matlab R2019a.

V. RESULTS AND DISCUSSION

**A. PERFORMANCE OF THE PROPOSED DIAGNOSIS SYSTEM**

The confusion matrix of the testing datasets (Fig. 4) was used to evaluate the performance of the trained classification model in determining the different bearing faults [30]. For instance, as shown in the first row (actual horizontal) in confusion matrix for testing dataset (1), a total of 749 sub-spectrograms were used to actual classify IR007, where 749 sub-spectrograms were accurately classified and 0 sub-spectrograms were misclassified as B007, therefore the recall (sensitivity) of IR007 was 99.7%. Based on this, the overall accuracy of the testing datasets (1) and (2) was 99.9% and 99.7%, respectively. Moreover, a comparison with other bearing fault diagnostics method based on the CWRU dataset is shown in Table 6 [38]–[41]. The results show that our method achieves the highest classification accuracies, which indicates that it is suitable for bearing fault diagnosis. Importantly, in this study, the training dataset and the testing dataset (2) were completely independent, and the number of sub-spectrograms used for testing was 3-5 times greater than those used for training. Meanwhile, although the five bearing faults in the testing datasets (1) and (2) were sourced from different motor loads and speeds, the extremely high classification accuracy indicates the proposed method can diagnose the correct type of the faults and avoid the problems of load and speed fluctuations that affect fault classification accuracy in practical applications. Therefore, these strongly demonstrates the proposed method is stable, reliable and can be used in practical applications. Furthermore, the training time of the proposed method was 88.20 s, and the testing times for the testing datasets (1) and (2) were 7.39 s and 10.60 s, respectively.

Moreover, an anonymous reviewer suggested us to test more sets of faults samples at different motor loads in order to evaluate generalizability of the proposed method. Therefore, we added testing dataset (3) and (4), corresponding to 2 HP and 3 HP motor loads in Table 5, respectively. Their confusion matrix for testing datasets are shown in Fig.5, and the overall accuracy of testing datasets (3) and (4) was 99.5% and 98.2%, respectively. In conclusion, all the above results suggest that the proposed method achieves excellent performance in diagnosing different bearing faults.

**B. FEATURE VISUALIZATION**

In the previous studies [42]–[44], feature visualization is typically generated by the t-distributed stochastic neighbor embedding (t-SNE) technique, which is used to display the


### TABLE 6. Comparison of classification accuracy with different methods from the literature.

| Classifier                          | Feature                                      | Fault classification accuracy |
|-------------------------------------|----------------------------------------------|------------------------------|
| SVM [2]                             | Statistical Locally Linear Embedding          | 77.8% - 94.1%               |
| DNN with temporal coherence [38]    | features directly from the training data     | 94.4%                        |
| Compact 1D CNN (with no fine tuning) [39] | features directly from the training data     | 93.2%                        |
| Improved deep recurrent neural network [40] | features directly from the training data     | 94.75% (1750 rpm), 96.53% (1797 rpm) |
| AlexNet with transfer learning in this study | features directly from the training data     | 99.9% (1797 rpm), 99.7% (1772 rpm) |

---

C. STRONGEST ACTIVATIONS

As shown in Fig. 8, a sample sub-spectrogram of each bearing fault was randomly selected from the testing dataset (2) and then fed into the trained classification model to display the strongest activations of the last convolution layer (conv 5).

Due to the trained classification model used features that automatically learned from the training dataset to classify the convolutional layers and fully connected layer. However, unlike the feature visualization generated by the t-SNE technique, the feature visualization in this study (Fig. 7) is used for interpreting to humans how convolutional neural network build an understanding of spectrograms of fault signals in the training dataset. Specifically, Fig. 7 is the feature visualization of the last fully connected layer (FCL 3) of the trained classification model. The last fully connected layer (FCL 3) of the trained classification model can generate the feature images that most closely resembles each class, representing their common and high-level features [30]. Because the convolutional layers toward the beginning of the trained classification model have only a small size of the receptive field and just learn small, low-level features from the training dataset, while the deeper layers can learn high-level combinations of features learned by the earlier layers [45].

---

Fig. 6 is cited as an example from [42], this feature visualization means the features or data is divisible or not divisible in the mapping of high-dimensional features into a 2-D space. Fig. 6 is cited as an example from [42], this feature visualization means the features or data is divisible or not divisible in the mapping of high-dimensional features into a 2-D space.
sub-spectrograms in the testing dataset, so the purpose of this part of the work was to verify that the trained classification model learned the correct features by comparing the activation areas in the strongest activations images with the corresponding sub-spectrograms. In the strongest activations images (Fig. 8), white pixels represent strong positive activation and black pixels represent strong negative activation [46]. It was obvious that the white areas in the strongest activations identified by the trained classification model, in agreement with the yellow areas (target areas) shown in the corresponding sub-spectrograms and the same as we expected. These results indicated that the trained classification model had learned the correct features of each bearing fault and further demonstrated that the proposed method was effective in diagnosing bearing faults.

VI. CONCLUSION

In this study, a generic intelligent bearing fault diagnosis system based on AlexNet with transfer learning was proposed to automatically identify and classify different bearing faults. Five bearing faults at two different motor loads and speeds from the CWRU bearing dataset were employed to validate the performance of the proposed method. The selected data were first pre-processed into spectrograms and then cut into sub-spectrograms for training and testing. Compared to the previous methods, the proposed methods achieved excellent performance with the following advantages: 1) The use of cross-domain transfer learning successfully avoided the overfitting problem of the deep network when the training dataset was limited. 2) The high overall accuracy of the testing dataset (1) and (2) was 99.9% and 99.7%, respectively. 3) The fast training time was 88.20 s, and the fast testing times for the testing dataset (1) and (2) were 7.39 s and 10.60 s, respectively. Also, the time resolution of the trained model reached 10 ms. 4) Feature visualizations showed the common and high-level features of sub-spectrograms of vibration signals that learned by the trained classification model. 5) Strongest activations of the trained network demonstrated the proposed method had learned the correct features of each fault. In summary, all the results indicate that the proposed method is effective and feasible and has good prospects for application in bearing fault diagnosis.

DECLARATION OF INTERESTS

The authors declare no conflict of interest.

REFERENCES

[1] A. Heng, S. Zhang, A. C. C. Tan, and J. Mathew, “Rotating machinery prognostics: State of the art, challenges and opportunities,” Mech. Syst. Signal Process., vol. 23, no. 3, pp. 724–739, Apr. 2009.
[2] W. A. Smith and R. B. Randall, “Rolling element bearing diagnostics using the case western reserve university data: A benchmark study,” Mech. Syst. Signal Process., vol. 66–65, pp. 100–131, Dec. 2015.
[3] X. Guo, L. Chen, and C. Shen, “Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis,” Measurement, vol. 93, pp. 490–502, Nov. 2016.
[4] M. He and D. He, “Deep learning based approach for bearing fault diagnosis,” IEEE Trans. Ind. Appl., vol. 53, no. 3, pp. 3057–3065, May 2017.
[5] 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 Process., vol. 130, pp. 377–388, Jan. 2017.
[6] R. D. Evans, “Wear of bearings,” in Encyclopedia Tribology, Q. J. Wang and Y.-W. Chung, Eds. Boston, MA, USA: Springer, 2013, pp. 4061–4066.
[7] M. Farajzadeh-Zanjan, R. Razavi-Far, M. Saif, J. Zarei, and V. Palade, “Diagnosis of bearing defects in induction motors by fuzzy-neighborhood density-based clustering,” in Proc. IEEE 14th Int. Conf. Mach. Learn. Appl. (ICMLA), Dec. 2015, pp. 935–940.
[8] G. Wang, F. Zhang, B. Cheng, and F. Fang, “DAIMER: A novel diagnosis aggregation method with evidential reasoning rule for bearing fault diagnosis,” J. Intell. Manuf., pp. 1–20, Mar. 2020.
[9] D.-T. Huang and H.-L. Kang, “A survey on deep learning based bearing fault diagnosis,” Neurocomputing, vol. 335, pp. 327–335, Mar. 2019.
[10] D.-S. Huang, Systematic Theory of Neural Networks for Pattern Recognition, Beijing, China: House of Electronic Industry of China, 1996.
[11] J. Wang, S. Li, Z. An, X. Jiang, W. Qian, and S. Ji, “Batch-normalized deep neural networks for achieving fast intelligent fault diagnosis of machines,” Neurocomputing, vol. 329, pp. 53–65, Feb. 2019.
[12] Z.-Q. Zhao, D.-S. Huang, and B.-Y. Sun, “Human face recognition based on multi-features using neural networks committee,” Pattern Recognit. Lett., vol. 25, no. 12, pp. 1351–1358, Sep. 2004.
[13] C. Li, R.-V. Sanchez, G. Zurita, M. Cerrada, D. Cabrera, and R. E. Vásquez, “Multimodal deep support vector classification with homologous features and its application to gearbox fault diagnosis,” Neurocomputing, vol. 168, pp. 119–127, Nov. 2015.
[14] F. Jia, Y. Lei, J. Lin, X. Zhou, and N. Lu, “Deep neural networks: A promising tool for fault-characteristic mining and intelligent diagnosis of rotating machinery with massive data,” Mech. Syst. Signal Process., vol. 72–73, pp. 303–315, May 2016.
[15] L. Zhou, C. Zhang, F. Liu, Z. Qiu, and Y. He, “Application of deep learning in food: A Review,” Comprehensive Rev. Food Sci. Food Saf., vol. 18, no. 6, pp. 1793–1811, Nov. 2019.
[16] S. Zhang, S. Zhang, B. Wang, and T. G. Habetler, “Deep learning algorithms for bearing fault diagnostics–A review,” in Proc. IEEE 12th Int. Symp. Diagn. Electr. Mach., Power Electron. Drives (SDEMPED), Aug. 2019, pp. 257–263.
[17] O. Janssens, V. Slavkovikj, B. Vervisch, K. Stockman, M. Loccufier, S. Verstockt, R. Van de Walle, and S. Van Hoecke, “Convolutional neural network based fault detection for rotating machinery,” J. Sound Vib., vol. 377, pp. 331–345, Sep. 2016.
[18] J. Wang, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A large-scale hierarchical image database,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2009, pp. 248–255.
[19] W. Zhao, “Research on the deep learning of the small sample data based on transfer learning,” AIP Conf. Proc., vol. 1864, no. 1, 2017, Art. no. 020018.
[20] N. V. Chawla, K. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic minority over-sampling technique,” in Proc. IEEE 12th Int. Conf. Comp. Sci. Eng., Jun. 2009, pp. 248–255.
[21] S. Shao, P. Wang, and R. Yan, “Generative adversarial networks for data augmentation in machine fault diagnosis,” J. Sound Vib., vol. 30, pp. 41–47, Mar. 2010.
[22] H. Zhiyi, S. Haidong, J. Lin, C. Junsheng, and Y. Yu, “Transfer fault diagnosis of bearing installed in different machines using enhanced deep auto-encoder,” Measurement, vol. 152, Feb. 2020, Art. no. 107395.
[28] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in Proc. Adv. Neural Inf. Process. Syst. (NIPS), 2012, pp. 1097–1105.

[29] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, “A survey of the recent architectures of deep convolutional neural networks,” Artif. Intell. Rev., pp. 1–62, Apr. 2020.

[30] T. Lu, F. Yu, C. Xue, and B. Han, “Identification, classification, and quantification of three physical mechanisms in oil-in-water emulsions using AlexNet with transfer learning,” J. Food Eng., vol. 288, Jan. 2021, Art. no. 110220.

[31] C. Olah, A. Mordvintsev, and L. Schubert, “Feature visualization,” Distill, vol. 2, no. 11, p. e7, Nov. 2017.

[32] S. Zhang, S. Zhang, B. Wang, and T. G. Habetler, “Deep learning algorithms for bearing fault diagnosis—a comprehensive review,” IEEE Access, vol. 8, pp. 29857–29881, 2020.

[33] Case Western Reserve University (CWRU) Bearing Data Center. Accessed: Dec. 2018. [Online]. Available: https://csegroups.case.edu/bearingdatacenter/pages/welcome-case-western-reserve-university-bearing-data-center-website

[34] J. Lee, H. Qiu, G. Yu, J. Lin, and Rexnord Technical Services, “Bearing data set,” IMS, Univ. Cincinnati, NASA Ames Prognostics Data Repository, NASA Ames Res. Center, Moffett Field, CA, USA, Data Set, 2007. [Online]. Available: https://ti.arc.nasa.gov/project/prognostic-data-repository

[35] Paderborn University. Bearing Data Center. Accessed: Dec. 2018. [Online]. Available: https://mb.uni-paderborn.de/kat/forschung/datacenter/bearing-datacenter/

[36] IEEE PHM 2012 Data Challenge Bearing Dataset. Accessed: Dec. 2018. [Online]. Available: http://data-acoustics.com/measurements/bearing-faults/bearing-6

[37] H. Qiu, J. Lee, J. Lin, and G. Yu, “Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics,” J. Sound Vibrat., vol. 289, nos. 4–5, pp. 1066–1090, Feb. 2006.

[38] X. Wang, Y. Zheng, Z. Zhao, and J. Wang, “Bearing fault diagnosis based on statistical locally linear embedding,” Sensors, vol. 15, no. 7, pp. 16225–16247, Jul. 2015.

[39] R. Zhang, Z. Peng, L. Wu, B. Yao, and Y. Guan, “Fault diagnosis from raw sensor data using deep neural networks considering temporal coherence,” Sensors, vol. 17, no. 3, p. 549, Mar. 2017.

[40] L. Eren, T. Ince, and S. Kiranyaz, “A generic intelligent bearing fault diagnosis system using compact adaptive 1D CNN classifier,” J. Signal Process. Syst., vol. 91, no. 2, pp. 179–189, Feb. 2019.

[41] H. Jiang, X. Li, H. Shao, and K. Zhao, “Intelligent fault diagnosis of rolling bearings using an improved deep recurrent neural network,” Meas. Sci. Technol., vol. 29, no. 6, Jun. 2018, Art. no. 065107.

[42] W. Zhang, C. Li, G. Peng, Y. Chen, and Z. Zhang, “A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load,” Mech. Syst. Signal Process., vol. 100, pp. 439–453, Feb. 2018.

[43] 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.

[44] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, “A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals,” Sensors, vol. 17, no. 2, p. 425, Feb. 2017.

[45] M. Beale, M. Hagan, and H. Demuth, MATLAB Deep Learning Toolbox User’s Guide: PDF Documentation for Release R2019a. Natick, MA, USA: MathWorks, 2019.

[46] R. Sivaramakrishnan, S. Antani, Z. Xue, S. Candemir, S. Jaeger, and G. R. Thoma, “Visualizing abnormalities in chest radiographs through salient network activations in deep learning,” in Proc. IEEE Life Sci. Conf. (LSC), Dec. 2017, pp. 71–74.