End-to-End Continuous/Discontinuous Feature Fusion Method with Attention for Rolling Bearing Fault Diagnosis

Jianbo Zheng 1,2, Jian Liao 1,2,* and Zongbin Chen 1,2

1 Institute of Vibration and Noise, Naval University of Engineering, Wuhan 430033, China
2 Naval Key Laboratory of Ship Vibration and Noise, Naval University of Engineering, Wuhan 430033, China
* Correspondence: jl_zss@163.com

Abstract: Mechanical equipment failure may cause massive economic and even life loss. Therefore, the diagnosis of the failures of machine parts in time is crucial. The rolling bearings are one of the most valuable parts, which have attracted the focus of fault diagnosis. Many successful rolling bearing fault diagnoses have been made based on machine learning and deep learning. However, most diagnosis methods still rely on complex signal processing and artificial features, bringing many costs to the deployment and migration of diagnostic models. This paper proposes an end-to-end continuous/discontinuous feature fusion method for rolling bearing fault diagnosis (C/D-FUSA). This method comprises long short-term memory (LSTM), convolutional neural networks (CNN) and attention mechanism, which automatically extracts the continuous and discontinuous features from vibration signals for fault diagnosis. We also propose a contextual-dependent attention module for the LSTM layers. We compare the method with the other simpler deep learning methods and state-of-the-art methods in rolling bearing fault data sets with different sample rates. The results show that our method is more accurate than the other methods with real-time inference. It is also easy to be deployed and trained in a new environment.

Keywords: fault diagnosis; rolling bearing; deep learning; LSTM; CNN; attention

1. Introduction

The fault diagnosis of mechanical equipment is vital in modern industry. Once the failure of mechanical equipment occurs, it will cause huge damage to the economy and property and even bring casualties. Therefore, finding a better fault diagnosis method is necessary to ensure the normal operation of the machine [1]. Rolling bearing is the most commonly used part in mechanical equipment, known as the joint of the machinery. It has the advantages of high efficiency, small friction resistance, convenient assembly and easy lubrication, so it is widely used in rotating machinery. As one of the core components of rotating machinery, such as gearbox and turbine machinery, the health of rolling bearings significantly influences the machine’s stability and life [2]. In the process of working, rolling bearings may be damaged by the outer raceway, inner raceway and rolling body due to lubricant pollution, overload and other reasons. Therefore, an effective fault diagnosis method is crucial to the stability of rolling bearings [3].

Most fault diagnosis methods used for rolling bearings are based on vibration signals. By detecting and analyzing the vibration data of rolling bearings, this method can diagnose various faults in real time [4]. Model-based [5] and data-driven [6] methods can diagnose rolling bearing faults based on vibration data. Data-driven methods generally use machine learning to learn bearing vibration data and identify different types of fault modes. This method can effectively and quickly process mechanical signals, requires less prior expertise, and can provide an accurate diagnosis [7]. It has become a common method for rolling bearing fault diagnosis. K-nearest neighbor (KNN), support vector machine (SVM), self-organizing mapping (SOM) networks and other machine learning algorithms have been
successfully applied to the intelligent diagnosis of rolling bearing faults [8–10]. Fault diagnosis methods based on traditional machine learning need to extract the artificial features to enhance the diagnosis performance, such as time-domain statistical features (root mean square, kurtosis, skewness and spectral kurtosis) [11–15], fast Fourier transform (FFT) spectrum, power spectrum, empirical mode decomposition (EMD) features, variational mode decomposition (VMD) features and other features [16–18]. For example, Xiao et al. proposed a fault diagnosis method based on the kurtosis criterion VMD and an SOM neural network. The method used the VMD algorithm to decompose the gear vibration signal, selected the intrinsic mode functions (IMFs) most relevant to the original vibration signals according to the average instantaneous frequency, and extracted the kurtosis of IMFs as the feature. This feature has been proven to diagnose gear faults accurately [10]. Song et al. proposed a feature-extracting method for rolling bearing vibration fault signals combining statistical filtering (SF), wavelet packet transform (WPT) and the moving peak value preservation (M-PH) method, which could identify fault types based on bearing diagnostic features in the frequency domain. This decision tree trained on the features could accurately diagnose the fault types of rolling bearings [18]. However, most rolling bearing fault diagnosis algorithms based on traditional machine learning rely on complex and well-designed signal feature extraction methods to accurately diagnose. It is necessary to adapt the feature extraction algorithm for specific situations or bearing types to achieve an accurate diagnosis. Therefore, this diagnosis method is susceptible to specific rolling bearings, experimental environment and feature adaptability, and it is difficult to achieve a general rolling bearing fault diagnosis system [19,20].

In recent years, deep learning [21–25] has been successfully applied in computer vision [26], natural language processing [27], speech recognition [28], mechanical control [29] and many other fields. This kind of algorithm integrates feature engineering and pattern recognition through the combination of multilayer artificial neural network layers. It can extract suitable features adaptively without complex artificial features for most data training. Therefore, various model structures of deep learning algorithms have been applied to rolling bearing fault diagnosis and achieved good performance, including deep belief networks (DBN), fully connected networks, long short-term memory (LSTM) models [30], convolutional neural networks (CNNs), transformers and others [7,31–36]. These studies proposed deep learning models with different structures or multimodal features to diagnose bearing faults. For example, Shao et al. proposed an optimized DBN for rolling bearing fault diagnosis. They used the stochastic gradient descent algorithm based on the energy functions to fine-tune all of the connection weights of the constrained Boltzmann machine (RBM) after pretraining effectively. The method improved the classification accuracy of DBN. Furthermore, an optimal DBN was designed using the particle swarm optimization algorithm to determine the best structure of the trained DBN. The method was applied to the simulation signal and experimental signal analysis of a rolling bearing and achieved higher precision and robustness than other intelligent methods [31]. Chen et al. proposed a model composed of a multiscale CNN and an LSTM (MCNN-LSTM) for rolling bearing fault diagnosis. This model consisted of a feature extractor and a classifier, allowing raw data to be input directly into the model without preprocessing. The feature extractor contained two CNNs with different kernel sizes, which were used to automatically extract the feature representation of the rolling bearing fault vibration signals. The extracted features were then input into a stacked LSTM network for bearing fault assessment [36]. The proposed method achieved 98.46% accuracy, which exceeded some state-of-the-art algorithms. Ding et al. proposed a new time-frequency transformer (TFT) model. This model adopted synchrosqueezed wavelet transform (SWT) [37] to obtain time-frequency representation. They designed an encoder composed of transformer blocks to construct the hidden time-frequency representation for rolling bearing faults. The method was proven to be superior in comparison with other methods [7]. Although the above algorithms and methods have achieved satisfactory results, there are some limitations.
1. Most current studies still need various signal processing methods to extract features. Therefore, the effectiveness of fault diagnosis heavily depends on the quality of manually extracted features. A suitable intelligent diagnosis algorithm is needed for adaptive feature extraction and selection.

2. Some deep learning algorithms still need to cooperate with many complex signal processing methods to adapt to rolling bearing fault detection. These methods have a lot of manual parameters to adjust and are difficult to deploy and train.

3. Some deep learning models improve the diagnosis effect by combining overly complex structures, but this usually increases the cost of calculation and the risk of overfitting.

In order to improve the above limitations, this paper proposes an end-to-end continuous/discontinuous feature fusion method with CNN, LSTM and attention (C/D-FUSA) for rolling bearing fault diagnosis, which does not require complex data processing and feature engineering. The proposed method mainly combines multilayer CNN and a multilayer LSTM network for adaptive feature extraction and the fusion of different features. A multilayer CNN network extracts the discontinuous large-scale features of rolling bearing vibration signals with several one-dimensional convolutional layers and one-dimensional pooling layers. Multilayer LSTM networks extract continuous features with short- and long-term memory by composing multiple bidirectional LSTM layers. Rolling bearing faults can be diagnosed by the fusion of continuous and discontinuous features. This work has the following contributions:

1. An end-to-end deep learning model is proposed for rolling bearing fault diagnosis. Without manual design features or complex data processing, this model can accurately extract and screen continuous/discontinuous signal features to diagnose rolling bearing faults.

2. Compared with the simple deep learning model, the proposed model has a higher classification accuracy (99.87%), and the inference time does not significantly increase.

3. This method could be easily deployed and migrated to a new environment or a new type of rolling bearing because of the absence of complex data processing and manual feature engineering.

4. The proposed model can achieve accurate multitype fault diagnosis, and the experiment proved that it could accurately diagnose 10 types of working states of rolling bearings.

The rest of this article is arranged as follows. Section 2 introduces the background and the details of the C/D-FUSA approach proposed by this work. The experimental setup and results are described in Section 3. The experimental results are discussed in Section 4. Finally, we give the conclusion in Section 5.

2. Materials and Methods

2.1. Types of Rolling Bearing Faults

A typical rolling bearing structure is shown in Figure 1, including an outer race mounted on a bearing house, an inner race mounted on a rotating shaft, a rolling element, and a supporting cage [38]. Rolling bearings are the most easily damaged parts in rotating machinery. Any damage will lead to the rapid degradation of the bearing due to wear and tear of the inner ball. There are many reasons for bearing faults, including improper size selection, leakage, excessive load and others. Once a bearing has a fault, it will produce periodic abnormal vibrations, whose amplitude is determined by the type of fault. The faults could happen in the bearings at the drive end or fan end. In most situations, the fault happens at the inner raceway, rolling element (ball) and outer raceway of a bearing. The faults usually range from 0.007 inches to 0.040 inches in diameter [39]. The types of faults can be diagnosed by the frequency features calculated by fault size, shaft speed, the load, and fault location, such as the inner circle rotation frequency, relative rotation frequency
of the inner and outer circles, cage rotation frequency and the revolution frequency of the rolling body [40]. Some of the vibration frequencies are listed as follows:

\[
    f_i = \frac{n}{60}
\]  

(1)

\[
    f_r = f_i - f_0
\]  

(2)

\[
    f_c = 0.381 - 0.4 \times f_r
\]  

(3)

\[
    f_b = \frac{1}{2} \left(1 - \frac{d}{D \cos \alpha}\right) f_r
\]  

(4)

\[
    f_{be} = \frac{1}{2} N_b \left(1 - \frac{d}{D \cos \alpha}\right) f_r
\]  

(5)

\[
    f_{bi} = \frac{1}{2} N_b \left(1 + \frac{d}{D \cos \alpha}\right) f_r
\]  

(6)

![Figure 1. Rolling bearing structure, including outer race, inner race, rolling element and supporting cage.](image)

The \( n \) is the speed of bearing inner circle; \( f_i \) is the inner circle rotation frequency; \( f_0 \) is the outer circle rotation frequency; \( f_r \) is the relative rotation frequency of the inner and outer circles; \( f_c \) is the cage rotation frequency; \( f_b \) is the revolution frequency of the rolling body; and \( f_{be} \) and \( f_{bi} \) are the outer race and inner race fault frequencies, respectively. The frequency domain features obtained by FFT can also help to diagnose faults. This was used in the cepstral analysis, which takes the logarithm of the Fourier transform of the original waveform and then takes the inverse Fourier transform of the logarithm [41]. The inverse Fourier transform results could be different for the normal and fault states, as shown in Figure 2.

In traditional methods, these features are transformed into specific numerical forms, such as kurtosis, mean and variance, and then fed into machine learning algorithms for classification. Although the frequency features summarized from experiences can be used to infer the specific fault types or locations of the rolling bearings in some cases, the fault diagnosis method based on manually designed features is difficult for widespread use. The end-to-end diagnosis model based on deep learning proposed in this work does not depend on any experience, data statistics or processing, and can directly infer specific fault types from raw data.

2.2. Framework of the Proposed C/D-FUSA

This paper proposes an end-to-end continuous/discontinuous feature fusion method with CNN, LSTM and attention (C/D-FUSA) for rolling bearing fault diagnosis. The framework of the proposed C/D-FUSA is shown in Figure 3. This method contains three subnetworks. (1) The subnet for the continuous features is designed to extract sequential associated continuous features using LSTM layers with a context-dependent attention
mechanism. (2) The subnet for the discontinuous features is designed to extract different level discontinuous features using convolutional layers with an attention mechanism. (3) The continuous and discontinuous features are concatenated and used to calculate the probability of different fault or normal states of rolling bearings in the subnet for classification.

Figure 2. Signal cepstral analysis diagram for a normal state and a fault state.

2.2. Framework of the Proposed C/D-FUSA

This paper proposes an end-to-end continuous/discontinuous feature fusion method with CNN, LSTM and attention (C/D-FUSA) for rolling bearing fault diagnosis. The framework of the proposed C/D-FUSA is shown in Figure 3. This method contains three subnetworks. (1) The subnet for the continuous features is designed to extract sequential associated continuous features using LSTM layers with a context-dependent attention mechanism. (2) The subnet for the discontinuous features is designed to extract different level discontinuous features using convolutional layers with an attention mechanism. (3) The continuous and discontinuous features are concatenated and used to calculate the probability of different fault or normal states of rolling bearings in the subnet for classification.

Figure 3. Framework for the proposed end-to-end continuous/discontinuous feature fusion method for rolling bearing fault diagnosis. This method contains three different subnetworks for the continuous features, the discontinuous features and classification. The subnetwork for discontinuous features used several convolutional modules with attention on extracting abstract discontinuous local features. The subnetwork for continuous features extracts the features used an LSTM layer with attention. The discontinuous and continuous features were concatenated and put into the subnetwork for classification, which contains several fully connected layers, to output the probability of each category.

2.3. Subnet for Continuous Features

The subnet for the continuous features is composed of LSTM layers with an attention mechanism. We proposed a context-dependent attention module for weighting the output of the features by the LSTM layer according to previous outputs.
2.3.1. Long Short-Term Memory

LSTM is a network with a long-term memory function that is improved based on the recurrent neural network (RNN) [42], as shown in Figure 4. Due to its characteristics, LSTM is widely used in fitting time series, which is very important for natural language processing, speech recognition, handwriting recognition and other applications. An LSTM cell comprises a forget gate, an input gate and an output gate. The forget gate controls information in the previous time steps remaining in the current cell state. The input gate controls the new information input into the current cell state. Finally, the output gate, based on the cell states, controls the output of the LSTM cell. These three gates are described in Equations (7)–(9):

\[
f_t = \sigma(W_f^T \cdot [h_{t-1}, x_t] + b_f) \tag{7}
\]

\[
i_t = \sigma(W_i^T \cdot [h_{t-1}, x_t] + b_i) \tag{8}
\]

\[
O_t = \sigma(W_o^T \cdot [h_{t-1}, x_t] + b_o) \tag{9}
\]

![Figure 4. An LSTM cell.](image)

In the above three equations, \( f_t \) is the output of the forget gate at time step \( t \); \( i_t \) is the output of the input gate; \( O_t \) is the output of the output gate; \( W_f, W_i, \) and \( W_o \) are trainable weight matrices; \( b_f, b_i, \) and \( b_o \) are the trainable bias in these gates; \( t-1 \) represents the previous time step; \( h_{t-1} \) is the hidden state at the previous time step; \( x_t \) is the input features at the time step \( t \); and \( \sigma \) is the sigmoid function. After information filtering through the forget gate and input gate, the LSTM cell state is adjusted by Equations (10) and (11):

\[
C_t = \tanh(W_c^T \cdot [h_{t-1}, x_t] + b_c) \tag{10}
\]

\[
c_t = f_t \times C_{t-1} + i_t \times C_t \tag{11}
\]

where \( C_t \) is the information for updating the cell state; \( W_c \) is the trainable weight matrix; \( b_c \) is the trainable bias; and \( c_t \) is the updated cell state controlled by the forget gate and the input gate. Finally, the hidden state at the current time step is defined by:

\[
h_t = O_t \times \tanh(c_t) \tag{12}
\]

2.3.2. Context-Dependent Attention

Since the output of the last time step of the LSTM network in C/D-FUSA was used for the fault diagnosis, the attention module based on time series was unavailable [43]. Therefore, this work proposes a context-dependent attention module for LSTM. The structure of the context-dependent attention module is shown in Figure 5. This module calculates the attention weights of the output features by integrating the outputs of \( L \) time points before
the last time point. First of all, the average output features of the contextual outputs are calculated by:

\[ o_m = \frac{\sum_{t=T-L}^{T-1} O_t}{L}, \quad o_m \in \mathbb{R}^{n \times 1} \]  

(13)

2.3.2. Context-Dependent Attention

Since the output of the last time step of the LSTM network in C/D-FUSA was used for the fault diagnosis, the attention module based on time series was unavailable [43]. Therefore, this work proposes a context-dependent attention module for LSTM. The structure of the context-dependent attention module is shown in Figure 5. This module calculates the attention weights of the output features by integrating the outputs of \( L \) time points before the last time point. First of all, the average output features of the contextual outputs are calculated by:

\[ o_m = \frac{\sum_{t=T-L}^{T-1} O_t}{L}, \quad o_m \in \mathbb{R}^{n \times 1} \]  

(13)

\[ h_1 = O_1 \times \tanh (c_1) \]  

(12)

Figure 5. Context-dependent attention module for LSTM.

In the equation, \( o_m \) is the average output features of the contextual outputs from the time step \( T - L \) to \( T \); \( T \) is the number of time steps; \( L \) is the length of the contextual window; and \( n \) is the number of output features in the LSTM layer. Then, the average output features are converted into an attention weight vector by Equations (14) and (15):

\[ a_l = g(W_{l2} \times (W_{l1} \times o_m)) \]  

(14)

where \( a_l \) is the attention weight vector; \( W_{l1} \) is a trainable weight matrix, \( W_{l1} \in \mathbb{R}^{k \times n} \); \( W_{l2} \) is a trainable weight matrix, \( W_{l2} \in \mathbb{R}^{n \times k} \); and \( g \) is the softmax function. The softmax function converts the result of the products to the range between 0 and 1. Finally, the output of LSTM is weighted by the attention weight vector by:

\[ o_{l,T} = a_l o_T \]  

(15)

where \( o_{l,T} \) is the result vector of an LSTM layer with context-dependent attention; and \( o_T \) is the output of the LSTM layer at the time step \( T \). The \( T \) used in the proposed method was equal to the length of the input vibration signals. The number of output features \( n \) was 128.

2.4. Subnet for Discontinuous Features

The subnet for the discontinuous features is composed of several convolutional modules with attention. As shown in Figure 6, the convolutional module comprises the sequentially connected 1D convolutional layer, batch normalization operation, ReLU activation function, and squeeze-and-excitation (SE) attention module [44] and 1D max-pooling layer.
The convolutional layer in the convolutional module is calculated with the same padding and the kernel size is 3. The kernel of the max-pooling layer is 2. Therefore, for an \( N \times M_i \) input matrix, the convolutional module could obtain an \( \frac{N}{2} \times M_o \) output matrix, where \( N \) is the length of the input matrix and \( M_i \) is the number of input features. \( M_o \) is the number of output features.

The SE attention module is one of the most popular attention modules in computer vision. In this module, the input matrix is averaged first, as described in Equation (16):

\[
P_{m,j} = \frac{\sum_{i=1}^{N} p_{i,j}}{N}, \quad 1 \leq j \leq M
\]  

where \( P_{m,j} \) is the average value for feature \( j \); \( p_{i,j} \) is the feature value of the input matrix; and \( P_m \) is the average vector, \( P_m \in \mathbb{R}^{M \times 1} \). Then, the \( C_m \) is squeezed by:

\[
C_s = \text{ReLU}(W_s \times P_m + b_s)
\]  

where \( C_s \) is the squeezed vector, \( C_s \in \mathbb{R}^{S \times 1} \); \( W_s \) is the trainable weight matrix, \( W_s \in \mathbb{R}^{S \times M} \); and \( b_s \) is the trainable bias. After the squeeze, the vector is then excited by:

\[
C_e = \sigma(W_e \times C_s + b_e)
\]  

where \( C_e \) is the excited vector, \( C_e \in \mathbb{R}^{M \times 1} \); \( W_e \) is the trainable weight matrix, \( W_e \in \mathbb{R}^{M \times S} \); \( b_e \) is the trainable bias; and \( \sigma \) is the sigmoid function. Finally, the original output \( P \) of the previous layer is weighted by \( C_e \):

\[
P_a = P \cdot C_e
\]

where \( P_a \) is the output of the attention module.

The C/D-FUSA uses three convolutional modules in the subnet for the discontinuous features. In the first convolutional module, the number of output features in each module is 32, 16 and 8, respectively. The squeeze coefficient \( S \) in each module is 8, 4 and 2, respectively.

2.5. Subnet for Classification

In the subnet for classification, the outputs of the discontinuous and continuous features by convolutional modules and the LSTM module with attention are concatenated. However, the output feature number of the convolutional modules (\( T \)) was much larger than that of the LSTM module 128. Too many discontinuous features could impair the diagnosis performance. Therefore, the output of the final convolutional module \( o_c \) is converted by:

\[
o'_c = \sigma(W_c \times o_c + b_c)
\]  

where \( o'_c \) is a vector with 16 features and \( W_c \) is a trainable matrix, \( W_c \in \mathbb{R}^{16 \times T} \). After the concatenation of \( o'_c \) and \( o_{l,T} \), two fully connected layers were used for classification. The first fully connected layer converted the 144 concatenated features into 64 features. The
last fully connected layer converted the 64 features into $N_C$ values. The $N_C$ is the number of categories. Finally, a softmax function converted these $N_C$ values into $N_C$ probabilities, ranging from 0 to 1. Each of the values represented the probability of a fault state or normal state of the rolling bearing.

3. Experiments and Results

3.1. Experiment Setups

We tested the proposed C/D-FUSA method on the experimental data set from the rolling bearing data center of Case Western Reserve University (CWRU) [45]. The CWRU data set was collected using a controllable motor of 2 HP (power: 1.5 kW), and the acceleration data measurement could be placed near or away from the motor bearing. The experimental motor’s actual test conditions and bearing failure state were recorded in the data set.

In this experiment, we tested the diagnosis performance of the C/D-FUSA for the drive-end bearing fault data at 12,000 samples/s. The details of the tested data set are shown in Table 1. This experiment considered ten fault types, including one normal type (no-fault) and nine fault types. The nine fault types included three different locations (ball, inner raceway, outer raceway) and three fault diameters (0.007 mm, 0.014 mm, 0.021 mm). The details of the ten types are listed below:

1. Normal type: no fault was found in these samples;
2. Location = Ball, Diameter = 0.007: the fault occurred on the ball, the fault diameter was 0.007 in;
3. Location = Ball, Diameter = 0.014: the fault occurred on the ball, the fault diameter was 0.014 in;
4. Location = Ball, Diameter = 0.021: the fault occurred on the ball, the fault diameter was 0.021 in;
5. Location = Inner Raceway, Diameter = 0.007: the fault occurred on the inner raceway, the fault diameter was 0.007 in;
6. Location = Inner Raceway, Diameter = 0.014: the fault occurred on the inner raceway, the fault diameter was 0.014 in;
7. Location = Inner Raceway, Diameter = 0.021: the fault occurred on the inner raceway, the fault diameter was 0.021 in;
8. Location = Outer Raceway, Diameter = 0.007: the fault occurred on the outer raceway, the fault diameter was 0.007 in;
9. Location = Outer Raceway, Diameter = 0.014: the fault occurred on the outer raceway, the fault diameter was 0.014 in;
10. Location = Outer Raceway, Diameter = 0.021: the fault occurred on the outer raceway, the fault diameter was 0.021 in.

Table 1. Details of the drive-end bearing fault data at 12,000 samples/s tested in our experiment.

| Fault Location | Diameter | Number of Samples | Number of Samples in CWRU-512 | Number of Samples in CWRU-6000 |
|----------------|----------|-------------------|-----------------------------|-------------------------------|
| Normal Normal  | 0.007    | 487,093           | 950                         | 80                            |
|                | 0.014    | 488,109           | 951                         | 80                            |
|                | 0.021    | 487,964           | 951                         | 80                            |
| Ball           | 0.007    | 488,309           | 952                         | 80                            |
|                | 0.014    | 487,239           | 948                         | 80                            |
|                | 0.021    | 487,529           | 950                         | 80                            |
| Inner Raceway  | 0.007    | 1,465,051         | 2855                        | 240                           |
|                | 0.014    | 487,819           | 950                         | 80                            |
|                | 0.021    | 1,465,487         | 2856                        | 240                           |
| Outer Raceway  | 0.007    | 487,964           | 951                         | 80                            |
|                | 0.014    | 487,529           | 950                         | 80                            |
|                | 0.021    | 1,465,051         | 2855                        | 240                           |
The numbers of samples with different types ranged from 487,093 to 2,182,450. To evaluate the performance of the proposed method under high- and low-frequency detection, we sampled 512 and 6000 consecutive and non-overlapping vibration signals as one sample and created new data sets: CWRU-512 and CWRU-6000, for training and diagnosis evaluation. The CWRU-512 contains 948–4261 samples for each category. The CWRU-6000 contains 80–360 samples for each category.

In the experiment, the number of categories \( N_C \) was set as 10. The number of time steps \( T \) was set as 512 in CWRU-512 and \( T \) was 6000 in CWRU-6000. The subnet for the discontinuous features contained three convolutional modules and two LSTM layers were used in the subnet for the continuous features.

We conducted three comparisons in this work. In the first comparison, we compared the performance of the proposed C/D-FUSA with the end-to-end continuous/discontinuous feature fusion method without attention (C/D-FUS) and LSTM models. The C/D-FUS has the same structure as the C/D-FUSA, which did not use the context-dependent attention module for LSTM layers and SE attention modules for convolutional modules. The LSTM models contained two LSTM layers with 128 hidden cells. The features output in the last time step \( T \) were used for classification. Two fully connected layers followed the LSTM layers for classification. The first fully connected layer converted 128 features into 64 features. The last fully connected layer converted 64 features into 10 values. Like the C/D-FUSA, the LSTM model used the softmax function to convert the 10 values into 10 probabilities for each fault type. We used five-fold cross-validation for each model and each data set. We compared each method’s accuracy and loss curves and calculated the average accuracy, macro-precision, macro-recall and macro-f1-score for each model. In the second comparison, we compared the inference time for each model. Finally, in the last experiment, we compared the performance of C/D-FUSA with the state-of-the-art methods proposed in the other studies.

The accuracy, precision, recall and F1 scores were calculated in each data set. These measurements were calculated using Equations (21)–(24):

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FN + FP} \tag{21}
\]
\[
\text{Precision} = \frac{TP}{TP + FP} \tag{22}
\]
\[
\text{Recall} = \frac{TP}{TP + FN} \tag{23}
\]
\[
\text{F1 score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{24}
\]

where \( TP \), \( TN \), \( FN \), and \( FP \) are the numbers of true positive, true negative, false negative and false positive samples.

The hyper-parameters used in all methods trained in the CWRU-512 or CWRU-6000 are the same, as shown in Table 2. The models for CWRU-512 and CWRU-6000 were trained on 200 epochs with an Adaptive Moment Estimation (Adam) optimizer [46] and cross-entropy loss function. The learning rates for CWRU-512 and CWRU-6000 were 0.002 and 0.001, respectively. Moreover, the max sequence length for these two data sets was 512 and 6000, respectively. The experiments were conducted using Python 3.6 (Python Software Foundation, Wilmington, DE, USA, http://www.python.org) and PyTorch 1.8 (Facebook AI Research, New York, NY, USA, https://pytorch.org/) on an Nvidia RTX 2060 GPU (Nvidia Corporation, Santa Clara, CA, USA).

3.2. Results
3.2.1. Performance Comparison of Different Models

The accuracy and loss curves are shown in Figures 7 and 8. It can be found that the LSTM model in CWRU-512 needed at least 30 times of training to improve the classification
accuracy to about 90%, and this model in CRWU-6000 could not be well fitted even after 200 times of training. Compared with LSTM, the training performance of C/D-FUSA and C/D-FUS was significantly better, and the test classification accuracy of both data sets could stabilize at more than 95% after about 20 training sessions. However, the accuracy curve of C/D-FUS was less stable than that of C/D-FUSA, and the gap between the training and test accuracy of C/D-FUS was larger than that of C/D-FUSA, indicating that C/D-FUS was underfitting. In summary, through the analysis of the accuracy curves, it can be seen that the C/D-FUSA proposed by us is superior to the other two methods in terms of fitting speed, diagnosis performance and stability, no matter the data set.

### Table 2. Hyper-parameters used in each data set.

| Data Set   | Hyper Parameter       | Value                                      |
|------------|-----------------------|--------------------------------------------|
| CWRU-512   | Learning Rate         | 0.002                                      |
|            | Epoch                 | 200                                        |
|            | Max length            | 512                                        |
|            | Optimizer             | Adaptive Moment Estimation (Adam) [46]     |
|            | Loss Function         | Cross-entropy loss                         |
| CWRU-6000  | Learning Rate         | 0.001                                      |
|            | Epoch                 | 200                                        |
|            | Max length            | 6000                                       |
|            | Optimizer             | Adam                                       |
|            | Loss Function         | Cross-entropy loss                         |

![Accuracy curves for each model](image)

**Figure 7.** Accuracy curves for each model: (a–c) are the accuracy curves for C/D-FUSA, C/D-FUS and LSTM models for CWRU-512, respectively; (d–f) are the accuracy curves for C/D-FUSA, C/D-FUS and LSTM models for CWRU-6000, respectively.

We also calculated the accuracy, precision, recall and F1 scores for each model, shown in Table 3. According to this result, the diagnostic accuracy, precision, recall rate and F1 scores of C/D-FUSA were higher than the other methods in both CRWU-512 and CRWU-6000. This result indicates that the method proposed in this study performs better than the C/D-FUS method (the fusion of LSTM and CNN) and the LSTM method.
3.2.2. Performance Comparison with Other Studies

We compared the performance of C/D-FUSA with other state-of-the-art methods in the CRUW data set, including Sohaib et al., (2017) [33], Le et al., (2019) [34], Lei et al., (2016) [47], Wang et al., (2022) [48], Yan et al., (2022) [49], Zhang et al., (2022) [50] and Zhao et al., (2021) [51]. The results are shown in Table 4. It can be found that the C/D-FUSA method proposed by us has better diagnostic performance than all other algorithms. The diagnostic model based on data preprocessing, LSTM, CNN and the attention mechanism had the best performance except for C/D-FUSA [34]. However, it required additional data processing, such as converting data to a frequency spectrum and envelope spectrum, and it also needed 5000 training epochs to achieve 99.74% accuracy. However, the C/D-FUSA is an end-to-end diagnostic algorithm, which does not require any complex data conversion or processing. This advantage would make it easier for use in different situations of bearing fault diagnosis. Moreover, it only needs 30–200 times of training to achieve 99.87% accuracy. The number of training samples differs between all of the methods because of the sampling frequency. This method achieves the highest accuracy when using a similar number of training samples as other methods.

3.2.3. Inference Time Comparison with Other Studies

Although the performance of C/D-FUSA exceeds that of all methods compared in this work, the structure of C/D-FUSA is more complex than those of the C/D-FUS and...
LSTM models. Therefore, it is important that the inference speed of this model could be fast enough to be used in practical applications. We compared the inference time of the three models for a single sample, and the results are shown in Table 5. It can be found that although the inference time of C/D-FUSA for a single sample was longer than that of the other two models, the difference was not significant. Moreover, the single sample secondary sampling time of CRWU-512 is 0.043 s, and that of CRWU-6000 is 0.5 s. The inference time for one sample of C/D-FUSA is far shorter than the sampling time, indicating that the reasoning speed of the proposed algorithm can fully meet the requirements of practical use.

Table 4. Diagnosis performance comparison among the state-of-the-art methods and proposed method.

| Method                  | Type                                | Number of Fault Classes | Number of Training Samples | Accuracy (%) |
|-------------------------|-------------------------------------|-------------------------|---------------------------|--------------|
| Sohaib et al., (2017) [33] | Hybrid features + sparse stacked autoencoder | 10                      | 710                        | 99.10        |
| Li et al., (2019) [34]   | Preprocessing + attention + LSTM + CNN | 10                      | -                          | 99.74        |
| Lei et al., (2016) [47]  | Signal fraction + deep learning     | 10                      | 20,000                     | 99.66        |
| Wang et al., (2022) [48] | SSAE and softmax classifier         | 10                      | 4163                       | 99.15        |
| Yan et al., (2022) [49]  | Markov transition field and residual network | 10                      | 6600                       | 98.52        |
| Zhang et al., (2022) [50] | Transfer learning                   | 10                      | 9518                       | 99.80        |
| Zhao et al., (2021) [51] | Adaptation network with adversarial learning | 10                      | 7000                       | 99.24        |
| C/D-FUSA                | Attention + LSTM + CNN              | 10                      | 13,299                     | 99.85        |

Table 5. Inference time of each model in CRWU-512 and CRWU-6000.

| Data Set | Model  | Inference Time (s) |
|----------|--------|--------------------|
| CRWU-512 | C/D-FUSA | 0.034               |
|          | C/D-FUS | 0.031               |
|          | LSTM    | 0.028               |
| CRWU-6000| C/D-FUSA | 0.228               |
|          | C/D-FUS | 0.220               |
|          | LSTM    | 0.208               |

4. Discussion

In this work, we proposed an end-to-end continuous/discontinuous feature fusion method with CNN, LSTM and attention (C/D-FUSA) for rolling bearing fault diagnosis. We conducted experiments to test the diagnosis performance and the inference time of the proposed method.

In comparing the diagnosis performance, we compared C/D-FUSA with the other two simpler deep learning models and the state-of-the-art diagnosis methods in previous studies. According to the accuracy curves and loss curves of different models in the training process, it can be found that although the LSTM is found to be effective in some studies [52,53], these methods might require other optimization or extra data processing to be efficient. Therefore, the adaptability of the LSTM model is still insufficient, and it would be difficult to be well trained quickly in many situations. Therefore, using the LSTM model alone could not achieve an efficient end-to-end fault diagnosis. The performance curves of C/D-FUSA and C/D-FUS were excellent in both CRWU-512 and CRWU-6000. However, comparing the curves, it could be seen that C/D-FUSA was superior to C/D-FUS in terms of stability and training performance in the training process. The C/D-FUSA had extra attention modules compared with the C/D-FUS. This better performance indicates that both the context-dependent attention module proposed in this work and the SE attention
module used in the convolutional modules effectively improve the diagnostic performance of the algorithm.

The excellent performance of C/D-FUSA was also proven in comparison with other advanced algorithms. Although some excellent work fused multiple signal processing, deep learning, transfer learning, and other advanced technologies, our work still had higher fault diagnosis performance than theirs [31–34,45,47,51]. In addition, these methods may need to transform the vibration signals to other spaces and use statistical methods to extract various features to achieve accurate fault diagnosis. On the one hand, these data-processing methods may consume more time, making it difficult to achieve real-time fault diagnoses of rolling bearings. On the other hand, they could be overly dependent on some designed features, which would be adjusted or modified for different rolling bearings. These disadvantages would make it difficult for these methods to be rapidly applied to a new work environment. Therefore, in an actual situation, the cost of these models may be higher. However, our proposed algorithm does not use complex data processing or feature engineering methods, and the deep learning model guides the whole diagnostic process. This end-to-end diagnostic approach can significantly reduce the difficulty and cost of model deployment and scenario migration in practice.

In addition to the performance of diagnostic accuracy, the method’s inference speed is also crucial. If the inference speed were too slow, it would be challenging to keep up with the signal sampling, making it unable to realize real-time fault diagnosis and, thus, miss some critical monitoring periods. The slow inference speed could bring missing faults and huge losses. We calculated the inference speed of our model at different diagnostic frequencies (inference per 512/6000 signals). The results showed that the inference speed of the proposed model was much faster than the signal-sampling frequency and was not significantly different from that of such simple model structures like LSTM and C/D-FUS. These results demonstrated that our proposed C/D-FUSA model could achieve the real-time, accurate, easily deployable and portable end-to-end rolling bearing fault diagnosis.

5. Conclusions

This paper proposed an end-to-end continuous/discontinuous feature fusion method for rolling bearing fault diagnosis (C/D-FUSA). This method comprises subnets for continuous features, discontinuous features and classification. The subnet for continuous features was composed of long short-term memory (LSTM) models with a proposed context-dependent attention module. This attention module weighted the final outputs of the LSTM according to the contextual outputs. This subnet was used for extracting the continuous features with contextual information. The subnet for discontinuous features was composed of several convolutional modules: a convolutional layer, a batch normalization layer, a ReLU activation function, a squeeze-and-excitation attention module and a max-pooling layer. The subnet was designed to extract features with discrete relations. Finally, the continuous and discontinuous features were input into the subnet for classification and transformed into probabilities of different fault types. By comparing the diagnosis performance and inference time with the other, simpler deep learning methods and other state-of-the-art methods, the proposed method was proven to be accurate (99.85% of accuracy and 99.87% of F1 score) and in real time (0.034 s for one sample with 512 signals). In practical applications, our proposed model can be deployed in neural network chips or embedded systems, access multiple vibration sensor detection information, and directly output diagnostic information. The advantages of this method also make it easy to be deployed and migrated in practice.

Author Contributions: Algorithm research and thesis writing, J.Z.; thesis revision and correction, J.L.; algorithm research, Z.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 anonymous reviewers whose insightful comments have helped to improve the quality of this paper considerably.

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

References
1. Liu, R.; Yang, B.; Zio, E.; Chen, X. Artificial intelligence for fault diagnosis of rotating machinery: A review. Mech. Syst. Signal Process. 2018, 108, 33–47. [CrossRef]
2. Bediaga, I.; Mendizabal, X.; Arnaiz, A.; Munoo, J. Ball bearing damage detection using traditional signal processing algorithms. IEEE Instrum. Meas. Mag. 2013, 16, 20–25. [CrossRef]
3. Duan, Z.; Wu, T.; Guo, S.; Shao, T.; Malekian, R.; Li, Z. Development and trend of condition monitoring and fault diagnosis of multi-sensors information fusion for rolling bearings: A review. Int. J. Adv. Manuf. Technol. 2018, 96, 803–819. [CrossRef]
4. Tandon, N.; Choudhury, A. A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings. Tribol. Int. 1999, 32, 469–480. [CrossRef]
5. Frank, P.M.; Ding, S.X.; Marcu, T. Model-based fault diagnosis in technical processes. Trans. Inst. Meas. Control 2000, 22, 57–101. [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. Ding, Y.; Jia, M.; Xiao, D.; Ding, S.X. A novel time–frequency Transformer based on self-attention mechanism and its application in fault diagnosis of rolling bearings. Mech. Syst. Signal Process. 2021, 168, 108616. [CrossRef]
8. Dong, S.; Xu, X.; Chen, R. Application of fuzzy C-means method and classification model of optimized K-nearest neighbor for fault diagnosis of bearing. J. Braz. Soc. Mech. Sci. Eng. 2015, 38, 2255–2263. [CrossRef]
9. Yan, X.; Jia, M. A novel optimized SVM classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing. Neurocomputing 2018, 313, 47–64. [CrossRef]
10. Xiao, D.; Ding, J.; Li, X.; Huang, L. Gear Fault Diagnosis Based on Kurtosis Criterion VMD and SOM Neural Network. Appl. Sci. 2019, 9, 5424. [CrossRef]
11. Antoni, J. The spectral kurtosis: A useful tool for characterising non-stationary signals. Mech. Syst. Signal Process. 2006, 20, 282–307. [CrossRef]
12. Barszcz, T.; Randall, R. Application of spectral kurtosis for detection of a tooth crack in the planetary gear of a wind turbine. Mech. Syst. Signal Process. 2009, 23, 1352–1365. [CrossRef]
13. Zimroz, R.; Bartelmus, W. Application of Adaptive Filtering for Weak Impulsive Signal Recovery for Bearings Local Damage Detection in Complex Mining Mechanical Systems Working under Condition of Varying Load. Solid State Phenom. 2011, 180, 250–257. [CrossRef]
14. Rzeszucinski, P.; Orman, M.; Pinto, C.T.; Tkaczyk, A.; Sulowicz, M. Bearing Health Diagnosed with a Mobile Phone: Acoustic Signal Measurements Can be Used to Test for Structural Faults in Motors. IEEE Ind. Appl. Mag. 2018, 24, 17–23. [CrossRef]
15. Orman, M.; Rzeszucinski, P.; Tkaczyk, A.; Krishnamoorthi, K.; Pinto, C.T.; Sulowicz, M. Bearing fault detection with the use of acoustic signals recorded by a hand-held mobile phone. In Proceedings of the 2015 International Conference on Condition Assessment Techniques in Electrical Systems (CATCON), Bangalore, India, 10–12 December 2015; pp. 252–256.
16. Yu, Y.; Dejie, Y.; Junsheng, C. A roller bearing fault diagnosis method based on EMD energy entropy and ANN. J. Sound Vib. 2006, 294, 269–277. [CrossRef]
17. Seryasat, O.; Shoorehdeli, M.A.; Honavar, F.; Rahmani, A. Multi-fault diagnosis of ball bearing using FFT, wavelet energy entropy mean and root mean square (RMS). In Proceedings of the 2010 IEEE International Conference on Systems, Man and Cybernetics, Istanbul, Turkey, 10–13 October 2010; pp. 4295–4299. [CrossRef]
18. Song, L.; Wang, H.; Chen, P. Vibration-Based Intelligent Fault Diagnosis for Roller Bearings in Low-Speed Rotating Machinery. IEEE Trans. Instrum. Meas. 2018, 67, 1887–1899. [CrossRef]
19. Almounajied, A.; Sahoo, A.K.; Kumar, M.K.; Alsebai, M.D. Investigation Techniques for Rolling Bearing Fault Diagnosis Using Machine Learning Algorithms. In Proceedings of the 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 6–8 May 2021; pp. 1290–1294. [CrossRef]
20. Manikandan, S.; Duraivelu, K. Fault diagnosis of various rotating equipment using machine learning approaches—A review. Proc. Inst. Mech. Eng. Part E J. Process Mech. Eng. 2020, 235, 629–642. [CrossRef]
21. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. Nature 2015, 521, 436–444. [CrossRef]
22. Wang, Z.; Cao, L.; Zhang, Z.; Gong, X.; Sun, Y.; Wang, H. Short time Fourier transformation and deep neural networks for motor imagery brain computer interface recognition. Concurr. Comput. Pract. Exp. 2018, 30, e4413. [CrossRef]
23. Wang, Z.; Sun, Y.; Shen, Q.; Cao, L. Dilated 3D Convolutional Neural Networks for Brain MRI Data Classification. IEEE Access 2019, 7, 134388–134398. [CrossRef]
24. Wang, Z.; Zhu, Y.; Shi, H.; Zhang, Y.; Yan, C. A 3D multiscale view convolutional neural network with attention for mental disease diagnosis on MRI images. Math. Biosci. Eng. 2021, 18, 6978–6994. [CrossRef] [PubMed]
25. Wang, Z.; Zhang, Y.; Shi, H.; Cao, L.; Yan, C.; Xu, G. Recurrent spiking neural network with dynamic presynaptic currents based on backpropagation. Int. J. Intel. Syst. 2021, 37, 2242–2265. [CrossRef]
26. Voulodimos, A.; Doulamis, N.; Doulamis, A.; Protopapadakis, E. Deep Learning for Computer Vision: A Brief Review. *Comput. Intell. Neurosci.* 2018, 2018, 7068349. [CrossRef] [PubMed]

27. Otter, D.W.; Medina, J.R.; Kalita, J.K. A Survey of the Usages of Deep Learning for Natural Language Processing. *IEEE Trans. Neural Netw. Learn. Syst.* 2020, 32, 604–624. [CrossRef]

28. Nassif, A.B.; Shahin, I.; Attili, I.; Azzeh, M.; Shaalan, K. Speech Recognition Using Deep Neural Networks: A Systematic Review. *IEEE Access* 2019, 7, 19143–19165. [CrossRef]

29. Chen, X.; Gohl, J. Industrial Robot Control with Object Recognition based on Deep Learning. *Procedia CIRP* 2018, 76, 149–154. [CrossRef]

30. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* 1997, 9, 1735–1780. [CrossRef]

31. Shao, H.; Jiang, H.; Zhang, X.; Niu, M. Rolling bearing fault diagnosis using an optimization deep belief network. *Meas. Sci. Technol.* 2015, 26, 115002. [CrossRef]

32. Fuan, W.; Hongkai, J.; Haidong, S.; Wenjing, D.; Shuaipeng, W. An adaptive deep convolutional neural network for rolling bearing fault diagnosis. *Meas. Sci. Technol.* 2017, 28, 095005. [CrossRef]

33. Sohaib, M.; Kim, C.-H.; Kim, J.-M. A Hybrid Feature Model and Deep-Learning-Based Bearing Fault Diagnosis. *Sensors* 2017, 17, 2876. [CrossRef]

34. Li, X.; Zhang, W.; Ding, Q. Understanding and improving deep learning-based rolling bearing fault diagnosis with attention mechanism. *Signal Process.* 2019, 161, 136–154. [CrossRef]

35. Alexakos, C.; Karnavas, Y.; Drakaki, M.; Tziafettas, I. A Combined Short Time Fourier Transform and Image Classification Transformer Model for Rolling Element Bearings Fault Diagnosis in Electric Motors. *Mach. Learn. Knowl. Extr.* 2021, 3, 228–242. [CrossRef]

36. Chen, X.; Zhang, B.; Gao, D. Bearing fault diagnosis base on multi-scale CNN and LSTM model. *J. Intell. Manuf.* 2020, 32, 971–987. [CrossRef]

37. Daubechies, I.; Lu, J.; Wu, H.-T. Synchronized wavelet transforms: An empirical mode decomposition-like tool. *Appl. Comput. Harmon. Anal.* 2011, 30, 243–261. [CrossRef]

38. Zhang, S.; Zhang, S.; Wang, B.; Habetler, T.G. Deep learning algorithms for bearing fault diagnosis—a comprehensive review. *IEEE Access* 2020, 8, 29857–29881. [CrossRef]

39. Case Western Reserve University. B.D.C. Seeded Fault Test Data. 2017. Available online: https://engineering.case.edu/bearingdatacenter (accessed on 10 May 2022).

40. Yu, X.; Dong, F.; Ding, E.; Wu, S.; Fan, C. Rolling Bearing Fault Diagnosis Using Modified LFDA and EMD with Sensitive Feature Selection. *IEEE Access* 2017, 6, 3715–3730. [CrossRef]

41. Jiang, F.; Ding, K.; He, G.; Du, C. Sparse dictionary design based on edited cepstrum and its application in rolling bearing fault diagnosis. *J. Sound Vib.* 2020, 490, 115004. [CrossRef]

42. Yu, Y.; Si, X.; Hu, C.; Zhang, J. A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures. *Neural Comput.* 2019, 31, 1235–1270. [CrossRef]

43. Wang, Y.; Huang, M.; Zhu, X.; Zhao, L. Attention-based LSTM for aspect-level sentiment classification. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin, TX, USA, 1–5 November 2016; pp. 606–615. [CrossRef]

44. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141. [CrossRef]

45. Loparo, K.A. Bearings Vibration Data Set; The Case Western Reserve University Bearing Data Center. Available online: http://www.eecs.cwru.edu/laboratory/bearing/download.htm (accessed on 1 May 2022).

46. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* 2014, arXiv:1412.6980. Available online: https://arxiv.org/abs/1412.6980 (accessed on 10 May 2022).

47. Lei, Y.; Jia, F.; Lin, J.; Xing, S.; Ding, S.X. An Intelligent Fault Diagnosis Method Using Unsupervised Feature Learning towards Mechanical Big Data. *IEEE Trans. Ind. Electron.* 2016, 63, 3137–3147. [CrossRef]

48. Wang, J.; Zhan, C.; Yu, D.; Zhao, Q.; Xie, Z. Rolling bearing fault diagnosis method based on SSAE and softmax classifier with improved K-fold cross-validation. *Meas. Sci. Technol.* 2022, 33. [CrossRef]

49. Yan, J.; Kan, J.; Luo, H. Rolling Bearing Fault Diagnosis Based on Markov Transition Field and Residual Network. *Sensors* 2022, 22, 3936. [CrossRef] [PubMed]

50. Zhang, R.; Gu, Y. A Transfer Learning Framework with a One-Dimensional Deep Subdomain Adaptation Network for Bearing Fault Diagnosis under Different Working Conditions. *Sensors* 2022, 22, 1624. [CrossRef] [PubMed]

51. Zhao, K.; Jiang, H.; Wang, K.; Pei, Z. Joint distribution adaptation network with adversarial learning for rolling bearing fault diagnosis. *Knowl. Based Syst.* 2021, 222, 106974. [CrossRef]

52. Zou, F.; Zhang, H.; Sang, S.; Li, X.; He, W.; Liu, X. Bearing fault diagnosis based on combined multi-scale weighted entropy morphological filtering and bi-LSTM. *Appl. Intell.* 2021, 51, 6647–6664. [CrossRef]

53. Gu, K.; Zhang, Y.; Liu, X.; Li, H.; Ren, M. DWT-LSTM-Based Fault Diagnosis of Rolling Bearings with Multi-Sensors. *Electronics* 2021, 10, 2076. [CrossRef]