

Fault Diagnosis of Bearing Based on Convolutional Neural Network Using Multi-Domain Features

Xiaorui Shao¹, Lijiang Wang², Chang Soo Kim^{1*}, and Ilkyeun Ra³

¹Information system, Pukyong National University
Busan, 608737 South Korea

[e-mail: shaoxiaorui@pukyong.ac.kr, cskim@pknu.ac.kr]

²Global and area studies, Pukyong National University
Busan, 608737 South Korea

[e-mail: wanglijiang0901@gmail.com]

³Department of Computer Science and Engineering, University of Colorado Denver,
Denver, CO 80217, USA

[e-mail: ilkyeun.ra@ucdenver.edu]

*Corresponding author : Chang Soo Kim

*Received December 9, 2020; revised March 6, 2021; accepted April 12, 2021;
published May 31, 2021*

Abstract

Failures frequently occurred in manufacturing machines due to complex and changeable manufacturing environments, increasing the downtime and maintenance costs. This manuscript develops a novel deep learning-based method named Multi-Domain Convolutional Neural Network (MDCNN) to deal with this challenging task with vibration signals. The proposed MDCNN consists of time-domain, frequency-domain, and statistical-domain feature channels. The Time-domain channel is to model the hidden patterns of signals in the time domain. The frequency-domain channel uses Discrete Wavelet Transformation (DWT) to obtain the rich feature representations of signals in the frequency domain. The statistic-domain channel contains six statistical variables, which is to reflect the signals' macro statistical-domain features, respectively. Firstly, in the proposed MDCNN, time-domain and frequency-domain channels are processed by CNN individually with various filters. Secondly, the CNN extracted features from time, and frequency domains are merged as time-frequency features. Lastly, time-frequency domain features are fused with six statistical variables as the comprehensive features for identifying the fault. Thereby, the proposed method could make full use of those three domain-features for fault diagnosis while keeping high distinguishability due to CNN's utilization. The authors designed massive experiments with 10-folder cross-validation technology to validate the proposed method's effectiveness on the CWRU bearing data set. The experimental results are calculated by ten-time averaged accuracy. They have confirmed that the proposed MDCNN could intelligently, accurately, and timely detect the fault under the complex manufacturing environments, whose accuracy is nearly 100%.

Keywords: CNN, DWT, Fault Diagnosis, Multi-domain, Time series Classification

1. Introduction

With the rapid and continuous development of the manufacturing industry, the manufacturing environments are complex and changeable, causing a high incidence of failures. Those failures may cause disastrous accidents, including economic losses, environmental pollution, and even life casualties [1]. Finding one practical approach to detect potential process abnormalities and component faults as early as possible for enhancing the security and reliability of the whole control system has been extracted massive attention.

The current fault diagnosis method could be divided into three categories: model-based, signal-based, knowledge-based methods [2, 3]. The model-based methods require the operator to hand the principles and the constraints of the manufacturing industries before diagnosis. Then the observer gives the decision by comparing the consistency between the measured outputs and expected outputs [4]. Different from the model-based methods, signal-based methods detect the fault by measuring the consistency in inexplicit input-output. That is, they transform the raw signals into low dimensional time-domain or frequency-domain features for final judgement. e.g., the changes of root-mean-square current factor between healthy and faulty conditions are calculated for diagnosis power converters of switched reluctance motors [5]. Feng et al. employed the Fourier spectrum for the diagnosis of planetary gearboxes [6]. As described above, both model-based and signal-based methods need to judge the fault by hand and depend upon the expert's experience.

The knowledge-based methods, also known as the data-driven methods, including statistical analysis-, and learning-based methods, could learn some common patterns from historical data without expert's experience for fault diagnosis. Statistical analysis-based methods mainly contain principal component analysis (PCA) and independent component analysis (ICA). They convert the raw signals into low-dimensional representations to detect the fault. For instance, Kaistha et al. [7] applied PCA for fault detection and isolation of a pressurized water reactor; Harmouche et al. [8] proposed a fault diagnosis method based on Kullback–Leibler divergence using PCA for an eddy currents application; Yu et al. [9] developed a novel method based on ICA for fault diagnosis of rotating machinery. Unfortunately, transformed less-dimensional signals may lose some critical information, thereby decreasing the performance of diagnosis. Moreover, improper feature selection in PCA and ICA also influence the diagnosis's performance.

The learning-based methods could overcome the above issues by directly processing raw signals. That is, integrating feature selection and fault diagnosis in one single framework without any human interaction. Two learning-based methods, including shallow learning and deep learning methods, are used for fault diagnosis. For the application of the shallow learning-based method, Widodo et al. [10] employed a support vector machine (SVM) machine for condition monitoring and fault diagnosis. A tree-based shallow learning method, random forest (RF), was used for fault diagnosis in [11-13]. Furthermore, shallow artificial neural network (ANN) maps the raw signals into high-level feature maps in two or three hidden layers for fault diagnosis [14]. Even though shallow learning-based methods made a remarkable improvement for fault diagnosis, there still existed some limitations. Especially, SVM requires a large variety of memory to find the optimal plane, and it is easy to be overfitting. RF is sensitive to noisy data [15]. Shallow ANN is easy to fall into underfitting or overfitting as the extracted features are not “representative” enough.

Fortunately, recent bosting deep learning technology gives us a new view to couple with fault diagnosis problem. Chiefly, deep brief networks (DBN), recurrent neural networks (RNN), and CNN are utilized for fault diagnosis. e. g., Tao et al. [16] adopted DBN with

fourteen statistical variables for bearing fault diagnosis. Zhao et al. [17] applied one of the most potent RNNs-long short-term memory (LSTM) for fault diagnosis of Tennessee Eastman benchmark process. Lei et al. [18] proposed a novel multi-channel LSTM (MCLSTM) with multivariate time series for the wind turbine's fault diagnosis. Chen et al. [19] applied CNN for fault diagnosis of the gearbox and achieved averaged accuracy of 96.8% on twenty testing data set by inputting the fast Fourier transform (FFT) transformed spectrum. Zhang et al. [20] proposed a novel CNN-based method using wide first-layer kernels (WDCNN) for fault diagnosis of bearing, and it achieved 100% accuracy on 19800 training samples. Long et al. [21] applied a two-dimensional CNN (2D-CNN) based on the structure of LeNet-5 for fault diagnosis; the validation experiments proved its state-of-the-art performance. A novel multi-scale CNN (MSCNN) was developed to diagnose wind turbine gearbox [22]. They calculated the mean of window-signals as the different-scale inputs. The results indicate that four-scale CNN performs the best. Motivated by [22], Shao et al. [23] developed a multi-scale feature fusion CNN (MSFFCNN) for time series classification in a smart factory, which could be utilized for fault diagnosis by changing the input data. They directly use raw signals with different-scale convolutional operations to extract the multi-scale feature representations; extracted features are fused for classification. Moreover, a deep transfer learning approach based on Visual Geometry Group with 16-layer (VGG-16) and continuous wavelet transform (CWT) is proposed for bearing, induction motors, and gearbox fault diagnosis [24]. The authors give one table to see some essential deep learning-based methods for fault diagnosis in recent years, as shown in Table 1.

Table 1. Recent references for fault diagnosis using deep learning technology

Method	Year	Input data
CNN [19]	2015	Statistical variables
DBN [16]	2016	FFT spectrum
WDCNN [20]	2017	Raw signals
LSTM [17]	2018	Raw signals
MCLSTM [18]	2018	Multivariate sensor signals
LeNet-5 [21]	2018	2-D transformed signals
MSCNN [22]	2019	Raw signals
Deep transfer VGG-16 [24]	2019	CWT transformed images
MSFFCNN [23]	2020	Raw signals

As shown in Table 1, most of the current deep learning-based methods only applied single domain (time or frequency domain) features for fault diagnosis. For instance, WDCNN [20], LSTM [17], MCLSTM [18], LeNet-5 [21], MSCNN [22], and MSFFCNN [23] utilized raw signals (time-domain) and DBN [16] adopted FFT transformed frequency-domain features for fault diagnosis. They may ignore some essential feature representations in both time and frequency domains so that the fault diagnosis is still not satisfactory and can be improved. Even if deep transfer VGG-16 [24] employed CWT transformed time-frequency domain features for fault diagnosis, it is hard to deploy on light-weight terminals for real-time diagnosis due to its huge parameters (more than 138 million). Moreover, most of those works did not test in a complex working environment. Motivated by those, this manuscript presents a novel multi-domain CNN (MDCNN) to extract rich feature representations from time, frequency, and statistical domains for intelligent, accurate, and real-time fault diagnosis. In the proposed MDCNN, time-domain features are extracted from raw signals, and frequency-domain features are extracted from DWT transformed coefficients using CNN. Statistical

features are calculated by six statistical variables. Those three-domain features are fused as comprehensive features for fault diagnosis.

The main contributions of this manuscript, as summarized follows:

- We present a novel end-to-end fault diagnosis framework based on CNN using multiple domain features without any inner feature selection engineering. Comparative experiments have confirmed its effectiveness and priority.
- The proposed MDCNN has excellent anti-noise and transfer learning capacities. Therefore, it works well in a complicated, noisy manufacturing environment.
- The impact of each component in the MDCNN has been analyzed through an ablation study. Moreover, the inner feature of MDCNN was explored by using t-distributed stochastic neighbor embedding (t-SNE) technology.

The rest of this manuscript is arranged as follows. Section 2 gives some pre-knowledges of the proposed MDCNN, including CNN, DWT, and six statistical variables. Section 3 presents the proposed MDCNN for fault diagnosis. In Section 4, sufficient comparative experiments are carried out to verify MDCNN's effectiveness. Section 5 discussed the proposed MDCNN in-depth. The conclusions and future studies are presented in Section 6.

2. Methodology

2.1 CNN

CNN could extract the hidden patterns of the input data through various filters in a virtual structure, which has been widely used for 2-D image classification [25], 1-D energy consumption forecasting [15, 26], and object detection [27]. Chiefly, CNN consists of three critical components: convolution, pooling, and activation functions. Convolution and pooling operations occurred alternatively in the structure of CNN. The activation function is to activate some parts of features to enhance the feature's representability. The convolution operation is implemented using one convolutional layer with different filters, as described in (1).

$$X_j^l = f(\sum_{i \in M_j} X_i^{l-1} * F_{ij}^l + \varepsilon_j^l) \quad (1)$$

Where l^{th} layer's feature is calculated through convolution operation $*$ between $(l - 1)^{th}$ layer's input and filters F_{ij}^l with a bias vector ε_j^l . Moreover, the final feature map is activated by one activation function $f()$. One of the most famous activation functions is the Rectified Linear Unit (ReLU) [28]. In which the values less than zero will be deleted, as defined in (2). Obtained feature maps are processed by one pooling operation to reduce the complexity and speed up the network. There are three pooling operations, including maximum, minimum, and average pooling. After processing of CNN, the obtained features have the priority of translation, scaling, and rotation invariance, which is very suitable for fault diagnosis.

$$a_j^l = \begin{cases} 0, & \text{if } X_i^{l-1} < 0 \\ X_i^{l-1}, & \text{Otherwise} \end{cases} \quad (2)$$

The structure and workflow of 2-D CNN with two 2-D convolutional and pooling layers is described in Fig. 1. The term "Conv", "Pooling", and "ReLU" are corresponding to convolution, pooling, and activation functions. Moreover, obtained features need to be flattened with the "Flatten" operation. Those features are utilized to predict the output in a linear mode.

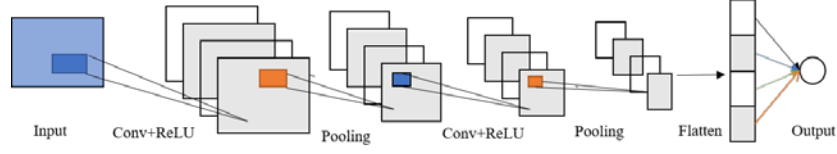


Fig. 1. The workflow of 2-D CNN with two convolutional and pooling layers

2.2 DWT

DWT could decompose the raw time-series $signal(t)$ into frequency-domain features with time localization using a set of discrete wavelets $DW_{a,b}(t)$ with scaling factor a , and time localization b , as given in (3). Where $DW_{a,b}(t)$ is discretized with individual mother wavelet such as Daubechies, Harr, Symlets, Coiflets, as defined in (4). Involving a , and b , we could convert DWT into one scaling function S (5) and one discrete wavelet function D (6). Furthermore, the coefficients of DWT, including two parts: “approximation” part $a_{\frac{j}{2}}(k)$ works as one low-frequency pass filter, as given in (7); “details” part $d_{\frac{j}{2}}(k)$ works as one high-frequency pass filter, as shown in (8). Where $D(t)_{j,k}^*$ is the complex conjugate of $D(t)_{j,k}$. Multi-resolution frequency-domain features could be obtained by multi-level DWT decomposition. Enormously, one-level DWT (1-DWT) decomposes the signal into (a_1, b_1) corresponding to the low-frequency part and high-frequency part. Two-level DWT (2-DWT) decomposes the 1-DWT generated low-frequency part a_1 into (a_2, b_2) . Repeating the above operations n times will get n – resolution frequency-domain features $(a_n, b_n, b_{n-1}, \dots, b_2, b_1)$.

$$DWT(signal(t))_{a,b} = \int_{-\infty}^{+\infty} signal(t) DW_{a,b}(t) dt \quad (3)$$

$$DW_{a,b}(t) = a^{-1/2} \left(\frac{t-b}{a} \right), \text{ where } \begin{cases} a = 2^j, j \in Z \\ b = k2^j, j, k \in Z \end{cases} \quad (4)$$

$$S(t)_{j,k} = \frac{1}{\sqrt{2^j}} S(2^{-j}t - k) \quad (5)$$

$$D(t)_{j,k} = \frac{1}{\sqrt{2^j}} D(2^{-j}t - k) \quad (6)$$

$$a_{\frac{j}{2}}(k) = \int signal(t) S(t)_{j,k} dt \quad (7)$$

$$d_{\frac{j}{2}}(k) = \int signal(t) D(t)_{j,k}^* dt \quad (8)$$

2.3 Statistical Variables

Statistical variable reflects the macro changing trend, which is very useful for distinguishing each fault type due to the signals having low cohesion and high coupling. The mean, maximum (max), and minimum (min) values could be used to detect the outliers existing in the signal. The standard deviation (std) reflects the stability of the signal. Kurtosis (Kurt) and

skewness (skew) reflect the distribution of signals. More detailed, kurtosis describes the steepness and slowness of a signal, while skewness indicates the distribution's symmetry. The calculations of each statistical variable, as given in (9)-(14). Where N is the length of the signal. The reference [26] also adopted those six statistic variables for power consumption forecasting.

$$mean = \frac{1}{N} \sum_{t=1}^N signal(t) \quad (9)$$

$$max = \max (signal(t)) \quad (10)$$

$$min = \min (signal(t)) \quad (11)$$

$$std = \sqrt{\frac{1}{N} \sum_{t=1}^N (signal(t) - mean)^2} \quad (12)$$

$$skew = E \left[\left(\frac{signal(t) - mean}{sd} \right)^3 \right] \quad (13)$$

$$kurt = E \left[\left(\frac{signal(t) - mean}{sd} \right)^4 \right] \quad (14)$$

3. The proposed MDCNN for Fault Diagnosis

This manuscript presents a multi-domain CNN for fault diagnosis, as shown in Fig. 2. It contains five steps: Input construction (step 1) is to obtain the time-domain features “raw”, frequency-domain and stactical-domain inputs. Step 2 is to extract the high-level features using CNN from time and frequency-domain inputs. Step 3 merged extracted time and frequency-domain features. Step 4 makes a fusion between the time-frequency and statistical-domain features. The fusion features will be used for fault diagnosis. Step 5 gives fault diagnosis result and update the whole networks. The detailed description is given in the following subsections.

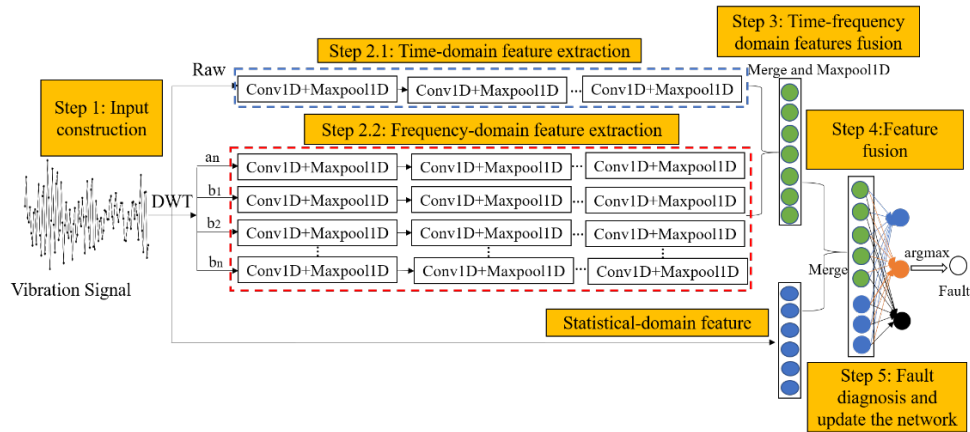


Fig. 2. The framework of the proposed MDCNN for fault diagnosis. Five steps are contained in the proposed framework. They are input construction, CNN feature extraction, time-frequency domain features fusion, feature fusion, fault diagnosis and updates the neural network. Notably, CNN feature extraction includes time-domain feature extraction, marked with a blue dotted frame; and frequency-domain feature extraction, marked with a red dotted frame.

3.1 Input Construction

Assuming that we collected N signals denoted as $Sig = \{sig_1, sig_2, sig_3, \dots, sig_N\}$, each signal sig contains T values, denoted as $sig = \{x_1, x_2, x_3, \dots, x_t, \dots, x_T\}$, where t is time step. Moreover, we know the fault type of each signal given by experts, which is formalized as $L = \{l_1, l_2, l_3, \dots, l_N\}$. The historical samples $\{Sig, L\}$ will be utilized to train the MDCNN for fault diagnosis. This manuscript adopted three-domain inputs to model the hidden patterns of each signal fully. The raw signal is the time-domain input. The raw signal is decomposed by DWT using (7) and (8) to obtain multi-resolution frequency-domain expression, which could be written as $D_{frequency} = DWT(sig) = \{a_n, d_1, d_2, \dots, d_n\}$. Here, n is the DWT decomposition level. Six statistical variables defined in (9)-(14) are utilized to distinguish signals in the statistical domain. Consequently, the input of the proposed MDCNN could be written as (15), where $statistic(sig)$ is to calculate six statistical variables.

$$Input = \{sig, DWT(sig), statistic(sig)\} \quad (15)$$

3.2 Domain Features Extraction

Time and frequency-domain inputs are processed by CNN individually to extract high-level hidden patterns. This manuscript primarily applies 1-D CNN for feature extraction, in which 1-D convolution operation (Conv1D) and maximum pooling operation (Maxpool1D) alternatively occurred. To easily understand the structure of the proposed MDCNN, the authors named the combination of Conv1D and Maxpool1D as "Block." The time-domain features extracted by CNN are denoted as (16), and the frequency-domain features are written as (17). The $Block_m$ represents the depth of the "Block", m depends on the length of signal T . Larger m could extract more abstract feature representations while it will increase the model's parameters. To make a trade-off, the authors set m as three. Moreover, the hidden nodes in the Conv1D layer increase from 16 to 64, and the pooling size is set as two. A more detailed setting of hyperparameters is discussed in the experiment part.

$$Features_{time} = Block_m(sig) \quad (16)$$

$$Features_{frequency} = [Block_m(a_n), Block_m(d_1), Block_m(d_2), \dots, Block_m(d_n)] \quad (17)$$

3.3 Feature Fusion

To obtain the fusion features, the proposed MDCNN employed merge operation to integrate three domain features. First, MDCNN merges the time-domain and frequency-domain features, then fused time-frequency domain features are processed by one max pooling layer to reduce its dimensions and speed up the network again. This process is defined as (18). The comprehensive features integrate time-frequency domain features and statistic-domain features, as shown in (19).

$$Features_{time-frequency} = Maxpool1D(concate(Features_{time}, Features_{frequency})) \quad (18)$$

$$Features = concate(Features_{time-frequency}, statistic(sig)) \quad (19)$$

3.4 Output and Update the Network

The comprehensive features are utilized to predict the probability of each fault in a linear mode. The maximum probability corresponding to the label is selected as the signal's fault type using the "argmax" operation. The process of fault diagnosis using the proposed MDCNN could be formalized as:

$$Fault = MDCNN(Input) = MDCNN(\{sig, DWT(sig), statistic(sig)\}) \quad (20)$$

Where $MDCNN()$ is the trained model. As can be seen from (20), the proposed MDCNN could directly detect the fault using three-domain features without any feature selection operations. Moreover, MDCNN updates the whole network by calculating the loss of categorical cross-entropy between predicting values and ground truth with "Adam" [29] optimizer. All activation functions are "ReLU," except for the last is "Linear."

4. Experimental Verification

The authors implement the proposed MDCNN based on the operating system of ubuntu 16.04, 64 bits with 23.4 gigabytes, Intel (R) i7-700 CPU. The deep learning framework is TensorFlow backend Keras, and the programming language is python.

4.1 Data

This manuscript adopts a bearing data set collected from the CWRU bearing data center to validate the effectiveness of the proposed MDCNN. The authors adopt 12k Drive End Bearing fault data and Normal Baseline data for the experiment. Each data sample is collected at 12,000 samples/second under four different loads, including 0 power horse (PH), 1 PH, 2 PH, and 3 PH corresponding subset A, B, C, and D. Each subset (loading) consists of four different faulty diameters: 7, 14, 21, and 28 mils. One diameter fault is caused by different components: inner race (IR), ball, outer race (OR) of the bearing except for 28 mils without OR. Consequently, eleven faulty data under four loadings are collected. Besides, a standard baseline data set is used for identifying faulty bearing to normal bearing under each loading. That is, twelve subclassifications existed in the data set, including IR7, IR14, IR21, IR28, Ball 7, Ball 14, Ball 28, OR7, OR14, OR21, and Normal. Moreover, the original faulty data is a long time series; the authors transformed each faulty type into 685 samples with the sample rate of 2048 using algorithm 1. The reason for that is CNN requires equal-length time series as the input and considering the influence of data imbalance. Zhang et al. also utilized 2048 as a sample rate to process the data [20]. Therefore, each subset contains $8200=685 \times 12$ samples. Furthermore, subset E integrates A, B, C, and D, including $32880=8200 \times 4$ samples, which is generated to verify the proposed MDCNN reasonably. The description of the data is described in Table 2.

Algorithm 1: Overlap algorithm to generate the data samples

Input: A long time series T, the number of samples N (defined is 685), and the length of sample K (defined is 2048).

Output: The transformed samples.

1. Compute the stride $= \frac{\text{len}(T)-K}{N-1}$.
 2. Initial each sample's starting position start = 0, and ending position end = K.
 3. Define a variable Data to save transformed time series T.
-

-
4. for i in $0:N\{$
 5. $Data[i] = T[start:end]$
 6. $start = start + stride$
 7. $end = start + K \}$
 9. return Data
-

Table 2. Each subset description.

Data (loads)	Faulty types (faulty components + faulty diameters)												Samples	
	IR7	IR14	IR21	IR28	Ball7	Ball14	Ball21	Ball28	OR7	OR14	OR21	Normal		
A (0PH)	√	√	√	√	√	√	√	√	√	√	√	√	√	8200
B (1 PH)	√	√	√	√	√	√	√	√	√	√	√	√	√	8200
C (2PH)	√	√	√	√	√	√	√	√	√	√	√	√	√	8200
D (3PH)	√	√	√	√	√	√	√	√	√	√	√	√	√	8200
E (0, 1, 2, 3PH)	√	√	√	√	√	√	√	√	√	√	√	√	√	32880

4.2 Modeling

Before modeling, the authors did an exploratory analysis to illustrate the difficulties for fault diagnosis under complicated environments, as shown in Fig. 3. All types of signals fluctuate frequently and randomly. It is tough to identify Ball 7, Ball 21, IR 7, IR 28, OR 14, and normal even though they are in the same loading; IR 14 is very similar to OR 21. Also, the fault types of Ball 7 in 0 PH, Ball 21 in 2 PH, IR 28 in 3 PH, and OR 14 in 1 PH are very similar. Those strong similarities and randomness increase the difficulties for fault diagnosis of bearing. Therefore, this manuscript applies CNN to extract high-level hidden patterns of each signal from three-domain inputs to increase distinguishability. To efficiently modeling, the authors convert the 12 fault types into digits [0, 11].

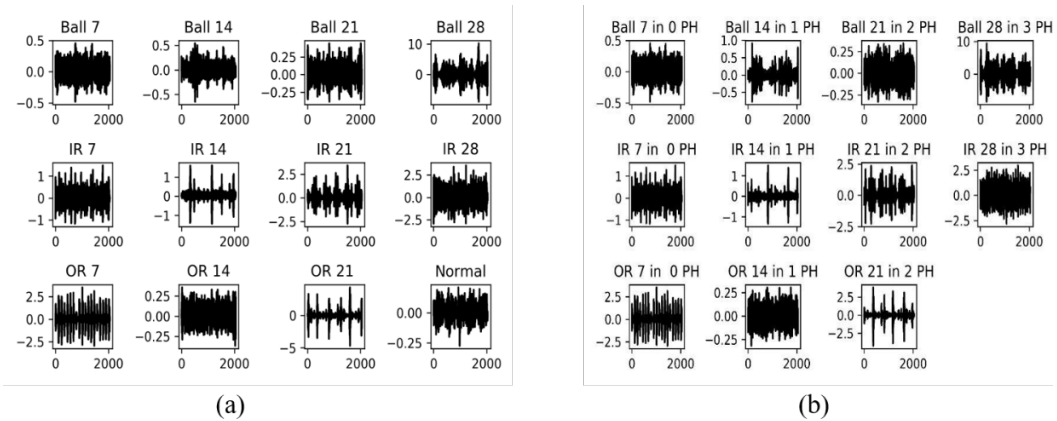


Fig. 3. The visualization of each fault. (a) is for different faults in the same loading 0 PH; (b) is for different faults in different loadings.

Two issues need to consider carefully: which wavelet is better for fault diagnosis? How many levels of DWT should be selected? The authors calculated the mean square error (MSE) between wavelets reconstructed signals and the ground truth using three kinds of DWT: 1-DWT, 2-DWT, and three-level DWT (3-DWT). Five wavelets including “db2”, “bior 1.1”, “coif 2”, “harr” and “sym2” are utilized for validation, the results as shown in Table 3. The results indicate that “db2” is the most suitable for reconstructing the bearing signals, which outperforms others in all kinds of DWT. Moreover, 1-DWT outperforms other levels. Thereby

1-DWT of “db2” is selected for building the model and consequential analysis. Moreover, the authors give one example of 3-DWT using “db2” to explain its working principle, as shown in Fig. 4. In the 1-DWT process, the original signal is decomposed into approximation part a_1 and details part d_1 . Then the approximation part of 1-DWT a_1 is decomposed into approximation part a_2 and details part d_2 . After three-time decomposition, we could obtain four subseries: a_3, b_3, b_2, b_1 to model the signal’s changing pattern in the frequency domain. Decomposed components are more identifiable than the original signal due to it reduces some random vibration.

Table 3. The MSE between DWT reconstructed signals and ground truth in different levels

Wavelet	MSE (1-DWT)	MSE (2-DWT)	MSE (3-DWT)
Db2	9.30×10^{-33}	1.89×10^{-32}	1.99×10^{-32}
Bior 1.1	1.48×10^{-32}	3.17×10^{-32}	3.48×10^{-32}
Coif 2	1.66×10^{-32}	3.05×10^{-32}	3.20×10^{-32}
Harr	1.48×10^{-32}	3.17×10^{-32}	3.49×10^{-32}
Sym2	6.09×10^{-25}	7.08×10^{-25}	7.62×10^{-25}

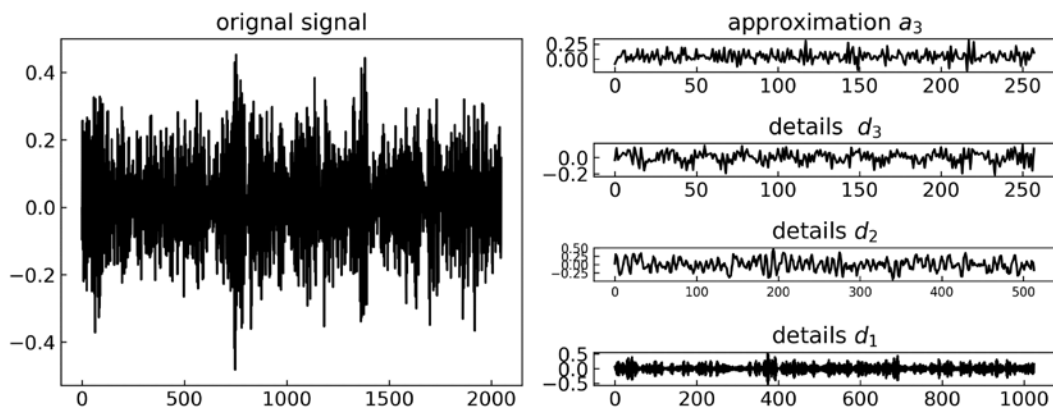


Fig. 4. One example of 3-DWT. The original signal is decomposed into four subseries: approximation part a_3 , details part d_3, d_2, d_1 .

4.3 Comparative Analysis

To validate the effectiveness of the proposed MDCNN for fault diagnosis, we compared it with some leading deep learning-based methods, including WDCNN [20], MSCNN [22], and MSFFCNN [23] on subset E. The detailed configurations of those comparative methods, as shown in Table 4. Each deep model runs ten times using a 10-folder cross-validation approach to evaluate fairly, and the early stop strategy is applied to find the best model as soon as possible. Specially, we split the data into three parts: train, validation, and test with the ratio of 7:1:2. The training part is to train the model. The validation part is to find the best model in the given training steps within the early stop strategy. If the validation accuracy did not increase within ten steps, the training process would be ended. Furthermore, the model with the highest accuracy on validation data will be saved as the best model.

Table 4. The configuration information for comparative methods

Method	Input	Description
WDCNN [20]	Raw signals	Structure: Input-Conv1D(16,64,24)-Maxpool1D(2)-Conv1D(32,3,1)-Maxpool1D(2)-Conv1D(64,3,1)-Maxpool1D(2)-Conv1D(64,3,1)-Maxpool1D(2)-Dropout(0.5)-Dense(100)-Dense(classes) Activation functions are “ReLU”, optimizer is “Adam”, loss function is “cross-entropy”
MSCNN [22]	Input1: Raw signals Input2: mean(2) Input3: mean(3)	Structure: Input-Conv1D(16)-Max1D(2)-Conv1D(32)-Max1D(2)-Conv1D(64)-Max1D(2) Three sub-outputs: Sub1, Sub2, Sub3 Output: Concatenate(Sub1,Sub2,Sub3)-Dense(100)-Dense(classes) Activation functions are “ReLU”, optimizer is “Adam”, loss function is “cross-entropy”
MSFFCNN [23]	Raw signals	The author’s code

The comparison results in terms of accuracy, as shown in **Table 5**. The findings show that the proposed method wins five times with the highest accuracy of 100%. Moreover, the results show that the proposed method outperforms others in terms of average accuracy, which is up to 99.97%. Furthermore, the proposed method has good robustness, which can be conducted from the standard error. Those three methods could be ranked as: The proposed>MSFFCNN>MSCNN>WDCNN. To quantify the difference between those methods, the authors did a *t*-test, as shown in **Table 6**. The results indicate that the significant difference existed between the proposed method and WDCNN, MSFFCNN due to the *p*-value of the *t*-test is less than 0.05; By the contrary, there is no significant difference between the proposed MDCNN and MSCNN. The reason for that is MSCNN reached the highest accuracy of 100% four times while the other are two or less. However, MSCNN is still influenced by the random data splitting, which increases the risk for the fault diagnosis of bearing. In summary, the proposed method could accurately detect the fault of bearing using vibration signals with good robustness.

Table 5. The results of different deep models for bearing fault diagnosis on subset E (%)

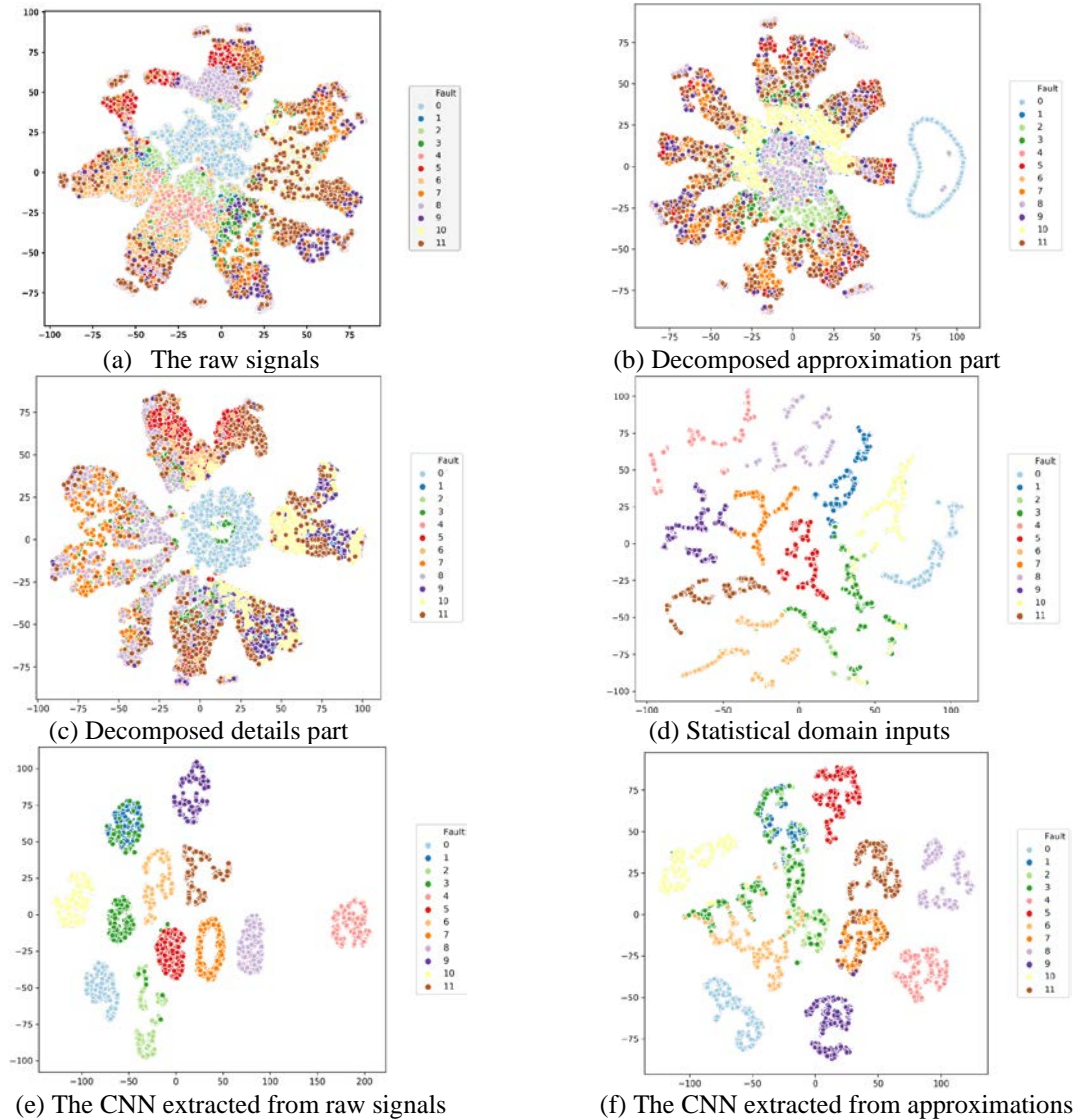
Method	Times										Average
	1	2	3	4	5	6	7	8	9	10	
WDCNN [20]	99.01	99.97	100	99.97	99.97	91.64	100	91.67	91.67	91.67	96.56 ±4.01
MSCNN [22]	91.67	99.97	99.97	99.91	100	100	91.67	100	100	99.91	98.31 ±3.32
MSFFCNN [23]	99.85	99.97	98.80	98.75	99.79	98.69	99.97	99.97	99.82	99.91	99.55 ±0.53
The proposed	100	99.94	99.94	100	100	99.94	100	99.88	99.97	100	99.97 ±0.04

Table 6. The *p*-value of *t*-test among four deep models

Method	WDCNN [20]	MSCNN [22]	MSFFCNN [23]	The proposed
WDCNN [20]	1	0.33	0.04	0.02
MSCNN [22]	0.33	1	0.28	0.15
MSFFCNN [23]	0.04	0.28	1	0.03
The proposed	0.02	0.15	0.03	1

4.4 Feature Extraction Capacity Analysis

To explore and understand the feature extraction capacity of the proposed MDCNN, the authors visualized some layers' output using t-SNE technology on subset A, as shown in Fig. 5. The raw signals (a) are mixed, inseparable, nonlinear and difficult to identify; 1-DWT decomposed "approximation" (b) and "details" parts (c) could identify fault 0 from others; Statistical components (d) enables those signals more separable. After the process of 1-D CNN, the features extracted from the time-domain becomes separable and identifiable, which can be conducted from (e). However, fault 2 and fault 3 are still mixed. Luckily, the CNN extracted features from DWT transformed the "approximation" part (f) makes fault 2 and fault 3 more discrete. Extracted features from DWT transformed "details" part (g) enables fault 2 far from fault 3. By combining those three domain features, CNN learned final features (h) are discriminable and independent, which satisfies the principle of classification: the distance between classes should be the largest while the distance within classes should be the smallest.



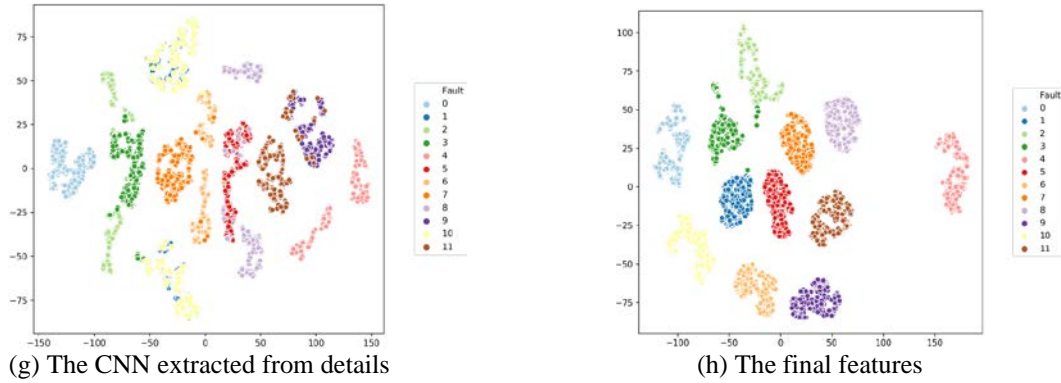


Fig. 5. The feature visualization results using t-SNE technology on subset A. (a) is the raw signals; (b) and (c) are 1-DWT decomposed “approximation” and “details” parts; (d) is six statistical variables; (e), (f), and (g) are CNN extracted features from raw signals, 1-DWT decomposed components, respectively. Different colours represent different fault types, as described in the legend.

4.5 Anti-noise Test

In the real production environment, collected data may contain noises, which increase the difficulty and uncertainty for fault diagnosis. It requires the model has adequate anti-noise capacity. This manuscript validates the proposed method’s anti-noise capacity on subset E. Especially, different-intensity white noises are added into raw signals for validation. The white noise’s intensity is measured with the signal-noise ratio (SNR), as defined (21). Where P_{signal} and P_{noise} are the power of signals and noise, whose unit is the decibel (dB). The authors added the white noise is from -4dB to 10dB. Three comparative methods defined in **Table 5** are used to illustrate the priority of the proposed method. The averaged accuracy, as shown in **Fig. 6**. The error bar is calculated with standard error, which is used to test each method’s robustness. The findings derive that the proposed MDCNN outperforms others; all cases are higher than 99%. Furthermore, it wins seven times expect for the MSFFCNN has the highest accuracy for -4dB. Moreover, the proposed method achieves the highest averaged accuracy of 99.62% simultaneously gets the lowest standard error of 0.26. Even though comparing to the none-noise case (see **Table 5**), the proposed method only reduced by 0.35% on average. Those pieces of evidence prove that the proposed method has an adequate anti-noise capacity for fault diagnosis.

$$SNR = 10 \log_{10} \frac{P_{signal}}{P_{noise}} \quad (21)$$

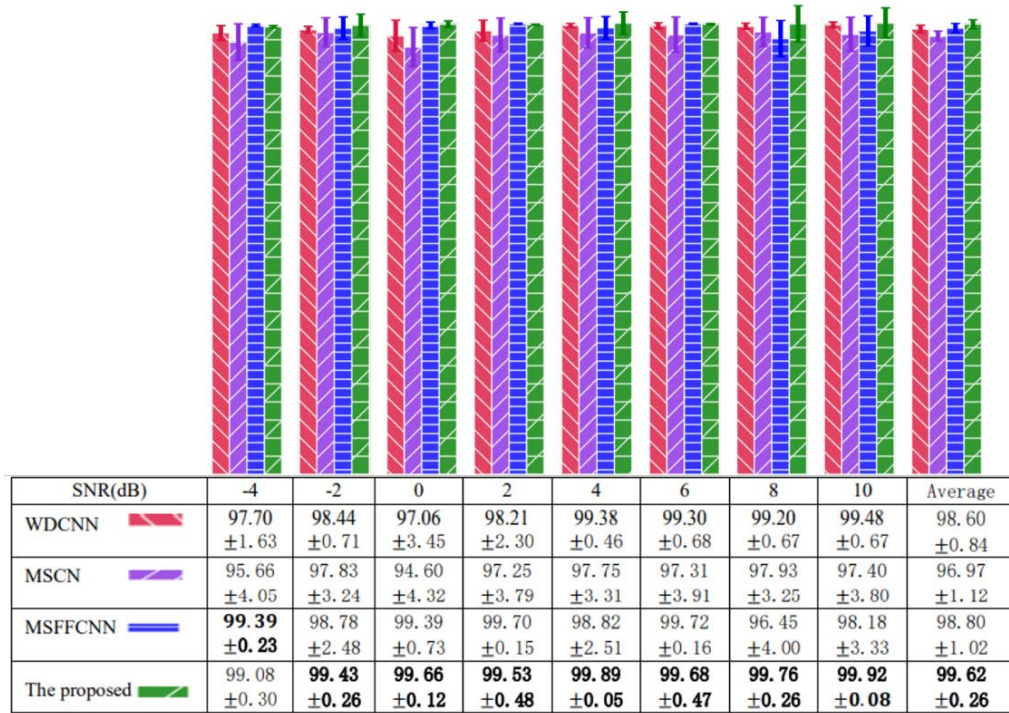


Fig. 6. The anti-noise test on subset E using different deep learning-based methods.

4.6 Ablation Study

To explore each component's effect in the proposed MDCNN for fault diagnosis, the authors did the ablation study with five experiments. We designed time-frequency domain CNN (TFCNN) to validate the effectiveness of the statistical inputs; Designed time-domain CNN (TCNN) and time-statistical CNN (TSCNN) to explore the impact of frequency-domain inputs; Designed frequency-domain CNN (FCNN) and frequency-statical CNN (FSCNN) to validate the effectiveness of time-domain inputs, respectively. All configurations of those experiments are the same as MDCNN. The ten-time results' distribution and averaged accuracy for each model, as shown in Fig. 7. Statistical-domain inputs have improved the fault diagnosis performance from 98.28% to 99.97% by comparing MDCNN and TFCNN; By comparing MDCNN with TSCNN, the frequency-domain has improved performance of 1.37%; The time-domain features contribute more than 0.08%, which can be conducted from MDCNN and FSCNN. Moreover, only using one or two domain inputs for fault diagnosis results in unstable distribution, which could be conducted from TFCNN, TCNN, TSCNN, and FCNN. Among them, FSCNN performs better compared to other combinational and signal models. Designed and combined each domain feature with CNN properly could effectively improve fault diagnosis's performance, including the accuracy and robustness, which could be found from MDCNN. In summary, statistical domain inputs provide robust feature expression of raw signals; DWT transformed frequency-domain features enable the proposed more accurate and robust; Time-domain features reflect the changing trend of raw signals in the time domain.

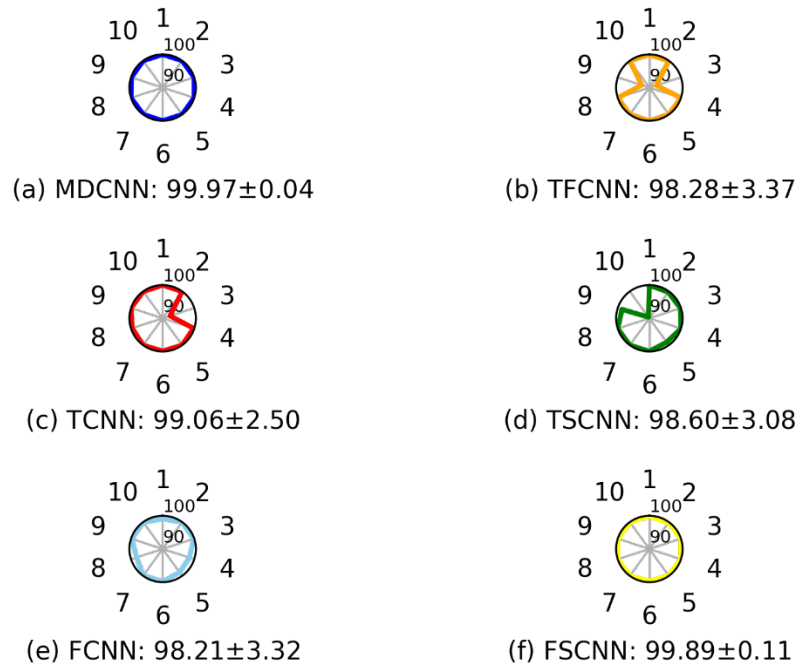


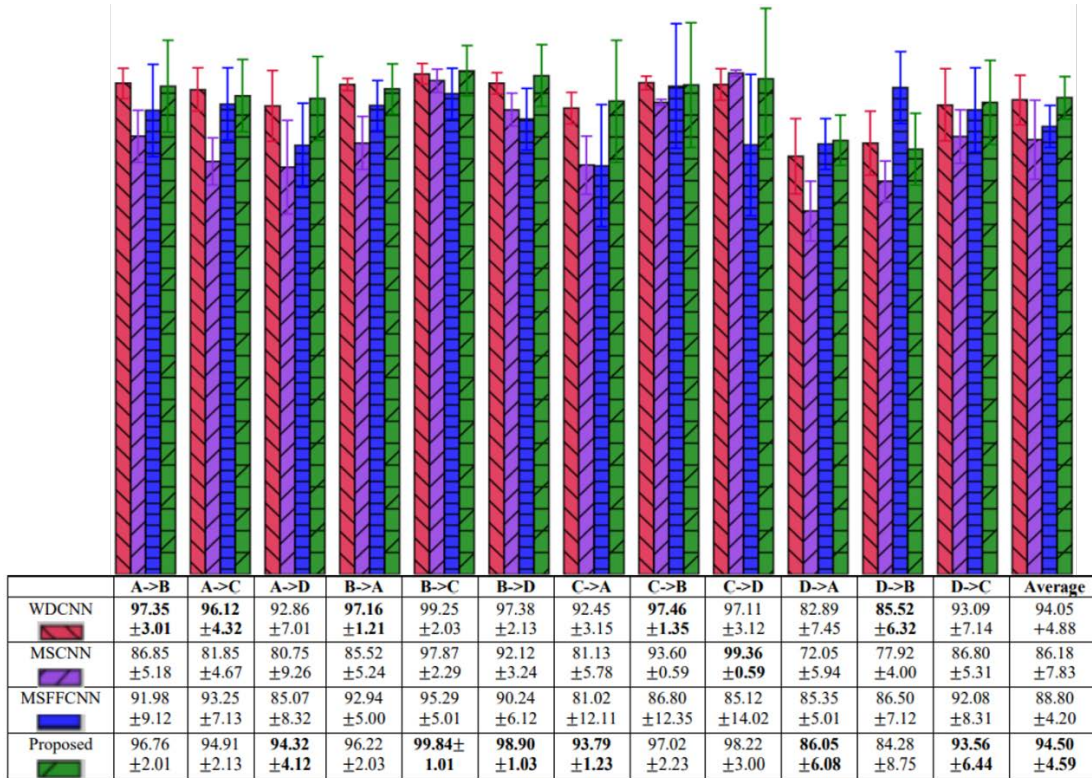
Fig. 7. The ten-time results' distribution and averaged accuracy using radar plotting for ablation study. The X-axis is training times, and the y-axis is the corresponding accuracy.

4.7 Transfer Learning Capacity

In a real production environment, too many factors may influence the fault diagnosis results, such as different currents, working loadings, and varying operator level. It is unrealistic to collect all these types of data to train each case. Therefore, the model has good transfer learning capacity is important. The authors designed four experiments to validate the proposed method's transfer learning capacity, as described in [Table 7](#). In each case, one subset is used to train the model while the other three are for testing. For instance, subset A is used as training data to train the proposed model; subsets B, C, and D are for testing in case 1, written as A->B, A->C, and A->D. We compared the proposed method with the other three leading methods, including WDCNN, MSCNN, and MSFFCNN, to validate its effectiveness. The comparative results using ten-time averaged accuracy and standard error, as shown in [Fig. 8](#). The findings show that the proposed method has won six times in making of accuracy for the following cases: A->D, B->C, B->D, C->A, D->A, and D->C. WDCNN won five times for A->B, A->C, B->A, C->B, D->B, and MSCNN won one time for C->D. Moreover, the proposed method outperforms others in terms of average accuracy, which is up to 94.50. Those four methods could be classed into two levels. The proposed method and WDCNN could be divided into the first level due to their averaged accuracy are higher than 90%, while the other two are less than it. Besides, the proposed method is more stable than WDCNN, which could be conducted from averaged error bar: the proposed (4.59) < WDCNN (4.88). The comparative analysis has proven that the proposed MDCNN has a good transfer learning capacity for fault diagnosis.

Table 7. Four experimental rigs used for the transfer learning capacity test

Experiments	Training data	Testing data		
Case 1	A	B	C	D
Case 2	B	A	C	D
Case 3	C	A	B	D
Case 4	D	A	B	C

**Fig. 8.** The comparative results for the transfer learning capacity test

5. Discussion

We have proposed a novel deep model named MDCNN for fault diagnosis of bearing, as shown in Fig. 2. The difficulties for fault diagnosis of the bearing are strong similarities existed in different kinds of signals, which could be seen from Fig. 3. Therefore, this manuscript used three-domain features to model the hidden patterns of signals fully. Moreover, CNN is employed to automatically extract in-depth, rich feature representations from three-domain inputs for intelligent, end-to-end fault diagnosis of bearing.

The authors have given one clear pipeline for building the model. Primarily, it utilized DWT to obtain frequency-domain inputs and calculated six variables as statistical domain features. Moreover, the authors calculated MSE to choose the best wavelet from five wavelets. The result indicated that "db2" performs the best within one-level DWT, as shown in Table 3. One 3-DWT example in Fig. 4 indicates that DWT decomposed components are more identifiable than the original signal.

The comparative experiments have confirmed the effectiveness and priority of the proposed method for fault diagnosis of bearing, which could be conducted from Table 5. Each model is trained ten times to evaluate fairly. The proposed method's accuracy is near to 100%, while

the standard is lower than 0.05.

Fig. 5 has confirmed that the proposed method has excellent feature extraction capacity. The ablation study has analyzed each component's impact in the MDCNN, as shown in **Fig. 7**. Statistical domain inputs provide robust feature expression of raw signals; DWT transformed frequency-domain features enable the proposed more accurate and robust; Time-domain features reflect the changing trend of raw signals in the time domain.

We added white noise with different intensities (from -6 dB to 10dB) into the raw signals to validate the anti-noise capacity of the proposed MDCNN. The results indicated that the proposed method has an adequate anti-noise capacity, and all cases' accuracies are higher than 99%, which could be found in **Fig. 6**. The comparative analysis in **Fig. 6** confirmed its priority again.

We designed four experiments to verify the proposed method's transfer learning capacity, as described in **Table 7**. The comparative analysis has illustrated that the proposed MDCNN has an excellent transfer learning capacity, is suitable for fault diagnosis of bearing in complex environments.

Moreover, the proposed method only takes 128 microseconds (μs) to process each sample, which can be used for real-time fault diagnosis. For the parameters of the proposed MDCNN, it has 116, 616 parameters and takes up 1.2 megabytes (MB), which could be easily deployed in light-weight mobile terminals for fault diagnosis.

6. Conclusion

In conclusion, this paper proposed a novel, accurate, and intelligent end-to-end framework named MDCNN for fault diagnosis of bearing. The proposed MDCNN has three-channel inputs, including raw time-domain signals, DWT transformed frequency-domain representations and statistical domain features. Time and frequency domains channels are fed into CNN individually to learn hidden patterns existed in the time-frequency domain. Extracted time-frequency hidden features are merged with six statistical variables as comprehensive features for fault diagnosis. Sufficient and detailed comparative analysis has confirmed the proposed method's effectiveness for fault diagnosis of bearing, whose accuracy is nearly 100%. It could detect the fault accurately in various complex environments such as noisy environments, different loadings due to its excellent feature extraction and transfer learning capacities. Moreover, the proposed method could be easily deployed in light-weight mobile terminals for real-time fault diagnosis as its parameters only take 12 MB.

In the future, we will validate the proposed MDCNN's generality on other kinds of signals. Besides, as described in **Fig. 8**, the averaged accuracy for different loads is 94.50%, which is not very high and can be improved by using domain adaption technology. Thereby, fault diagnosis using deep learning and domain adaption technologies is another interest.

Acknowledgement

We would like to thank CWRU bearing data center provides the data set.

References

- [1] S. Shao, W. Sun, R. Yan, "A Deep Learning Approach for Fault Diagnosis of Induction Motors in Manufacturing," *Chinese J. Mech. Eng.*, vol. 30, no. 6, pp. 1347–1356, 2017. [Article \(CrossRef Link\)](#)
- [2] Z. Gao, C. Cecati, and S. X. Ding, "A survey of fault diagnosis and fault-tolerant techniques-part I: Fault diagnosis with model-based and signal-based approaches," *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3757–3767, 2015. [Article \(CrossRef Link\)](#)
- [3] Z. Gao, C. Cecati, and S. X. Ding, "A Survey of Fault Diagnosis and Fault-Tolerant Techniques—Part I: Fault Diagnosis With Knowledge-Based and Hybrid/Active Approaches," *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3757–3767, 2015. [Article \(CrossRef Link\)](#)
- [4] L. Xu and H. E. Tseng, "Robust model-based fault detection for a roll stability control system," *IEEE Trans. Control Syst. Technol.*, vol. 15, no. 3, pp. 519–528, 2007. [Article \(CrossRef Link\)](#)
- [5] H. Chen and S. Lu, "Fault diagnosis digital method for power transistors in power converters of switched reluctance motors," *IEEE Trans. Ind. Electron.*, vol. 60, no. 2, pp. 749–763, 2013. [Article \(CrossRef Link\)](#)
- [6] Z. Feng and M. J. Zuo, "Fault diagnosis of planetary gearboxes via torsional vibration signal analysis," *Mech. Syst. Signal Process.*, vol. 36, no. 2, pp. 401–421, 2013. [Article \(CrossRef Link\)](#)
- [7] N. Kaistha and B. R. Upadhyaya, "Incipient fault detection and isolation in a PWR plant using principal component analysis," in *Proc. of Am. Control Conf.*, vol. 3, pp. 2119–2120, 2001. [Article \(CrossRef Link\)](#)
- [8] J. Harmouche, C. Delpha, and D. Diallo, "Incipient fault detection and diagnosis based on Kullback-Leibler divergence using principal component analysis: Part II," *Signal Processing*, vol. 109, pp. 334–344, 2015. [Article \(CrossRef Link\)](#)
- [9] G. Yu, X. Liang, and J. Wang, "Fault feature separation for fault diagnosis of rotating machinery using ICA with reference," in *Proc. of ICRMS'2011 - Saf. First, Reliab. Prim. Proc. 2011 9th Int. Conf. Reliab. Maintainab. Saf.*, no. 1, pp. 1010–1014, 2011. [Article \(CrossRef Link\)](#)
- [10] A. Widodo and B. S. Yang, "Support vector machine in machine condition monitoring and fault diagnosis," *Mech. Syst. Signal Process.*, vol. 21, no. 6, pp. 2560–2574, 2007. [Article \(CrossRef Link\)](#)
- [11] B. S. Yang, X. Di, and T. Han, "Random forests classifier for machine fault diagnosis," *J. Mech. Sci. Technol.*, vol. 22, no. 9, pp. 1716–1725, 2008. [Article \(CrossRef Link\)](#)
- [12] C. Aldrich and L. Auret, "Fault detection and diagnosis with random forest feature extraction and variable importance methods," *IFAC Proceeding Volumes*, vol. 43, no. 9, pp.79-86, 2010. [Article \(CrossRef Link\)](#)
- [13] Z. Chen et al., "Random forest based intelligent fault diagnosis for PV arrays using array voltage and string currents," *Energy Convers. Manag.*, vol. 178, no. October, pp. 250–264, 2018. [Article \(CrossRef Link\)](#)
- [14] X. Shao, C. Pu, Y. Zhang, and C. S. Kim, "Domain Fusion CNN-LSTM for Short-Term Power Consumption Forecasting," *IEEE Access*, vol. 8, pp. 188352–188362, 2020. [Article \(CrossRef Link\)](#)
- [15] X. Shao and C. S. Kim, "Multi-step Short-term Power Consumption Forecasting Using Multi-Channel LSTM With Time Location Considering Customer Behavior," *IEEE Access*, vol. 8, pp. 125263-125273, 2020. [Article \(CrossRef Link\)](#)
- [16] J. Tao, Y. Liu, and D. Yang, "Bearing Fault Diagnosis Based on Deep Belief Network and Multisensor Information Fusion," *Shock Vib.*, vol. 2016, 2016, Art. ID 9306205. [Article \(CrossRef Link\)](#)
- [17] H. Zhao, S. Sun, and B. Jin, "Sequential Fault Diagnosis Based on LSTM Neural Network," *IEEE Access*, vol. 6, pp. 12929–12939, 2018. [Article \(CrossRef Link\)](#)
- [18] J. Lei, C. Liu, and D. Jiang, "Fault diagnosis of wind turbine based on Long Short-term memory networks," *Renew. Energy*, vol. 133, pp. 422–432, 2019. [Article \(CrossRef Link\)](#)

- [19] Z. Chen, C. Li, and R. Sanchez, "Gearbox Fault Identification and Classification with Convolutional Neural Networks," *Shock Vib.*, vol. 2015, 2015, Art. ID 390134. [Article \(CrossRef Link\)](#)
- [20] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors (Switzerland)*, vol. 17, no. 2, 2017. [Article \(CrossRef Link\)](#)
- [21] L. Wen, X. Li, L. Gao, and Y. Zhang, "A New Convolutional Neural Network-Based Data-Driven Fault Diagnosis Method," *IEEE Trans. Ind. Electron.*, vol. 65, no. 7, pp. 5990–5998, 2018. [Article \(CrossRef Link\)](#)
- [22] G. Jiang, H. He, J. Yan, and P. Xie, "Multiscale Convolutional Neural Networks for Fault Diagnosis of Wind Turbine Gearbox," *IEEE Trans. Ind. Electron.*, vol. 66, no. 4, pp. 3196–3207, 2019. [Article \(CrossRef Link\)](#)
- [23] X. Shao, C. Soo Kim, and D. Geun Kim, "Accurate Multi-Scale Feature Fusion CNN for Time Series Classification in Smart Factory," *Comput. Mater. Contin.*, vol. 65, no. 1, pp. 543–561, 2020. [Article \(CrossRef Link\)](#)
- [24] S. Shao, S. McAleer, R. Yan, and P. Baldi, "Highly Accurate Machine Fault Diagnosis Using Deep Transfer Learning," *IEEE Trans. Ind. Info.*, vol. 15, no. 4, pp. 2446–2455, 2019. [Article \(CrossRef Link\)](#)
- [25] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Neural Information Processing Systems 2012*, pp. 1–9, 2012.
- [26] X. Shao, C.-S. Kim, and P. Sontakke, "Accurate Deep Model for Electricity Consumption Forecasting Using Multi-channel and Multi-Scale Feature Fusion CNN-LSTM," *Energies*, vol. 13, no. 8, p. 1881, Apr. 2020. [Article \(CrossRef Link\)](#)
- [27] S. Akcay, M. E. Kundegorski, C. G. Willcocks, and T. P. Breckon, "Using deep convolutional neural network architectures for object classification and detection within x-ray baggage security imagery," *IEEE Trans. Inf. Forensics Secur.*, vol. 13, no. 9, pp. 2203–2215, 2018. [Article \(CrossRef Link\)](#)
- [28] V. Nair and G. E. Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines Vinod," in *Proc. of the 27th International Conference on International Conference on Machine Learning*, pp. 807–814, 2010.
- [29] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *Proc. of 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–15, 2015. [Online]. Available: <https://arxiv.org/abs/1412.6980>



Xiaorui Shao received a B. S. of engineering degree in computer science and technology and a B. S. of economics degree in the economics department from the Yanbian University, Yanji, Jilin, China 2017. He received an M.S. degree in the Interdisciplinary Program of Information Systems from Pukyong National University (PKNU), Busan, Korea, in 2019. He is pursuing the Ph. D degree in the department of information system from PKNU. He has won the outstanding paper reward of ICONI 2019, 2020 and the PKNU early career researcher award in 2020. His research interest includes the smart grid, time series forecasting, fault diagnosis, and deep learning.



Lijiang Wang received the B. S. of management degree in marketing from the Yanbian University, Yanji, Jilin, China, in 2017. She is now pursuing an M.S. degree in global and area studies from Pukyong National University, Busan, Korea. Her research interest, including time-series processing, economic forecast.



Chang-Soo Kim received Ph. D. degree in the Department of Computer Engineering from Chung Ang University in 1991, Seoul, Korea. He became a professor in the Department of IT Convergence and Application Engineering, Pukyong National University, Pusan, South Korea, in 1992 and has continued until the present. He has been the Vice President of Korea Multimedia Society since 2011. His current research interests include scheduling of smart factories, big data simulation.



Ilkyeun Ra received a Ph.D. degree in Computer and Information Science from Syracuse University, USA, MS degree in Computer Science from University of Colorado Boulder, Colorado, USA, and BS degree and MS degree in Computer Science from Sogang University, Seoul, Korea. Currently, he is an associate professor in the Department of Computer Science and Engineering at the University of Colorado Denver. His main research interests include computer networks, IoT, cloud computing, large scale of distributed computing systems, and high-performance computing.