

Article

# A Fault Diagnosis Method of Rotating Machinery Based on One-Dimensional, Self-Normalizing Convolutional Neural Networks

Jingli Yang \*, Shuangyan Yin, Yongqi Chang and Tianyu Gao

School of Electrical Engineering and Automation, Harbin Institute of Technology, Harbin 150080, China; 19S101111@stu.hit.edu.cn (S.Y.); 18S001022@stu.hit.edu.cn (Y.C.); 18B901013@stu.hit.edu.cn (T.G.)

\* Correspondence: jinglidg@hit.edu.cn; Tel.: +86-451-864-03-083

Received: 6 June 2020; Accepted: 6 July 2020; Published: 9 July 2020



**Abstract:** Aiming at the fault diagnosis issue of rotating machinery, a novel method based on the deep learning theory is presented in this paper. By combining one-dimensional convolutional neural networks (1D-CNN) with self-normalizing neural networks (SNN), the proposed method can achieve high fault identification accuracy in a simple and compact architecture configuration. By taking advantage of the self-normalizing properties of the activation function SeLU, the stability and convergence of the fault diagnosis model are maintained. By introducing  $\alpha$ -dropout mechanism twice to regularize the training process, the overfitting problem is resolved and the generalization capability of the model is further improved. The experimental results on the benchmark dataset show that the proposed method possesses high fault identification accuracy and excellent cross-load fault diagnosis capability.

**Keywords:** one-dimensional convolutional neural networks; self-normalizing neural networks; SeLU;  $\alpha$ -dropout; fault diagnosis

## 1. Introduction

Due to its strong point of versatility, rotating machinery is currently widely adopted in various pieces of mechanical equipment and works in extremely complex environments. Damage will not only impede the normal operation of the equipment, but also cause huge economic losses and pose threats to personal safety. Consequently, performing studies on the fault diagnosis of rotating machinery has become an urgent issue in the field of machine health monitoring.

Generally speaking, the fault diagnosis techniques of rotating machinery are commonly classified into the three types: model-based, signal-based and machine learning-based [1,2]. In model-based methods, a dynamic model should be developed based on the characteristics of vibration responses and the fault generation mechanism of rotating machinery. Cheng et al. [3] established a torsional vibration model for the planetary gear system, and conducted a signal simulation under the normal state and different breaking modes of the sun wheel gear. Then, a new metric named sideband-amplitude ratio was extracted from the frequency spectrum to quantitatively evaluate the damage degree. Park et al. [4] constructed a lumped parametric model of planetary gears based on the relationship between the gear mesh stiffness and the transmission errors, and it was verified in detecting planetary gear failures. Model-based fault diagnosis methods can enhance the comprehension of the transmission behavior and vibration responses of rotating machinery under different failure modes. However, there exist many assumptions and constraints in the model building process; thus, the influences of various uncertainties cannot be fully considered. Furthermore, there is still a certain gap between the simulation and the actual situations, which also has an impact on the fault diagnosis results to some extent.

In signal-based methods, two techniques can be used to obtain fault information for the diagnosis of rotating machinery: signal filtering or signal decomposition methods to separate fault components from normal components, and time-frequency analysis methods to reveal the time-varying characteristics of the frequency components of vibration signals. Vibration signals are commonly used in the field of rotating machinery diagnosis because the machine sounds can reflect the most inherent information of health state and contain lots of information, such as working condition [5,6]. Teng et al. [7] extracted the sensitive components of vibration signals using the traditional narrow-band filtering technique at first; then, the inverted spectral analysis and Hilbert-yellow transform were applied to detect the obvious cracking faults; and finally, the Gaussian wavelet transform and multi-scale envelope spectral demodulation were employed to successfully detect the cracking faults in the composite fault mode. Feng et al. [8] combined a Kalman filter with a higher-order energy demodulation operator to effectively diagnose early faults with different signature frequencies. Although fault diagnosis methods based on the signal processing technology are intuitive, practical and physically meaningful, traditional signal decomposition and demodulation methods still have their own limitations. For example, the methods for extracting the envelope spectrum and the transient frequency spectrum components are still complex, as they rely on staff with rich engineering experience to discriminate the types of faults. Recently, the development of machine learning has pushed the fault diagnosis of rotating machinery towards intelligence. In machine learning-based methods, the fault diagnosis process is divided into the two stages [9]: the fault feature extraction stage and the health state recognition stage (i.e., fault classification). The feature extraction aims at extracting information about the health state from the raw signals, removing redundancy and facilitating the fault classification of rotating machinery. The fault classification involves the selection of an appropriate classification method to construct a fault diagnosis model for identifying faults of rotating machinery. Traditional machine learning methods—artificial neural networks (ANN) [10], k-nearest neighbor (KNN) [11], sparse representation-based classifier (SRC) [12,13], support vector machine (SVM) [14,15], etc.—artificially extract fault features from the collected data and select sensitive features for training to achieve fault type identification automatically. Chen et al. [16] utilized probabilistic neural networks to perform effective fault diagnosis of hydroelectric turbine generators. Pandya et al. [17] presented an improved KNN algorithm based on an asymmetric proximity function to enhance the diagnosis precision of bearings. However, these methods still rely on the artificial feature extraction and suffer from problems such as low accuracy and poor generalization performance.

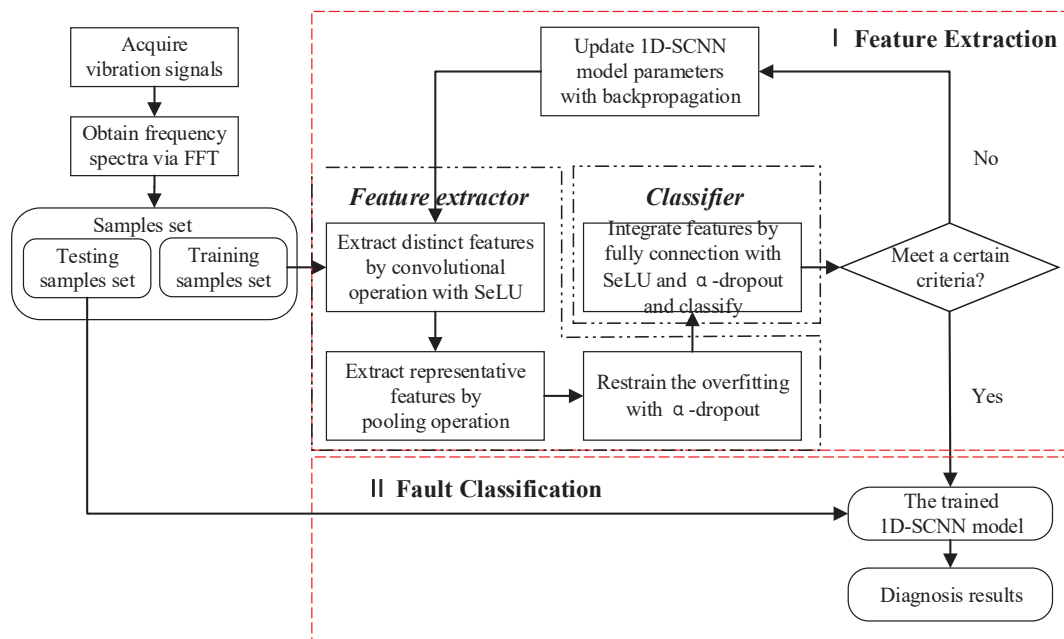
With the emergence of deep learning methods, ground-breaking solutions to the fault diagnosis problem of rotating machinery have been provided recently. Deep learning-based diagnosis approaches apply hierarchical networks to learn abstract fault features layer by layer, and then place an output layer after the last extraction layer to accomplish the fault identification task. Nowadays, depth network models are broadly used in the field of fault diagnosis of rotating machinery, such as depth neural networks (DNN), stacked automatic encoder (SAE), depth belief networks (DBN) and convolutional neural networks (CNN). Jia et al. [18] proposed a five-layer SAE model for the fault diagnosis of rotating machinery; the raw data were converted into the frequency-domain data and subsequently fed it into the model to implement the diagnosis task, and the effectiveness of the method was fully verified by experiments of rotating bearings and planetary gearboxes. Wang et al. [19] introduced the batch normalization layer (BN) into SAE, which could address the issue of internal covariance shifting while training multilayer networks. The experimental results showed that the proposed method could enhance fault identification accuracy and accelerate convergent speed of the training process. Yang et al. [20] presented a DNN-based automatic classification algorithm, which was evaluated by the sub-signals extracted in the frequency-domain. Shen et al. [21] adopted a feature robustness-enhanced adaptive fault diagnosis network based on a deep compression auto-encoder (CAE) and demonstrated its superiority on a gearbox databank. Yin et al. [22] optimized the network structure of DBN by a genetic algorithm and utilized it for the fault diagnosis of gear drive chains. Gan et al. [23] constructed a two-layer fault diagnosis network based on DBN, which adopted a wavelet transform for data

preprocessing and implemented fault type and fault degree identification respectively based on the two-layer network. Li et al. [24] combined a hybrid diagnosis model based on SAE and DBN. DBN was used to discriminate health states on the learning features extracted from SAE. Shang et al. [25] published a DBN-based fault diagnosis model for rolling bearings, which could evade the complex structure of deep neural nets to some extent. The proposed model has the merits of easily training and good fault diagnosis capability. However, most fault diagnosis models based on deep learning theory are regular fully-connected networks with low model generalization performance, and the parameters increase exponentially with the number of layers. Therefore, issues of overfitting and decreasing diagnosis ability still affect the actual applications of above methods.

Compared with the fully-connected DNN, CNN's characteristics such as sparse connection, weight sharing and pooling operations can reduce the number of parameters of the training network, and enhance the model's stability and generalization performance. The original version of CNN is a two-dimensional structure inspired by the visual system, which has achieved remarkable results in the field of image recognition because of its ability to describe the natural 2D spatial correlations in images [26,27]. Nowadays, the traditional 2D-CNN algorithm has been extensively applied in the field of fault diagnosis. Wen et al. [28] modified the LeNet-5 to perform fault diagnosis for bearings and centrifugal pumps, wherein the raw time-domain signal was transformed into a two-dimensional gray scale image to train the model. Guo et al. [29] adopted a time-frequency domain transformation of the original signal to obtain a continuous wavelet transform scale diagram (CWTS), on whose basis they adopted CNN to directly classify the fault signal. The validity and generality of the proposed method were confirmed on a rotor experiment platform. Zhao et al. [30] developed a planetary gearbox fault diagnosis strategy based on the synchro squeezing transform (SST) and the deep convolutional neural networks (DCNN), which showed the fault identification accuracy was up to 98.3%. Sun et al. [31] converted the raw signal to a 2D image; then they carried out automatic feature extraction from the image via CNN and completed the fault classification of bearings. Inspired by the successful applications in the fields of natural language processing and speech recognition [32,33], the 1D-CNN algorithm was introduced for fault diagnosis only recently. Liu et al. [34] improved the traditional LeNet-5 network, and their results showed that the improved 1D LeNet-5 network could achieve more significant performance on fault diagnosis. Eren et al. [35] utilized a compact adaptive 1D-CNN classifier to implement the fault diagnosis of induction bearings and proved the feasibility of the algorithm on a real dataset. Zhang et al. [36] presented an end-to-end fault diagnosis approach based on 1D-CNN, and the experimental results exhibited high accuracy, even in a noisy environment. Several studies have been conducted to prove the 1D-CNN has advantages over 2D-CNN in processing vibration signals for fault diagnosis. An et al. [37] compared the performances of 2D-CNN and 1D-CNN in bearing fault diagnosis, which demonstrated that the 1D-CNN model offered better feature extraction capability than the 2D-CNN model. Jing et al. [38] compared the fault diagnosis performances of different CNN models under three types of input data (i.e., the raw data, the spectrum data and the combined time-frequency data), and the results indicated that the 1D-CNN model is superior than others when the spectrum data is adopted as its input.

To be noted is that most of the current studies focusing on fault diagnosis methods of rotating machinery have been conducted on ideal data—that is, assuming the training samples cover the total working conditions of rotating machinery. However, in real engineering situations, it is infeasible to obtain the ideal data for all working conditions of rotating machinery because the working load changes constantly. Therefore, it is critical to utilize the data collected from limited conditions to establish fault diagnosis models. In other words, how to enhance the generalization capability of fault diagnosis models has emerged as one of the hot spots in practical industry applications.

Aiming at the shortfalls of existing methods, this paper proposes a fault diagnosis method based on one-dimensional, self-normalizing convolutional neural networks (1D-SCNN) to address the low accuracy and poor generalization capability of fault diagnosis for rotating machinery. As shown in Figure 1, the main innovation of this paper is summarized as follows.



**Figure 1.** The block diagram of the proposed method.

1. A fault diagnosis model based on 1D-SCNN is presented, which has a simple and compact architecture configuration with only a convolutional layer and a pooling layer. Compared with the conventional techniques, it can achieve competitive performance in terms of fault diagnosis accuracy and generalization capability.
2. The scaled exponential linear units (SeLU) are employed to strengthen the features of the fault signal. With the self-normalizing properties, activations can maintain normalization when propagating through layers of the network. Therefore, SeLU can maintain the stability and convergence of the network, and enhance the generalization capability of the model.
3. The  $\alpha$ -dropout algorithm is introduced into the feature extractor and classifier simultaneously, which not only can restrain the overfitting at the initial stage of training, but also should be able to accelerate the speed of the network's parameters updating and further boost the generalization capability of the model.
4. A series of experiments utilizing the Case Western Reserve University bearing dataset are conducted. The results demonstrate that the proposed method possesses good fault diagnosis accuracy and generalization capability, and provides an excellent solution for enhancing the reliability and maintainability of mechanical equipment.

## 2. One-Dimensional Convolutional Neural Networks

The one-dimensional convolutional neural network (1D-CNN) is a type of depth-feedback neural network with convolutional kernels of unit width. The literature [39] claims that CNNs dedicated to classification tasks each consist primarily of a feature extractor and a classifier.

### 2.1. Feature Extractor

The feature extractor is a composite structure used to map the raw signal into the feature space to characterize diverse fault features, and it is an essential element of 1D-CNN. The feature extractor is composed of the one-dimensional convolutional layer, the activation layer and the pooling layer.

### (1) One-dimensional convolutional layer

The one-dimensional convolutional layer extracts distinct fault features from the input data (or output features of the previous layer) through convolutional operations. It involves multiple feature maps with many neurons, and each neuron is sparsely connected to the feature map of the upper layer by a weight matrix (convolutional kernels). CNN weights are shared when the input feature map is the same as the output. Sparse connection and weight sharing can reduce network complexity and avoid the overfitting problem. The convolutional operation is shown in Equation (1) [36].

$$y^{l(i,j)} = \mathbf{K}_i^l * \mathbf{x}^{l(r^j)} = \sum_{j'=0}^{W-1} \mathbf{K}_i^{l(j')} \mathbf{x}^{l(j+j')} \quad (1)$$

where  $\mathbf{K}_i^{l(j')}$  denotes the  $j'$ -th weight or weights (if the depth of kernel is larger than 1) of the  $i$ -th convolutional kernel in the  $l$ -th convolutional layer,  $\mathbf{x}^{l(r^j)}$  denotes the local region where the  $j$ -th convolutional kernel in the  $l$ -th convolutional layer is convolved and  $W$  is the width of the convolutional kernel.

### (2) Activation layer

After the convolutional calculation, the activation function is adopted to perform a nonlinear transition, which can make it easier to distinguish the features of different types of faults. Rectified linear unit function (ReLU) is the commonly adopted activation function. The expression is given by the Equation (2) [36].

$$a^{l(i,j)} = f(y^{l(i,j)}) = \max\{0, y^{l(i,j)}\} \quad (2)$$

where  $a^{l(i,j)}$  is the activation of  $y^{l(i,j)}$ .

### (3) Pooling layer

The pooling layer conducts feature dimension reduction to achieve more representative characteristics. The pooling operation can enhance the stability and robustness of the features by eliminating the noise interference [26]. The mean-pooling layer and max-pooling layer are extensively utilized in convolutional networks. Max-pooling performs the local max operation on the perception domain of the output feature map, which is calculated by Equation (3) [36].

$$p^{l(i,j)} = \max_{(j-1)W+1 \leq t \leq jW} \{a^{l(i,t)}\} \quad (3)$$

where  $a^{l(i,t)}$  denotes the activation of the  $t$ -th neuron in the  $i$ -th frame of the  $l$ -th layer,  $W$  is the width of the pooling region and  $j$  denotes the  $j$ -th pooling kernel.

## 2.2. Classifier

The classifier involves full-connection layers and activation functions that integrate and classify fault features derived from the output of the pooling layer. Notably, the activation function for the hidden layer is ReLU, and for the last layer is softmax, as shown by Equation (4) [36]. Through softmax

regression, distinct fault features are converted into a normalized probability distribution, and the closer the value to 1, the more probable it is a true fault type.

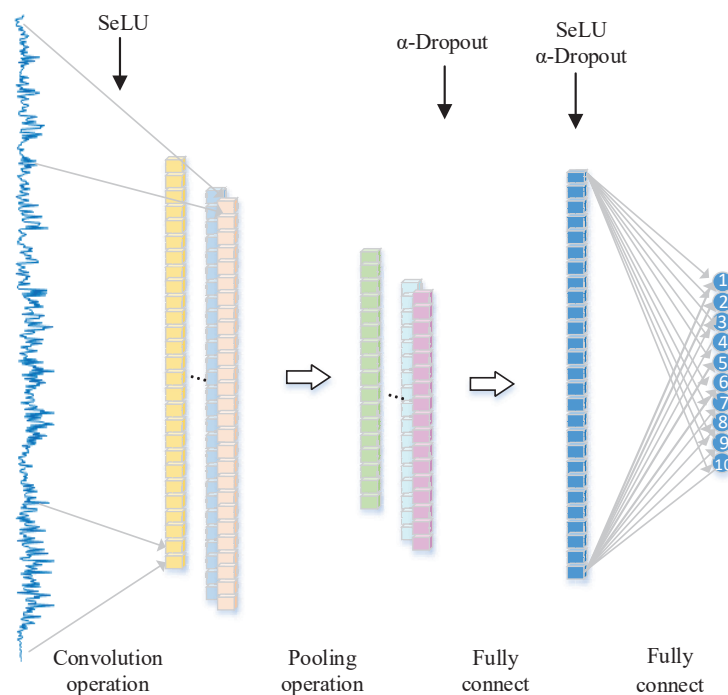
$$P(j) = \frac{e^{Z^l(j)}}{\sum_{k=1}^n e^{Z^l(k)}} \quad (4)$$

where  $Z^l(j)$  denotes the input of the  $j$ -th node in the  $l$ -th full-connection layer.

### 3. Methodology

#### 3.1. Overview

The proposed 1D-SCNN-based fault diagnosis method is concerned with increasing the fault diagnosis accuracy of rotating machinery and further promoting the generalization capability of fault diagnosis models, which can achieve fault type identification of rotating machinery even under non-ideal conditions. The architecture of the proposed method for fault diagnosis is shown in Figure 2.



**Figure 2.** The architecture of the proposed 1D-SCNN method for fault diagnosis.

The activation function SeLU is introduced in the convolutional layer. With the input feature of zero mean and unit variance, SeLU ensures that the output feature still has zero mean and unit variance. Therefore, it can make the network robust and avoid the overfitting problem. Details about the convolutional layer-based strategy are illustrated in Section 3.2.

Adding an  $\alpha$ -dropout layer after the pooling layer can improve the generalization capability of the model, wherein  $\alpha$ -dropout is improved by normalizing the output features upon dropout. Details about the pooling layer-based strategy are shown in Section 3.3.

Applying the activation functions SeLU and  $\alpha$ -dropout to the full-connection layer can not only enhance the fault features, but also further improve the generalization capability of the model. Details about the full-connection layer-based strategy are described in Section 3.4.

### 3.2. SeLU-Based Enhancement of Convolutional Fault Feature Extraction

The structure of the self-normalizing neural networks (SNN) is shown in Figure 3. Assuming  $\{x_i\}_{i=1}^n$  is the training set of SNN, the sample  $x_i$  of the random variable  $X$  obeys a distribution with the mean  $\mu$  and variance  $\sigma^2$ . The linear transformation of  $x_i$  yields the pre-activation vector  $z_i = w_i x_i$ , where  $w_i$  is the weight matrix and  $z_i$  is the implementation of the random variable  $Z$ . The mean and variance of  $Z$  are shown by Equations (5) and (6).

$$E(Z) = \sum_{i=1}^n w_i E(x_i) = \sum_{i=1}^n w_i \mu = \mu \bar{w} \quad (5)$$

$$\text{Var}(Z) = \text{Var}\left(\sum_{i=1}^n w_i x_i\right) = \sum_{i=1}^n w_i^2 \sigma^2 = \sigma^2 \tau \quad (6)$$

where  $\bar{w} = \sum_{i=1}^n w_i$  and  $\tau = \sum_{i=1}^n w_i^2$ . According to the central limit theorem,  $Z$  obeys the normal distribution:  $N(\mu \bar{w}, \sqrt{\sigma^2 \tau})$  with  $p_N(Z; \mu \bar{w}, \sqrt{\sigma^2 \tau})$ .

Klambauer [40] et al. presented a novel activation function named SeLU as follows.

$$\text{selu}(x) = \lambda \begin{cases} x, & \text{if } x > 0 \\ \alpha e^x - \alpha, & \text{if } x \leq 0 \end{cases} \quad (7)$$

The function performs the mapping of the mean and variance of the random variable from the former layer to the next layer by mapping  $F$  [40]:

$$F: \begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix} \mapsto \begin{pmatrix} \tilde{\mu} \\ \tilde{\sigma}^2 \end{pmatrix}: \begin{aligned} \tilde{\mu}(\mu, \bar{w}, \sigma^2, \tau) &= \int_{-\infty}^{\infty} \text{selu}(z) p_N(z; \mu \bar{w}, \sqrt{\sigma^2 \tau}) dz \\ \tilde{\sigma}^2(\mu, \bar{w}, \sigma^2, \tau) &= \int_{-\infty}^{\infty} \text{selu}(z)^2 p_N(z; \mu \bar{w}, \sqrt{\sigma^2 \tau}) dz - (\tilde{\mu})^2 \end{aligned} \quad (8)$$

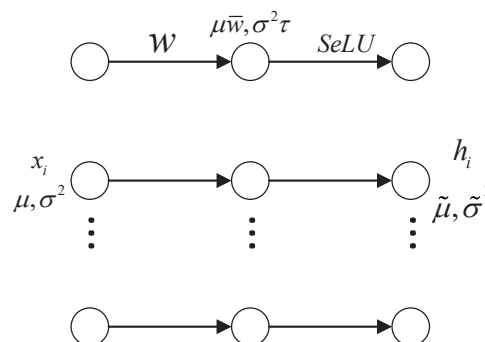


Figure 3. The block diagram of SNN.

After normalizing the weight matrix:  $\bar{w} = 0, \tau = 1$ , given the fixed point  $(\mu, \sigma^2) = (0, 1)$ , we have  $\tilde{\mu} = \mu = 0, \tilde{\sigma}^2 = \sigma^2 = 1$ . By solving Equation (8) for  $\alpha$  and  $\lambda$ , the solutions  $\alpha = 1.6732632423543772\dots$  and  $\lambda = 1.0507009873554804\dots$  can be obtained. At this point, the  $(2 \times 2)$ -Jacobian of  $F$  [40] is

$$J(\mu, \sigma^2) = \begin{pmatrix} \frac{\partial \mu^{new}(\mu, \sigma^2)}{\partial \mu} & \frac{\partial \mu^{new}(\mu, \sigma^2)}{\partial \sigma^2} \\ \frac{\partial \sigma^{new}(\mu, \sigma^2)}{\partial \mu} & \frac{\partial \sigma^{new}(\mu, \sigma^2)}{\partial \sigma^2} \end{pmatrix}, J(0, 1) = \begin{pmatrix} 0.0 & 0.088834 \\ 0.0 & 0.782648 \end{pmatrix} \quad (9)$$

The norm of  $J$  is equal to  $0.7877 < 1$ ; i.e., the mapping  $F$  is a contraction mapping and the corresponding fixed point is stable.

As mentioned above, SeLU can ensure that the activation passes through the network layers in a normalized state, and its output value tends to a stable fixed point. Once the perturbation and noise occur, the shift will be pulled back to the normalized state immediately, which can avoid the overfitting problem and offer good robustness. Compared with batch normalization, there is no need to calculate the variance and offset of the current input data before updating the network parameters, which means the reduction of the computational complexity. Furthermore, SeLU has no dead zone; i.e., neurons can still be activated when the input is less than 0, which means richer features can be obtained compared to ReLU. Consequently, applying the activation function SeLU to the convolutional layer of the 1D-SCNN model can incorporate the advantages of self-normalization properties into the feature extractor of 1D-SCNN, and enhance the feature extraction capability of the model. Moreover, the robustness of the network can also be guaranteed.

### 3.3. $\alpha$ -Dropout-Based Improvement of Pooling Layer Generalization Capability

Dropout is an effective technique to address the overfitting problem of deep neural nets [41], the key idea of which is to randomly drop units and their connections from the neural network to prevent co-adapting between units. The dropout rate  $q$  ( $0 < q < 1$ ) indicates the probability that each neuron in the network layer will stop working, and its value is generally set to be 0.5.

To enhance the generalization capability of the model, 1D-SCNN introduces the idea of dropout by adding it after the pooling layer. Furthermore, the  $\alpha$ -dropout improved algorithm is adopted to maintain the normalization of the values that pass across the network. In the dropout algorithm, the activations are scaled by  $1/q$  during the training process and remain constant during the testing process. Moreover, the dropout variable  $d$  obeys a binomial distribution  $B(1, q)$ .  $\alpha$ -Dropout can set the input to  $\alpha'$  because the default value of the low variance is  $\lim_{x \rightarrow -\infty} \text{selu}(x) = -\lambda\alpha = \alpha'$ , where the mean and variance are described as follows [40].

$$\begin{aligned} E(xd + \alpha'(1-d)) &= q\mu + (1-q)\alpha' \\ \text{Var}(xd + \alpha'(1-d)) &= q\left((1-q)(\alpha' - \mu)^2 + \nu\right) \end{aligned} \quad (10)$$

To ensure that the variance and mean remain constant after applying the  $\alpha$ -dropout algorithm, the affine transformation shown in Equation (11) [40] is adopted.

$$\begin{aligned} E(a(xd + \alpha'(1-d)) + b) &= \mu \\ \text{Var}(a(xd + \alpha'(1-d)) + b) &= \nu \end{aligned} \quad (11)$$

When  $\mu = 0, \nu = 1, a = (q + \alpha'^2 q(1-q))^{-1/2}, b = -(q + \alpha'^2 q(1-q))^{-1/2}((1-q)\alpha')$ , the values of the parameters  $a$  and  $b$  depend on  $q$  and  $\alpha'$ .

In the absence of the  $\alpha$ -dropout algorithm, the forward propagation process for each hidden layer parameters is described by Equation (12).

$$\begin{aligned} z_i^{(l+1)} &= w_i^{(l+1)} y^l + b_i^{(l+1)} \\ y_i^{(l+1)} &= \text{selu}\left(z_i^{(l+1)}\right) \end{aligned} \quad (12)$$

With the  $\alpha$ -dropout algorithm, the forward propagation process of each hidden layer parameter is shown in Equation (13).

$$\begin{aligned} r_j^{(l)} &\sim \text{Bernoulli}(q) \\ \tilde{y}^{(l)} &= r^{(l)} \times y^{(l)} \\ z_i^{(l+1)} &= w_i^{(l+1)} \tilde{y}^{(l)} + b_i^{(l+1)} \\ y_i^{(l+1)} &= \text{selu}\left(z_i^{(l+1)}\right) \end{aligned} \quad (13)$$



where  $z_i^{(l+1)}$  denotes the input of the  $(l + 1)$ -th layer and  $y_i^{(l+1)}$  denotes the output of the  $(l + 1)$ -th layer.  $r^{(l)}$  is a vector of the  $l$ -th layer, which consists of binomially distributed random variables with values of  $1 - q$ .

In short, the overfitting problem of 1D-SCNN can be addressed by applying the  $\alpha$ -dropout algorithm to the pooling layer of the feature extractor, and the generalization capability of the model can also be improved.

### 3.4. SeLU and $\alpha$ -Dropout-Based Advancement of Full-Connection Layer Generalization Capability

The full-connection layer is used to identify the fault features output by the feature extractor. The selection of the activation function plays a crucial role in the linear separability of the features and the stability of the network. Therefore, SeLU is selected as the activation function of the full-connection hidden layer, the principles and characteristics of which are described in Section 3.2. Moreover, to further improve the generalization capability of the model, the full-connection layer is improved via the  $\alpha$ -dropout algorithm, the principles and characteristics of which are detailed in Section 3.3.

### 3.5. Training of 1D-SCNN

The 1D-SCNN training process includes the forward propagation of the signal and the backward propagation of the error. The probability distribution of the sample type is obtained by the forward propagation of the signal, and the cross-entropy loss function is employed to evaluate the consistency of this output with the real sample label. Assuming that  $q_k^j$  represents the probability of the  $k$ -th sample predicted to be  $j$ -th class and  $p_k^j$  is the one-hot vector of the target sample label, the expression of the cross-entropy loss function is given by Equation (14) [36].

$$L = -\frac{1}{m} \sum_{k=1}^m \sum_j p_k^j \log q_k^j \quad (14)$$

where  $m$  denotes the size of the mini-batch. The backward propagation of the error is the main foundation for 1D-SCNN to optimize the weights, which takes the loss value as the source of the error and updates the gradient from back to front by the chain law.

The error backpropagation begins at the full-connection layer and is calculated similarly to ANN, and the process can be demonstrated by Equations (15)–(19) [36]. Here,  $z^{i+1(j)}$  is the logits value of the last layer.

$$\frac{\partial L}{\partial z^{i+1(j)}} = \sum_{k=1}^m p_k^j q_k^j - p_k^j \quad (15)$$

$$\frac{\partial L}{\partial \mathbf{w}_{ij}^l} = \frac{\partial L}{\partial z^{i+1(j)}} \cdot \frac{\partial z^{i+1(j)}}{\partial \mathbf{w}_{ij}^l} = \frac{\partial L}{\partial z^{i+1(j)}} \cdot a^{i(i)} \quad (16)$$

$$\frac{\partial L}{\partial \mathbf{b}_j^l} = \frac{\partial L}{\partial z^{l+1(j)}} \cdot \frac{\partial z^{l+1(j)}}{\partial \mathbf{b}_j^l} = \frac{\partial L}{\partial z^{l+1(j)}} \quad (17)$$

$$\frac{\partial L}{\partial a^{l(i)}} = \sum_j \frac{\partial L}{\partial z^{l+1(j)}} \cdot \frac{\partial z^{l+1(j)}}{\partial a^{l(i)}} = \sum_j \frac{\partial L}{\partial z^{l+1(j)}} \cdot \mathbf{w}_{ij}^l \quad (18)$$

$$L = -\frac{1}{m} \sum_{k=1}^m \sum_j p_k^j \log q_k^j \quad (19)$$

The weight  $\mathbf{W}$  and bias  $\mathbf{b}$  are updated iteratively as follows.

$$\begin{aligned}\mathbf{W}_{ij}^l &= \mathbf{W}_{ij}^l - a \frac{\partial L}{\partial \mathbf{W}_{ij}^l} \\ \mathbf{b}_i^l &= \mathbf{b}_i^l - a \frac{\partial L}{\partial \mathbf{b}_i^l}\end{aligned}\quad (20)$$

where  $a$  denotes the iteration rate.

The backpropagation calculation of the pooling layer is the inverse operation of the forward propagation. As for the max-pooling operation, only pooling units that deliver the maximum are involved. The specific propagation process is shown in the Equation (21) [36].

$$\frac{\partial L}{\partial a^{l(i,t)}} = \frac{\partial L}{\partial p^{l(i,j)}} \cdot \frac{\partial p^{l(i,j)}}{\partial a^{l(i,z)}} = \begin{cases} 0 & t \neq t_h \\ \frac{\partial L}{\partial p^{l(i,j)}} & t = t_h \end{cases}\quad (21)$$

where  $h$  indicates the maximal location of the pooling region.

The backpropagation calculation of the convolutional layer requires the gradient of each unit, which is calculated as follows.

$$\frac{\partial L}{\partial y^{l(i,j)}} = \frac{\partial L}{\partial a^{l(i,j)}} \cdot \frac{\partial a^{l(i,j)}}{\partial y^{l(i,t)}} = \begin{cases} 0 & y^{l(i,j)} \leq 0 \\ \frac{\partial L}{\partial a^{l(i,j)}} & y^{l(i,j)} > 0 \end{cases}\quad (22)$$

$$\frac{\partial L}{\partial \mathbf{x}^{l(j)}} = \sum_i \frac{\partial L}{\partial y^{l(i,j)}} \cdot \frac{\partial y^{l(i,j)}}{\partial \mathbf{x}^{l(j)}} = \sum_i \frac{\partial L}{\partial y^{l(i,j)}} \cdot \sum_{j'=0}^{W-1} \mathbf{K}_i^l(j')\quad (23)$$

$$\frac{\partial L}{\partial \mathbf{K}_i^l(j')} = \frac{\partial L}{\partial y^{l(i,j)}} \cdot \frac{\partial y^{l(i,j)}}{\partial \mathbf{K}_i^l(j')} = \frac{\partial L}{\partial y^{l(i,j)}} \cdot \sum_j \mathbf{x}^{l(j)}\quad (24)$$

### 3.6. Fault Diagnosis Process

The flow of the proposed 1D-SCNN-based fault diagnosis method of rotating machinery is shown in Figure 4. The method can adaptively extract fault features from the spectrum of the vibration signals, thereby enabling effective fault identification by features. Besides, the proposed fault diagnosis model has superior generalization capability which in turn provides good adaptive performance under cross-load levels. The algorithm is implemented as follows.

1. The spectra of the vibration signals are obtained using the fast Fourier transform (FFT) at the raw signal length without the windowing function, and are used as input samples for the 1D-SCNN model.
2. Randomly divide the input samples into a training sample set and a testing sample set with a ratio of 7:3. The training sample set serve as the input for the training stage of the model and the testing sample set is adopted for the testing stage.
3. Train the 1D-SCNN model by the forward propagation and the backward propagation operations, and save the trained model after meeting a certain criterion.
4. Load the trained fault diagnosis model and input the testing sample set into above diagnosis model to obtain the fault diagnosis results.

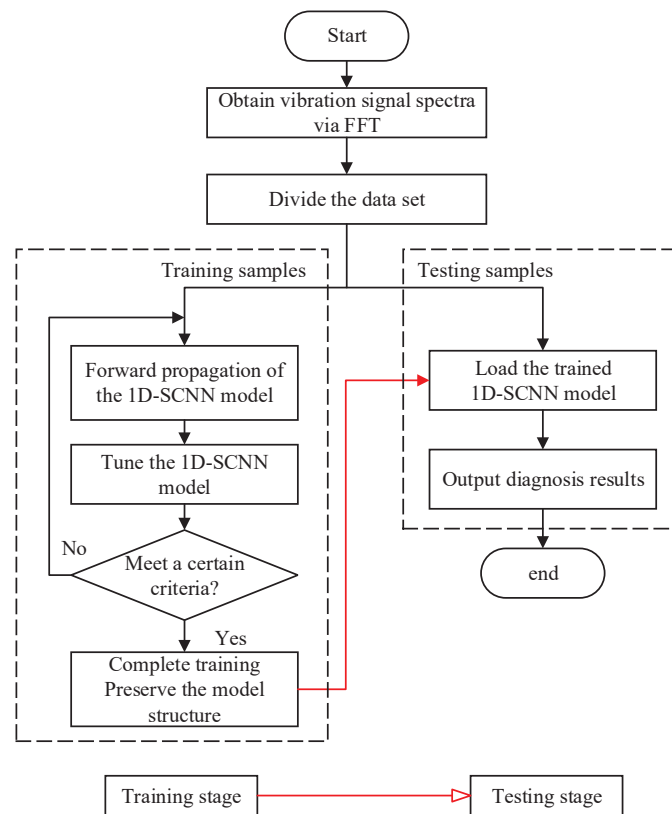


Figure 4. Flow diagram of 1D-SCNN-based fault diagnosis method of rotating machinery.

#### 4. Experimental Results and Analysis

The deep learning framework chosen in this study was Keras. By using Python as the programming language, and adopting Pycharm as the editor, the proposed method was developed with TensorFlow. The computer used during the experiments was configured with CPU i5-8265U, 16G memory and a 64-bit operating system. To verify the superior performance of the proposed 1D-SCNN-based fault diagnosis method, experiments for evaluating of the fault diagnosis accuracy and the generalization capability were conducted, respectively.

##### 4.1. Dataset Description

Rolling bearings from the Bearing Laboratory at Case Western Reserve University (USA) were selected as experimental subjects. In this paper, the vibration signal data of a deep groove ball bearing at the driven end with a sampling frequency of 12 kHz are selected for experimental analysis. The damaged parts of the bearing are: inner race, ball and outer race; the degrees of failure are 0.007, 0.014, 0.021 (unit: inch), respectively. The healthy state is labeled as 1, and labels for the inner race's damaged states are 2, 3 and 4, in ascending order of degree of damage; the ball and outer race states are labeled in the same way. In other words, above 10 labels represent the 10 health states of the bearing, i.e., fault types. The four datasets A, B, C and D were prepared with load levels of 0, 1, 2 and 3 (unit: hp), respectively. Each dataset contains 10 fault types described above, and each fault type consists of 100 samples. Furthermore, each sample involves 1024 consecutive data points, which were obtained by dividing the spectrum. Specific information on the experimental dataset is illustrated in Table 1.

**Table 1.** Description of the experimental dataset.

Damaged Parts	None	Inner Race				Ball		Outer Race		Load	
Label	1	2	3	4	5	6	7	8	9	10	
Damaged degree	0	0.007	0.014	0.021	0.007	0.014	0.021	0.007	0.014	0.021	
A	train	70	70	70	70	70	70	70	70	70	0
	test	30	30	30	30	30	30	30	30	30	
B	train	70	70	70	70	70	70	70	70	70	1
	test	30	30	30	30	30	30	30	30	30	
C	train	70	70	70	70	70	70	70	70	70	2
	test	30	30	30	30	30	30	30	30	30	
D	train	70	70	70	70	70	70	70	70	70	3
	test	30	30	30	30	30	30	30	30	30	

## 4.2. Results Analysis

### 4.2.1. Selection of Model Parameters

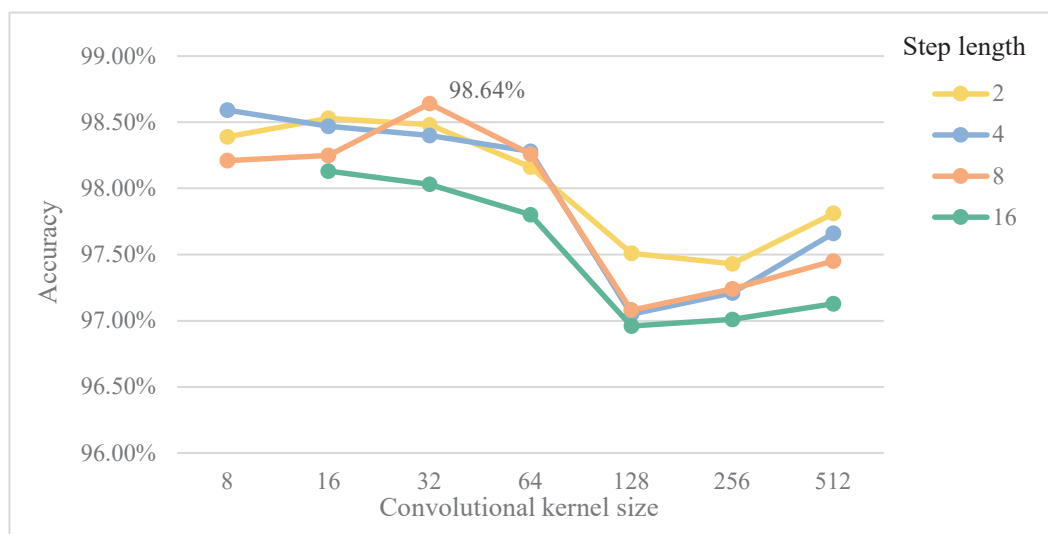
The hyperparameters of the 1D-SCNN model are set as follows: the convolutional kernel size is initialized by the LeCun uniform initializer (ensure that the output mean is 0 and the variance is 1), the pooling kernel size is 2, the pooling step length is 2, the number of neurons in the full-connection layer is 64, the  $\alpha$ -dropout rate is 0.5, the optimizer is Adam, the iteration number is set to 200 and the sample size for each training round is 100. The size and step length of convolutional kernels, and the numbers of convolutional kernels and layers play vital roles in the performance of the model, which was determined experimentally in this study.

The preliminary evaluation experiments on the fault diagnosis accuracy and generalization capability demonstrate that above three hyperparameters have few effects on the fault diagnosis accuracy under the same load level, whereas the model generalization capability under the cross-load level conditions is more affected. Accordingly, the average of the fault identification accuracy under cross-load level conditions is employed as the indicator for parameter optimization. To eliminate the influence of random factors, 30 groups of experiments are repeated.

#### (1) Convolutional kernel size and step length

Zhang et al. [36] stated that the first layer of a convolutional network should have large convolutional kernels for one-dimensional signals, which facilitates the automatic learning of network-oriented fault diagnosis features. Considering that the length of the data sample selected in this study was 1024, the effects of convolutional kernel size and step length on the generalization capability of the model are discussed with convolutional kernel sizes of 8, 16, 32, 64, 128, 256 and 512, respectively. First, the numbers of convolutional and pooling layers were fixed both to one with 16 convolutional kernels, and then the step length is multiplied from 2 to 16 under each convolutional kernel, where the maximum step length is 8 when the convolutional kernel is 8. Thus, the effects of convolutional kernel size and step length on the diagnosis accuracy are analyzed simultaneously. The experimental results are depicted in Figure 5.

As revealed by Figure 5, the impact of convolutional kernel size on diagnosis performance is greater than that of convolutional step length, which reflects the significance of convolutional kernel size selection in the first layer of the model. Since the features extracted from small convolutional kernels are not representative enough, large convolutional kernels can cause the computational volume to spike. In this paper, the convolutional kernel size of the first layer of is set to 32 and the step length is set to 8.



**Figure 5.** Effects of the convolutional kernel size and step length on the performance of 1D-SCNN.

### (2) Number of convolutional kernels

After determining the convolutional kernel size and the step length, the effects of the number of convolutional kernels on the generalization capability of the 1D-SCNN model are analyzed, and the experimental results are shown in Table 2.

**Table 2.** Effect of the number of convolutional kernels on the performance of 1D-SCNN.

Number of Convolutional Kernels	8	16	32	64
Accuracy	98.51%	98.64%	98.54%	98.44%

As shown in Table 2, as the number of convolutional kernels increases from 8 to 32, and the accuracy rises initially and falls later, and achieves the highest value when the number of convolutional kernels is 16; thus, the number of convolutional kernels of the first layer is selected as 16.

### (3) Number of layers

The number of convolutional kernels, the convolutional step length and the number of convolutional kernels were determined, after which the effects of alternating layers of convolutional and pooling on the generalization capability of the 1D-SCNN model were analyzed. Drawing on the design principle of the network structure in the literature [42], the first layer was devised as a large convolutional kernel, and all other layers adopted a convolutional kernel size of  $3 \times 1$ . The results of the experiment are displayed in Table 3.

**Table 3.** Effects of layer number on the performance of 1D-SCNN.

Network Structure	Monolayer	Bilayer	Trilayer	Four Layers	Five Layers
Accuracy	98.64%	97.6%	96.37%	93.62%	91.84%

As shown in Table 3, the fault diagnosis accuracy shows a decreasing trend as the number of network layers increases; i.e., the model possesses the best generalization capability when it contains only a single layer. Therefore, we utilizes one layer for the proposed model.

Based on above experimental results, the parameters of the 1D-SCNN model are determined as shown in Table 4. The model has only one convolutional and pooling layer with the convolutional

kernel size of 32 and the step length of 8. The number of units in the full-connection layer is 64, and the softmax layer yields 10 outputs, which corresponds to 10 bearing health states.

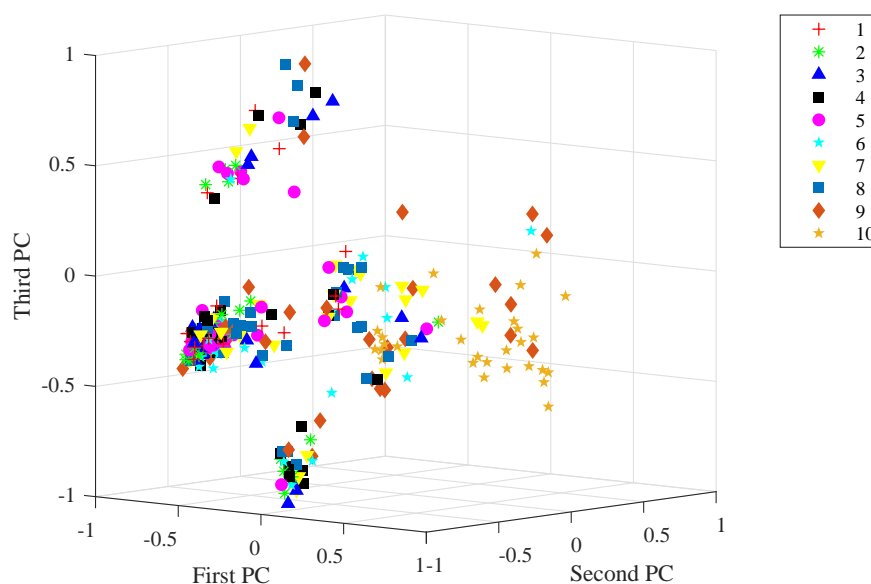
**Table 4.** Parameters of the 1D-SCNN model.

Number	Network Layer	Kernel Size/Step Length	Number of Kernels	Output Size
1	Convolution	$32 \times 1/8 \times 1$	16	$128 \times 16$
2	Pooling	$2 \times 1/2 \times 1$	16	$64 \times 16$
3	Full-connection	64	1	$64 \times 1$
4	Softmax	10	1	10

#### 4.2.2. Fault Diagnosis Accuracy Assessment

This section evaluates the fault diagnosis accuracy of the proposed model by describing a series of experiments using four datasets, A, B, C and D, respectively. For each dataset, 30 random sample experiments were conducted under same load level. During each experiment, 70 samples of each fault type were randomly selected as training samples, and the remaining were used as testing samples. In other words, the size of the training sample set was 700 and that of the testing sample set was 300. Furthermore, the average of the fault identification accuracy was taken to eliminate the influence of random factors.

Taking the data under 2hp load level as an example, the raw vibration signal and the feature extracted from the 1D-SCNN model were processed by principal component analysis (PCA). With visualizing techniques, the obtained top three principal components for the raw vibration signal and the extracted feature achieved by the feature extractor of the 1D-SCNN model are plotted in Figures 6 and 7, respectively. As depicted in Figure 6, the principal components of the raw vibration signal of the 10 different types of faults exhibit scattered overlap state. In contrast, Figure 7 depicts that the principal components of the extracted feature achieved by the 1D-SCNN model distinctly present non-overlapping discrete states in 3D space. Accordingly, the 1D-SCNN model possesses excellent adaptive fault feature extraction capability.



**Figure 6.** Raw vibration signals.

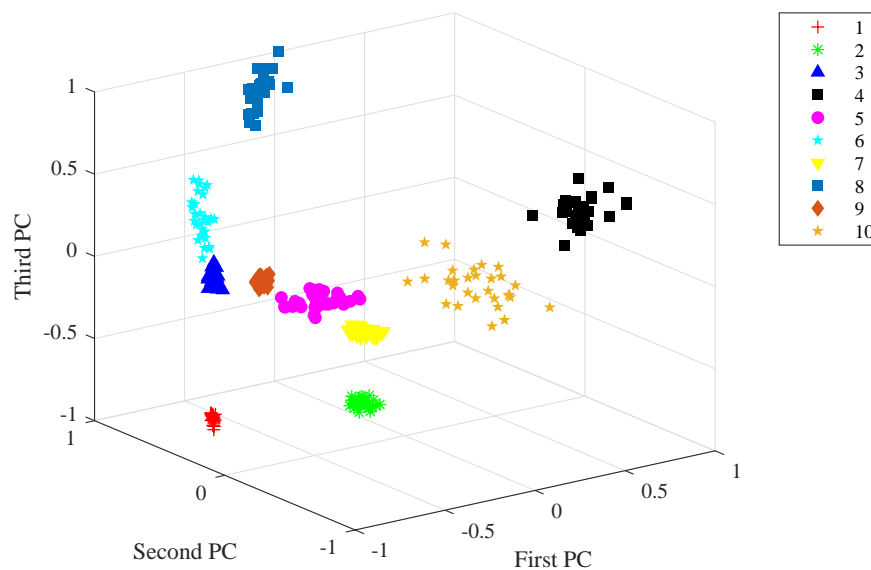


Figure 7. Features extracted from 1D-SCNN.

Figures 8 and 9 illustrate the changes in model loss and accuracy during the training process and the testing process under 0 hp load level. As shown in these figures, the loss of both the training process and testing process gradually decreases and converges to 0 when the number of iterations reaches about 200. In addition, the training accuracy increases along with the testing accuracy approaching 100% and stabilizes after several iterations.

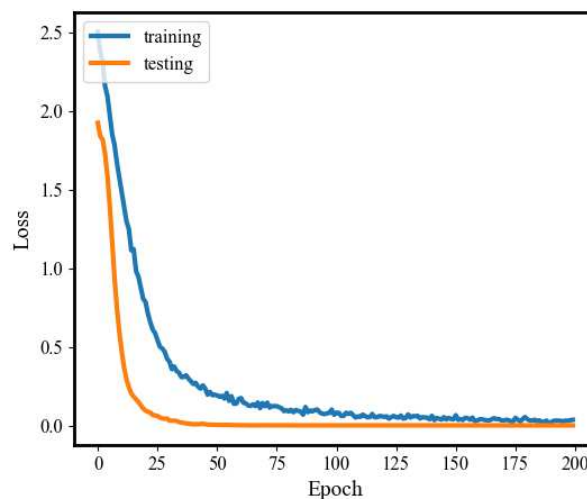


Figure 8. 1D-SCNN model loss.

To further evaluate the fault diagnosis performance of the proposed model, some representative methods in the field of fault diagnosis of rotating machinery (i.e., EMD+SVM [14], FFT+DNN [20], FFT+SDAE [18] and 1D-CNN [34]) were also implemented in these experiments. Among them, the number of neurons per layer was 1024, 800, 200, 200 or 10 for DNN; the number 1024, 800, 200 or 10 for SDAE. In addition, the numbers of mini-batches and iterations were set to 100 and 200, respectively. As shown in Table 5, the diagnosis performance of the traditional machine learning method (EMD+SVM) was worse than those of depth neural networks. Moreover, the 1D-CNN method does not perform well because its input is the time-domain signal, and its optimization time is usually longer than methods that adopt the frequency spectrum as input. Therefore, the 1D-CNN fault

diagnosis model is unstable under the specified number of iterations. Including the proposed method, the fault identification accuracy of the three deep learning-based fault diagnosis model is close to 100%, which exhibits superior fault diagnosis performance.

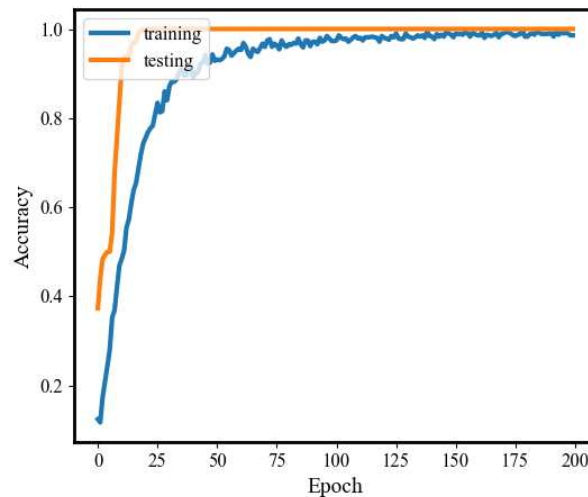


Figure 9. Fault diagnosis accuracy of 1D-SCNN.

Table 5. Comparison of experimental results of different methods.

Comparison Method	A	B	C	D	AVG
EMD+SVM	80.90%	84.77%	93.27%	96.57%	88.88%
1D-CNN	92.59%	85.22%	91.29%	92.82%	90.48%
FFT+DNN	100%	100%	100%	100%	100%
FFT+SDAE	100%	100%	100%	100%	100%
FFT+1D-SCNN	99.81%	100%	100%	100%	99.95%

In the actual applications of rotating machinery, the working conditions are complex and variable, which makes it impossible to obtain adequate data samples under each load level; thus, the diagnosis capability under cross-load level conditions is also an extremely significant performance indicator. The generalization capability of the proposed 1D-SCNN model is evaluated in this section. For each load level of the rolling bearing, 30 fault diagnosis models were generated by the random sample method. During the generation of each model, 70 samples of each fault type under the same load level were randomly selected as training samples. In other words, the size of the training sample set for each diagnosis model was 700. To validate the generalization capability of the generated fault diagnosis models, a series of experiments were conducted for each fault diagnosis model on three testing datasets under different load levels. Over the course of each experiment, 30 samples of each fault type under each load level were selected to form a testing sample set. In other words, the size of the testing sample set was 900, and the mean of the fault identification accuracy was taken to eliminate the influences of random factors. Let us take the 1D-SCNN model under 0 hp load level as an example. The model is used to identify data samples at 2 hp load level. By utilizing PCA on the features extracted from the 1D-SCNN model, the top three principal components obtained are visualized and shown in Figure 10. As shown in the figure, via the feature extraction operation of the 1D-SCNN model, the principal components of the 10 different types of faults under the cross-load level conditions exhibit a non-overlapping discrete state in 3D space, which means fault identification can be accomplished accurately with these extracted features.



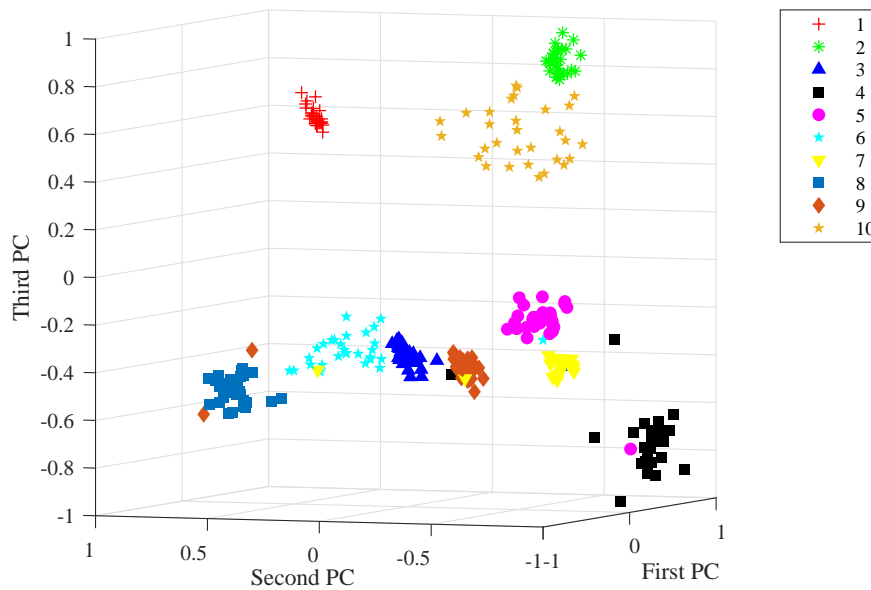


Figure 10. Features extracted by the model under cross-load level conditions.

#### 4.2.3. Model Generalization Capability Assessment

To quantitatively assess the generalization capability of the proposed method, the fault identification accuracy of the proposed 1D-SCNN-based fault diagnosis method is compared with that of some representative methods, such as EMD+SVM, FFT+DNN, FFT+SDAE and 1DCNN under cross-load level conditions. The experimental results are exhibited in Figure 11.

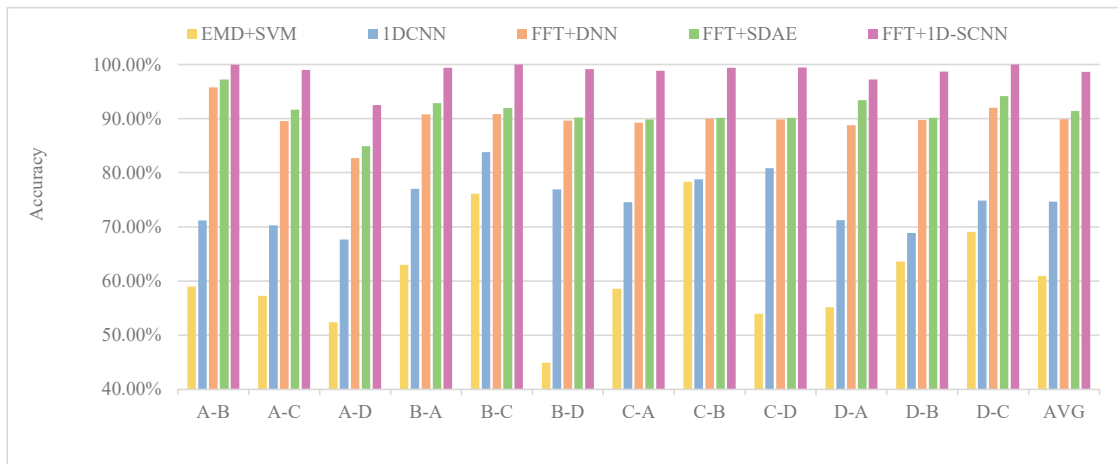


Figure 11. Fault diagnosis accuracy of different methods under cross-load level conditions.

As shown in Figure 11, the proposed 1D-SCNN model achieves the fault identification accuracy of more than 90% under each cross-load diagnosis scenarios, and it offers the best performance among all methods. Furthermore, the training model generated under a certain load level always shows high performance in the identification of samples collected under similar load levels, and its fault identification accuracy also decreases gradually as the load level span increases. The reason for above phenomenon is that the fault samples collected under near-load level have more similarities; i.e., 0 hp causes a similar degree of damage to 1 hp and gaps to 3 hp. Overall, the 1D-SCNN-based fault diagnosis method has an average fault identification accuracy of 98.64%, which is considerably higher than its counterparts. Therefore, it has excellent generalization capability and is suitable for cross-load fault diagnosis.

#### 4.2.4. Influence of SeLU and $\alpha$ -Dropout on Model Performance

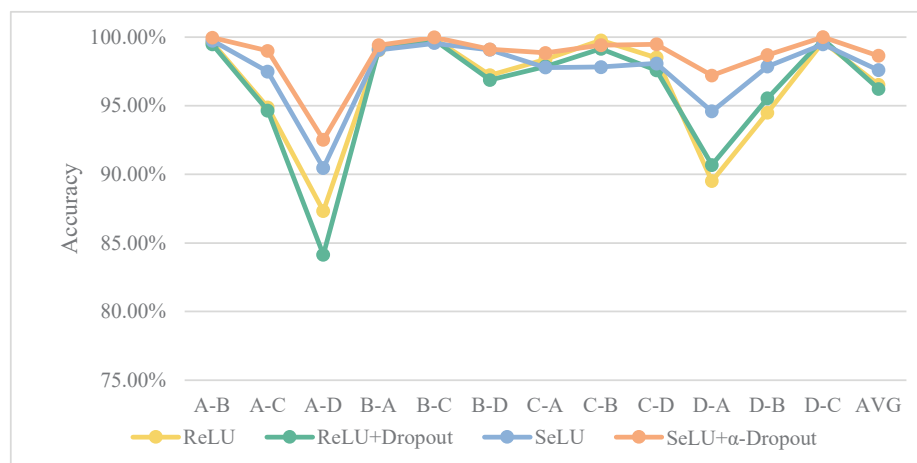
This section verifies the enhancement level of SeLU and  $\alpha$ -dropout on the performance of the 1D-SCNN model. Based on 1D-SCNN, the performance of the model is verified for four cases; i.e., the activation function is ReLU without dropout, the activation function is ReLU with dropout, the activation function is SeLU without  $\alpha$ -dropout and the activation function is SeLU with  $\alpha$ -dropout. The experimental results are presented in Table 6 and Figure 12.

**Table 6.** Effects of SeLU and  $\alpha$ -dropout on the diagnostic performance of 1D-SCNN.

Comparison Method	A	B	C	D	AVG
ReLU	99.78%	100%	100%	100%	99.95%
ReLU + Dropout	99.82%	100%	100%	100%	99.96%
SeLU	99.90%	100%	100%	100%	99.95%
SeLU + $\alpha$ -Dropout	99.81%	100%	100%	100%	99.95%

It can be seen that both changing the activation function from ReLU to SeLU and using a combination of SeLU and  $\alpha$ -dropout have few effects on model diagnosis performance under the same load level.

Figure 12 discusses the effects of SeLU and  $\alpha$ -dropout on the generalization capability of the 1D-SCNN model. Overall, the enhancing strategy of changing the activation function from ReLU to SeLU can improve fault identification accuracy of the model under cross-load level conditions to some extent. Combining ReLU with dropout offers no performance improvement over ReLU separately. By contrast, the combination of SeLU and  $\alpha$ -dropout provides a significant boost to the performance of fault diagnosis model. Consequently, it is confirmed that SeLU and  $\alpha$ -dropout can significantly enhance the generalization capability of the 1D-SCNN model.



**Figure 12.** Effects of SeLU and  $\alpha$ -dropout on the generalization capability of 1D-SCNN.

## 5. Conclusions

To improve the fault identification accuracy of rotating machinery under cross-load level conditions, a 1D-SCNN-based fault diagnosis method is proposed. The method involves feature extraction based on the frequency spectrum of the raw vibration signals. By introducing the self-normalizing properties of SNN into 1D-CNN through SeLU, and applying  $\alpha$ -dropout twice to regularize the model, the accuracy and generalization capability of the fault diagnosis model are greatly enhanced. Furthermore, the model is configured with a simple and compact architecture, which has a good computational complexity. The experimental results demonstrate that the proposed method not only can possess high fault identification accuracy under the same load level, but also should be able to achieve good performance under cross-load level conditions. Given that in actual engineering

situations, it is tricky to obtain ideal data under all working conditions of rotating machinery due to constant changes in speed and load level, the proposed method can utilize the data collected from limited working conditions to build a fault diagnosis model to identify fault types under other working conditions, which has significant engineering application value.

It should be noted that the actual running conditions of the rotating machinery are very complex; thus, it is impossible to obtain a training dataset including all the conditions. Considering the diversity of real engineering situations, future work will be devoted to taking more validation scenarios into account to test the generalization capability of the proposed method. For example, training the model using the datasets of A and C, and then testing the performance of the model using the datasets of B and D. Furthermore, most time-domain features and the simple frequency spectrum are insensitive for bearing fault diagnosis in actual situations. Considering the complexity of real engineering situations, we will try some other features, such as the squared envelope or the squared envelope spectrum, in our future work.

**Author Contributions:** Conceptualization, J.Y.; formal analysis, S.Y.; funding acquisition, J.Y.; methodology, J.Y.; resources, T.G.; software, T.G.; supervision, J.Y.; visualization, S.Y.; writing—original draft, S.Y.; writing—review and editing, Y.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Acknowledgments:** The authors would like to thank the editors and the reviewers for their helpful comments.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Frank, P.M. Analytical and Qualitative Model Based Fault Diagnosis—A survey and Some New Results. *Eur. J. Control* **1996**, *2*, 6–28. [[CrossRef](#)]
2. Gao, Z.; Cecati, C.; Ding, S.X. A Survey of Fault Diagnosis and Fault-Tolerant Techniques—Part II: Fault Diagnosis With Knowledge-Based and Hybrid/Active Approaches. *IEEE Trans. Ind. Electron.* **2015**, *62*, 3768–3774 [[CrossRef](#)]
3. Cheng, Z.; Hu, N. Quantitative damage detection for planetary gear sets based on physical models. *Chin. J. Mech. Eng.* **2012**, *25*, 1–7. [[CrossRef](#)]
4. Park, J.; Ha, J.M.; Oh, H.; Youn, B.D.; Choi, J.; Kim, N.H. Model-Based Fault Diagnosis of a Planetary Gear: A Novel Approach Using Transmission Error. *IEEE Trans. Reliab.* **2016**, *65*, 1830–1841. [[CrossRef](#)]
5. Kandukuri, S.T.; Klausen, A.; Karimi, H.R.; Robbersmyr, K.G. A review of diagnostics and prognostics of low-speed machinery towards wind turbine farm-level health management. *Renew. Sustain. Energy Rev.* **2016**, *53*, 697–708. [[CrossRef](#)]
6. Gao, Z.; Cecati, C.; Ding, S.X. A survey of fault diagnosis and fault-tolerant techniques—Part I: Fault diagnosis with model-based and signal-based approaches. *IEEE Trans. Ind. Electron.* **2015**, *62*, 3757–3767. [[CrossRef](#)]
7. Teng, W.; Ding, X.; Zhang, X.; Liu, Y.; Ma, Z. Multi-fault detection and failure analysis of wind turbine gearbox using complex wavelet transform. *Renew. Energy* **2016**, *93*, 591–598. [[CrossRef](#)]
8. Feng, Z.; Qin, S.; Liang, M. Time–frequency analysis based on Vold-Kalman filter and higher order energy separation for fault diagnosis of wind turbine planetary gearbox under nonstationary conditions. *Renew. Energy* **2016**, *85*, 45–56. [[CrossRef](#)]
9. Lei, Y.; Jia, F.; Kong, D.; Lin, J.; Xing, S. Opportunities and Challenges of Machinery Intelligent Fault Diagnosis in Big Data Era. *J. Mech. Eng.* **2018**, *54*, 94–104. [[CrossRef](#)]
10. Knapp, G.M.; Wang, H.P. Machine fault classification: A neural network approach. *Int. J. Prod. Res.* **1992**, *30*, 811–823. [[CrossRef](#)]
11. Zhu, X.; Zhang, Y.; Zhu, Y. Intelligent fault diagnosis of rolling bearing based on kernel neighborhood rough sets and statistical features. *J. Mech. Sci. Technol.* **2012**, *30*, 2649–2657. [[CrossRef](#)]
12. Sun, R.; Yang, Z.; Luo, W.; Qiao, B.; Chen, X. Weighted sparse representation based on failure dynamics simulation for planetary gearbox fault diagnosis. *Meas. Sci. Technol.* **2012**, *30*, 1361–6501. [[CrossRef](#)]
13. Wang, C.; Gan, M.; Zhu, C. Fault feature extraction of rolling element bearings based on wavelet packet transform and sparse representation theory. *J. Intell. Manuf.* **2015**, *29*, 937–951. [[CrossRef](#)]

14. Deng, W.; Yao, R.; Sun, M.; Zhao, H.M.; Luo, Y.L.; Dong, C. Study on a novel fault diagnosis method based on integrating EMD, fuzzy entropy, improved PSO and SVM. *J. Vibroeng.* **2017**, *19*, 2562–2577.
15. Wu, J.; Tang, T.; Chen, M.; Hu, T. Self-Adaptive Spectrum Analysis Based Bearing Fault Diagnosis. *Sensors* **2018**, *18*, 3312. [[CrossRef](#)] [[PubMed](#)]
16. Chen, X.; Zhou, J.; Xiao, H.; Wang, E.; Xiao, J.; Zhang, H. Fault diagnosis based on comprehensive geometric characteristic and probability neural network. *Appl. Math. Comput.* **2014**, *230*, 542–554. [[CrossRef](#)]
17. Pandya, D.H.; Upadhyay, S.H.; Harsha, S.P. Fault diagnosis of rolling element bearing with intrinsic mode function of acoustic emission data using APFKNN. *Expert Syst. Appl.* **2013**, *40*, 4137–4145. [[CrossRef](#)]
18. Jia, F.; Lei, Y.; Lin, J.; Zhou, X.; Lu, N. Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. *Mech. Syst. Signal Process.* **2016**, *72*, 303–315. [[CrossRef](#)]
19. Wang, J.; Li, S.; Han, B.; An, Z.; Xin, Y.; Qian, W.; Wu, Q. Construction of a batch-normalized autoencoder network and its application in mechanical intelligent fault diagnosis. *Meas. Sci. Technol.* **2018**, *30*, 1361–6501. [[CrossRef](#)]
20. Yang, Y.; Fu, P.; He, Y. Bearing fault automatic classification based on deep learning. *IEEE Access.* **2018**, *6*, 71540–71554. [[CrossRef](#)]
21. Shen, C.; Qi, Y.; Wang, J.; Cai, G.; Zhu, Z. An automatic and robust features learning method for rotating machinery fault diagnosis based on contractive autoencoder. *Eng. Appl. Artif. Intell.* **2018**, *76*, 170–184. [[CrossRef](#)]
22. Yin, J.; Zhao, W. Fault diagnosis network design for vehicle on-board equipments of highspeed railway: A deep learning approach. *Eng. Appl. Artif. Intell.* **2016**, *56*, 250–259. [[CrossRef](#)]
23. Gan, M.; Wang, C.; Zhu, C. Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings. *Mech. Syst. Signal Process.* **2016**, *72*, 92–104. [[CrossRef](#)]
24. Chen, Z.; Li, W. Multisensor Feature Fusion for Bearing Fault Diagnosis Using Sparse Autoencoder and Deep Belief Network. *IEEE Trans. Instrum. Meas.* **2016**, *66*, 1693–1702. [[CrossRef](#)]
25. Shang, Z.; Liao, X.; Geng, R.; Gao, M.; Liu, X. Fault diagnosis method of rolling bearing based on deep belief network. *J. Mech. Sci. Technol.* **2018**, *32*, 5139–5145. [[CrossRef](#)]
26. Lecun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]
27. Bengio, Y. *Learning Deep Architectures for AI*; Foundations and Trends in Machine Learning Series; Now Publishers Inc.: Hanover, MA, USA, 2009; Volume 2, 1–127.
28. Wen, L.; Li, X.; Gao, L.; Zhang, Y. A New Convolutional Neural Network Based Data-Driven Fault Diagnosis Method. *IEEE Trans. Ind. Electron.* **2017**, *65*, 5990–5998. [[CrossRef](#)]
29. Guo, S.; Yang, T.; Gao, W.; Zhang, C. A Novel Fault Diagnosis Method for Rotating Machinery Based on a Convolutional Neural Network. *Sensors* **2018**, *18*, 1429. [[CrossRef](#)]
30. Zhao, D.; Wang, T.; Chu, F. Deep convolutional neural network based planet bearing fault classification. *Comput. Ind.* **2019**, *107*, 59–66. [[CrossRef](#)]
31. Sun, Y.; Gao, H.; Guo, L.; Hong, X.; Song, X.; Zhang, J. A New Intelligent Fault Diagnosis Method and Its Application on Bearings. In Proceedings of the 13th International Conference on Damage Assessment of Structures, Porto, Portugal, 9–10 July 2019; pp. 618–628.
32. Palaz, D.; Collobert, R.; Doss, M.M. Estimating phoneme class conditional probabilities from raw speech signal using convolutional neural networks. *Comput. Sci.* **2013**, 1766–1770.
33. Kim, Y. Convolutional neural networks for sentence classification. *arXiv* **2014**, arXiv:1408.5882.
34. Liu, X.; Zhou, Q.; Zhao, J.; Shen, H.; Xiong, X. Real-time and anti-noise fault diagnosis algorithm based on 1-D convolutional neural network. *J. Harbin Inst. Technol.* **2019**, *51*, 89–95.
35. Levent, E.; Turker, I.; Serkan, K.A. Generic Intelligent Bearing Fault Diagnosis System Using Compact Adaptive 1D CNN Classifier. *J. Signal Process. Syst.* **2019**, *91*, 179–189.
36. Zhang, W.; Peng, G.; Li, C.; Chen, Y. A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. *Sensors* **2017**, *17*, 425. [[CrossRef](#)] [[PubMed](#)]
37. An, J.; Ai, P.; Xu, S.; Liu, C.; Xia, J.; Liu, D. Intelligent fault diagnosis method for rotating machinery based on one dimensional convolution neural network. *J. Nanjing Univ. Nat. Sci.* **2019**, *55*, 133–142.
38. Jing, L.; Zhao, M.; Li, P.; Li, X. A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox. *Measurement* **2017**, *111*, 1–10. [[CrossRef](#)]

39. Liu, T.; Zhou, D.; Li, M.; Li, X. Multi-Resolution Fusion Convolutional Neural Network for Screw Locking Series. *J. Xi'an Jiaotong Univ.* **2020**, *54*, 161–178.
40. Klambauer, G.; Unterthiner, T.; Mayr, A.; Hochreiter, S. Self-Normalizing Neural Networks. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 971–980.
41. Srivastava, N.; Geoffrey, H.; Alex, K.; Ilya, S.; Ruslan, S.; Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
42. Zhang, W.; Li, C.; Peng, G.; Chen, Y.; Zhang, Z. A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. *Mech. Syst. Signal Process.* **2018**, *100*, 439–453. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).