

Article

Sensor Data-Driven Bearing Fault Diagnosis Based on Deep Convolutional Neural Networks and S-Transform

Guoqiang Li ¹ , Chao Deng ¹, Jun Wu ^{2,*} , Xuebing Xu ², Xinyu Shao ¹ and Yuanhang Wang ³

¹ School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China; lgq1211@tom.com (G.L.); dengchao@hust.edu.cn (C.D.); shaoxy@hust.edu.cn (X.S.)

² School of Naval Architecture and Ocean Engineering, Huazhong University of Science and Technology, Wuhan 430074, China; m13627290949@163.com

³ China Electronic Product Reliability and Environmental Testing Research Institute, Guangzhou 510610, China; wangyuanhang@ceprei.com

* Correspondence: wuj@hust.edu.cn

Received: 18 May 2019; Accepted: 17 June 2019; Published: 19 June 2019



Abstract: Accurate and timely bearing fault diagnosis is crucial to decrease the probability of unexpected failures of rotating machinery and improve the efficiency of its scheduled maintenance. Since convolutional neural networks (CNN) have poor feature extraction capability for sensor data with 1D format, CNN combined with signal processing algorithm is often adopted for fault diagnosis. This increases manual conversion work and expertise dependence while reducing the feasibility and robustness of the corresponding fault diagnosis method. In this paper, a novel sensor data-driven fault diagnosis method is proposed by fusing S-transform (ST) algorithm and CNN, namely ST-CNN. First of all, a ST layer is designed based on S-transform algorithm. In the ST layer, sensor data is automatically converted into 2D time-frequency matrix without manual conversion work. Then, a new ST-CNN model is constructed, and the time-frequency coefficient matrixes are inputted into the constructed ST-CNN model. After the training process of the ST-CNN model is completed, the classification layer such as softmax performs the fault diagnosis. Finally, the diagnosis performance of the proposed method is evaluated by using two public available datasets of bearings. The experimental results show that the proposed method performs the higher and more robust diagnosis performance than other existing methods.

Keywords: fault diagnosis; convolution neural networks; S-transform; sensor data; bearing

1. Introduction

Bearings are an extremely critical component in rotating machinery to reduce the friction between the moving parts and provide continuous effective support for the rotary axis. According to the investigation of rotating machinery failures, more than 45% of the machinery breakdown is caused by the bearing fault [1]. Thus, the bearing fault diagnosis has an enormous impact on maximizing the production efficiency of machinery, minimizing machinery downtime and the maintenance cost [2]. As a result, the bearing fault diagnosis has attracted considerable attentions.

In general, a typical bearing fault diagnosis includes three main steps: Monitoring signal acquisition, feature extraction, and fault diagnosis. In the monitoring signal acquisition step, sensor data such as vibration signals, acoustic emission signals, motor current signals, and temperature signals have been widely used in the field of fault diagnosis [3–5]. In addition, these signals are often collected by sensors mounted on the machinery. In the feature extraction step, a common practice is to extract time domain features such as the root mean square, skewness, kurtosis coefficient and

gap factor from collected signals using statistical methods [6–11]. To a certain extent, time domain features can effectively expose the fundamental differences under different conditions. In addition, frequency domain analysis methods [12,13] based on Fourier transform [14,15] are used to extract discernible frequency features from the collected signals. However, the monitoring signals of bearings obtained from practical industrial applications have non-linearity and non-stationary characteristics. The use of the statistical methods and frequency domain analysis methods to deal with the signals have their inherent limitations. To address this problem, time-frequency domain analysis methods such as wavelet transform (WT), empirical mode decomposition (EMD) and Hilbert–Huang transform (HHT) are introduced [16–19]. The time-frequency analysis methods decompose the collected signals into a set of time-frequency components, and these time-frequency components contain some fault features, which are useful for fault diagnosis [20–22]. In the fault diagnosis step, machine learning methods are the basis for effective fault diagnosis. For example, Yan et al. [23] proposed a novel fault diagnosis algorithm based on optimized support vector machines with multi-domain feature to achieve fault diagnosis of bearings. Zhao et al. [24] used the improved Euclidean weighted K-nearest neighbor (EW-KNN) classifier to monitor various health conditions of rolling bearings. Zhang et al. [25] introduced wavelet packet decomposition to improve EMD for time-frequency feature extraction. Singh et al. [26] presented a bearing fault diagnosis method based on the principles of EMD, envelope analysis and pseudo-fault signal. In this regard, the machine learning-based fault diagnosis method inevitably needs to rely on designed representative features and complex model tuning. However, the designed representative features such as wavelet coefficient features and intrinsic mode functions (IMFs) have their inherent limitations in extracting high-frequency components which have the obvious fault characteristics. In addition, strong background noise and other interference components of signals inevitably exist in collected monitoring signals under practical industrial environment. Moreover, traditional fault diagnosis methods based on machine learning are not always effective to eliminate the effect of strong background noise and other interference components of signals.

Deep learning (DL), as a new field of machine learning, provides an efficient way to automatically learn representative features from collected signals. Several common DL methods have been used in the field of fault diagnosis. For instance, Jia et al. [27] proposed a fault diagnosis model for rotating machinery based on deep neural networks (DNN). Sun et al. [28] proposed a fault classification model for induction motor fault classification based on sparse auto-encoder and DNN. However, fully-connected DNNs have limitations in solving more complex problems. The parameters of DNN are exponentially increasing when more layers are needed for data fitting. It leads to high computational complexity and possible overfitting problems.

Compared with the DNN-based method, a convolutional neural networks-based (CNN) [29,30] method is easier to train under the same available training set and computational resource. With the widespread application of CNN in fault diagnosis, CNN has shown the capability in extracting useful and robust features from monitoring signals. Ince et al. [31] proposed a 1D CNN-based approach that is directly applicable to the raw signal and achieved a more efficient fault detection system for real-time motor. However, the raw signals are interspersed with noise interference components, which increase the requirement of feature extraction capability of 1D CNN and increases the training cost of 1D CNN. Han et al. [32] developed the CNN-based model for gearbox fault diagnosis with constructed multi-level wavelet coefficients matrixes for reducing the interference of noise in vibration signal. However, common time-frequency analysis methods such as WT, short time Fourier transform (STFT), and HHT have the limitations when they are used to convert the signals into time-frequency coefficient matrixes for obtaining the sensitive fault components. One is that the strong correlation of the fault features extraction method-based on CNN with the quality of time-frequency information imposes big challenges in the noise interference background applications. The other is that the arduously obtained time-frequency information by manual work increases the complexity of the fault diagnosis method based on CNN.

In this paper, we propose a new sensor data-driven fault diagnosis method based on S-transform (ST) and CNN, namely ST-CNN. In order to obtain the appropriate inputs of CNN and enhance the quality of inputs, the ST is used to obtain the time-frequency matrix from sensor data. Several researchers have presented their investigations using ST and generalized ST to obtain time-frequency representative of signals for bearing fault detection [33,34]. Admittedly, the generalized ST can make up the poor energy concentration of ST in high frequency domain. However, in the proposed method, the ST is used to deal with sensor data for obtaining the useful time-frequency matrix due to it having poor energy concentration at the high frequency. In addition, its frequency-dependent window function produces higher frequency resolution at lower frequencies; hence the low frequency fault components could be enhanced [35]. Based on the above, the ST is introduced into CNN as the ST layer. Through the ST layer, the time-frequency complex matrix with 2D format is automatically extracted from sensor data without manual conversion, and this matrix with 2D format is a suitable input for the CNN. By integrating CNN with the ST layer, the ST-CNN could directly use original sensor data to realize the bearing fault diagnosis.

The remainder of this paper is organized as follows. Section 2 presents the proposed ST-CNN architecture. In Section 3, the procedure of the proposed fault diagnosis method is explained in detail. In Section 4, two real experiments are conducted to evaluate the effectiveness of the proposed method, and the results and discussions are presented. The conclusions are given in Section 5.

2. Proposed ST-CNN Architecture

Admittedly, neurons generally have the function of information extraction and discrimination in the human brain. Considering this, an architecture based on artificial neurons is designed. The sensor data is inputted into the designed architecture. Through the calculation and transfer of artificial neurons, the representative features of sensor data could be obtained, and then the condition of sensor data also could be achieved. As shown in Figure 1, the architecture of the proposed ST-CNN consists of five most important components, that is, the ST layer, convolutional layer, pooling layer, fully-connected layer and classification layer. In this architecture, the ST algorithm is integrated into the ST-CNN as the ST layer, which feed the appropriate inputs to the first convolutional layer. Then, the convolutional layers (i.e., Conv 1, Conv 2, Conv 3, Conv 4) and the pooling layers (i.e., MP 1, MP 2, MP 3), are used to extract the representative features. The batch normalization (i.e., BN) is used between MP 1 with Conv 2 and MP 2 with Conv 3, respectively. The fully-connected layer (i.e., FC 1) is adopted to non-linearly fit the representative features extracted from MP 3. The classification layer (i.e., SFM) is used to output the probability that the testing signal sample belong to different fault types. The details of the operation of each layer is as follows:

2.1. ST Layer

Recently, the merits of CNNs have been represented in the field of extracting representative feature of images. However, CNNs are not compatible with 1D time-series sensor data. In contrast, CNNs are well suited to obtain the representative features from input with the 2D format. Due to the differences of sensor data between different operating conditions could be reflected in time-frequency data, researchers have investigated the fault diagnosis based on the time-frequency analysis method (such as WT and STFT) and CNN. As the extension of the ideas of WT, ST is based on a moving and scalable localizing Gaussian window to obtain the desirable time-frequency feature, which is absent in WT [36]. In this regard, ST is used to deal with the sensor data to obtain the time-frequency matrix [37].

In general, time-frequency analysis of sensor data is completed under MATLAB, but which could not directly be inputted into the CNN. Based on this, the ST is designed and introduced into the ST-CNN as the ST layer. In the ST layer, 1D time-series sensor data samples are directly converted into

time-frequency matrixes with 2D format. No manual works are needed by this way. The operation of converting the sensor data $x(t)$ into a time-frequency matrix by the ST layer is described as:

$$s(\tau, f) = \int_{-\infty}^{+\infty} x(t) g(\tau - t, f) e^{-i2\pi f t} dt \quad (1)$$

where $x(t)$ is the sensor data. f is the sensor sampling frequency. $g(\tau - t, f)$ denotes a particular normalized Gaussian window function, and its formula can be expressed as:

$$g(t, f) = \frac{|f|}{\sqrt{2\pi}} e^{-\frac{t^2 f^2}{2}} \quad (2)$$

After Equation (1), the 1D time-series sensor data such as vibration signal is converted into a time-frequency complex matrix with 2D format (i.e., $s(\tau, f)$) and as the output of the ST layer, where the rows denote the frequencies and the columns indicate the time value. After the ST layer, the output of the ST layer is directly inputted into the CNN.

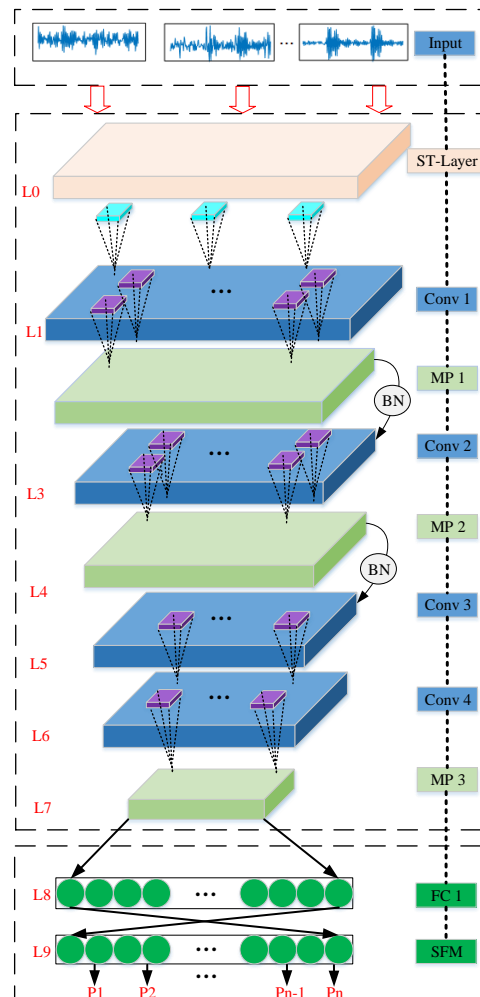


Figure 1. The architecture of the S-transform-convolutional neural networks (ST-CNN).

2.2. Convolutional Layer

For each convolutional layer, the number of filter kernel could be defined according to the need. In each convolutional layer, its kernel parameters are convolved with the data points of input. In this

paper, the input of the convolutional layer is $s(\tau, f) \in R^{A \times B}$, where A and B represents the length and width of $s(\tau, f)$ obtained from the ST layer, respectively. The output C_{cn} of the convolutional layer is formulated as:

$$C_{cn} = f(s(\tau, f) * w_{cn} + b_{cn}) \quad (3)$$

where $*$ is the convolutional operation, C_{cn} denotes cn -th feature map, cn represents the number of filter kernels, w_{cn} is the weight matrix of cn -th filter kernel of the current convolutional layer and cn -th filter kernel bias is b_{cn} . Typically, rectified linear units (ReLU) is selected to execute $f(\cdot)$ in Equation (4).

2.3. Pooling Layer

Admittedly, the dimensionality of output feature maps will increase after the convolutional layer, and the curse of dimensionality is easily caused with the increment of the convolutional layer. Based on this, the pooling layer is used to reduce the dimensionality of output feature maps. In the pooling layer, the dimensionality of the feature maps obtained from convolutional layer is eliminated by statistical methods such as max-pooling or average-pooling. This process is expressed as:

$$P_{cn} = f(\beta \text{down}(C_{cn}) + b) \quad (4)$$

where β is the multiplicative bias term, C_{cn} is the inputs, $\text{down}(C_{cn})$ denotes the pooling operation, b is the additive bias vector, and $f(\cdot)$ is the activation operation.

2.4. Fully-Connected Layer

Like traditional neural network, multiple neurons of fully-connected layer are used to non-linearly fit its input. All neurons are connected to all data point of feature maps from the last pooling layer such as MP 3. Its operation process is described as:

$$F(P_L) = f(wP + b) \quad (5)$$

where P is the outputs of the last pooling layer, $F(P_L)$ represents the outputs of current fully-connected layer, w and b denotes the weight and additive bias term, respectively. $f(\cdot)$ is the activation operation.

2.5. Classification Layer

Softmax is commonly used in the classification layer, which is the generalization of the logistic classifier for solving the multi-classification problem [38]. Sensor data sample x output f through above layers, and its predicted category is determined by $p(y = j|f)$. For the classification layer, its output is a vector with k -dimension, and the sum of the values of each element in this vector is 1. Its mathematical formula is given as:

$$h_{\gamma^T}(f^{(i)}) = \begin{bmatrix} p(y^{(i)} = 1|f^{(i)}; \gamma_1^T) \\ p(y^{(i)} = 2|f^{(i)}; \gamma_2^T) \\ \vdots \\ p(y^{(i)} = k|f^{(i)}; \gamma_k^T) \end{bmatrix} = \frac{1}{\sum_{j=1}^k e^{\gamma_j^T f^{(i)}}} \begin{bmatrix} e^{\gamma_1^T f^{(i)}} \\ e^{\gamma_2^T f^{(i)}} \\ \vdots \\ e^{\gamma_k^T f^{(i)}} \end{bmatrix} \quad (6)$$

where $\gamma_1^T, \gamma_2^T, \dots, \gamma_k^T$ are the parameters of regression model and $\frac{1}{\sum_{j=1}^k e^{\gamma_j^T f^{(i)}}}$ is to normalize the outputs.

Then, the cost function $M(\gamma^T)$ is defined as:

$$M(\gamma^T) = -\frac{1}{m} \left[\sum_{i=1}^m \sum_{j=1}^k 1\{y^i = j\} \log \frac{e^{\gamma_j^T f^{(i)}}}{\sum_{l=1}^k e^{\gamma_l^T f^{(i)}}} \right] \quad (7)$$

where $1\{\cdot\}$ is an indicative operation, which means that when the value of brace is true, its result is 1. Otherwise, its result is 0. The cost function is minimized by stochastic gradient descent algorithm.

3. Proposed Fault Diagnosis Method Based on ST-CNN

In this paper, a novel fault diagnosis method is proposed based on ST-CNN. With the capability of directly extracting features from original sensor data, no manual data conversion work is needed for bearing fault diagnosis. Figure 2 shows the flowchart of the proposed bearing fault diagnosis method. In the first step, sensor data of bearings in different conditions are collected. Data samples are obtained by a sampling window from sensor data. All the obtained data samples are then randomly divided into training, validation, and testing dataset. In the second step, the training dataset is used to train the proposed ST-CNN by reducing the training error. The validation dataset is used to verify the diagnostic performance of the trained ST-CNN and prevent possible overfitting and select the trained model. The testing dataset is adopted to evaluate the generalization capability of the proposed method.

In the training process of the ST-CNN, the training samples with 1D format are directly inputted into the proposed ST-CNN, and automatically converted into the time-frequency coefficient matrix by the ST operation. The output $L0$ of the ST operation is used as the input $s(\tau, f) \in R^{A \times B}$ for the first convolutional operation. After the input $s(\tau, f)$ is convolved with the filter kernel of convolutional operation, cn feature maps are obtained, which is formed as $L1$. The back-to-back pooling operation is used to reduce the dimensionality of $L1$. After the convolutional and pooling operation, the representative features are obtained from the training sample. In addition, the predicted result of the representative features could be obtained through classification stage, which is composed of one fully-connected layer (i.e., FC1) and the softmax layer (i.e., SFM). In SFM, the output neurons are transformed to the logits by Equation (6) to cater the form of probability distribution for the number of diagnosis types. Then, the training error of the ST-CNN will gradually minimize using Equation (7). After training, the ST-CNN directly extracts the representative fault features from sensor data. The fault diagnosis can be performed on new monitoring sensor data by the trained ST-CNN.

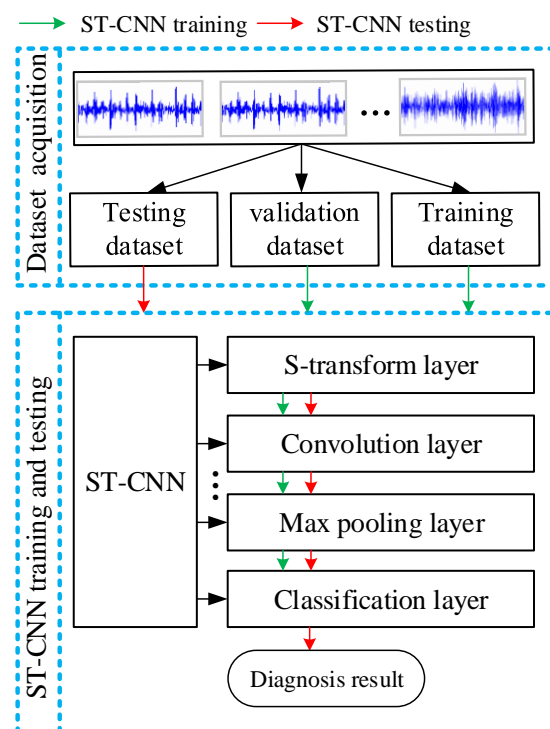


Figure 2. The flowchart of proposed fault diagnosis method.

4. Experiment Studies

This section uses two public available bearing datasets to evaluate the effectiveness of the proposed method.

4.1. Case One: Bearing Fault Diagnosis with Different Defect Severity

4.1.1. Experimental Setup and Data Description

In this case, a public available roller bearing dataset coming from the Case Western Reserve University (CWRU) motor drive system is analyzed [39]. As shown in Figure 3, the experimental setup main includes two hp motors (left), a torque transducer/encoder (center), a dynamometer (right), and control electronics (not shown). Two accelerometers are placed at the 12 o'clock position of the drive and fan end of the motor housing. Vibration signals of bearings are collected by the two accelerometers. Different degrees of single point fault diameters are introduced at outer race, inner race and ball of bearings by using electro-discharge device.

In this experiment, there are three types of bearing faults, which are inner ring fault (IRF), outer race fault (ORF) and ball fault (BF), as well as a normal bearing condition. Each bearing fault type contains three fault diameters: 0.007 inch, 0.014 inch, and 0.021 inch. There are three load conditions (1, 2 and 3 hp). Each load condition contains ten types of bearing dataset. In order to expand the number of training samples, a sample augmentation technique is employed in each bearing dataset. As shown in Figure 4, training samples are sliced with the window of 256 points. Then, 2000 samples are obtained from each bearing condition dataset.

In this study, there are four sets of datasets, which are dataset A, B, C and D. Dataset A, B and C correspond to three load conditions respectively, and each dataset obtains a total of 20,000 samples. Dataset D is generated by randomly selecting the data samples from the above three datasets, which consider the impact of three load condition. In each dataset, 14,000 samples are selected as training set, 3000 samples for validation set, and 3000 samples for testing set. The details of dataset D are shown in Table 1, and other datasets are similar. For each dataset, the training set is used to train the proposed ST-CNN model, and validation set is used to prevent possible overfitting and stop training process when the error rate decreases slightly or even starts to increase. Testing set is used to evaluate the performance of the proposed method.

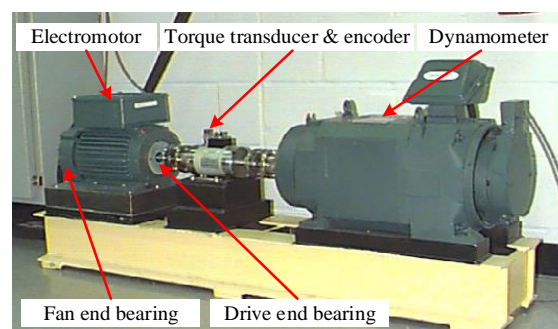


Figure 3. Experiment setup [39].

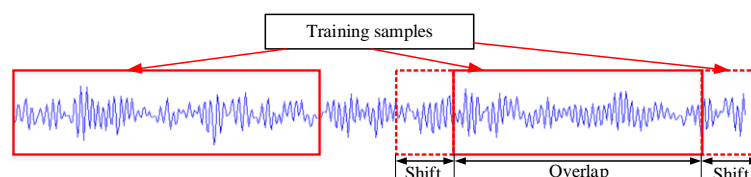


Figure 4. Data argument with overlap.

Table 1. Description of used dataset in case one.

Condition Type	Defect Severity (inch)	Dataset Division (Training/Validation/Testing)
Normal	0	1400/300/300
IRF	0.007	1400/300/300
IRF	0.014	1400/300/300
IRF	0.021	1400/300/300
ORF	0.007	1400/300/300
ORF	0.014	1400/300/300
ORF	0.021	1400/300/300
BF	0.007	1400/300/300
BF	0.014	1400/300/300
BF	0.021	1400/300/300

4.1.2. ST-CNN Testing Result

The parameters of the ST-CNN are shown in Table 2. Parameters are optimized by using mini-batch stochastic gradient descent with a batch size of 32. Table 3 shows the testing result of ten trials for the proposed method on four datasets. From the result, the proposed method is satisfactory in each dataset. In particular, the diagnosis performance on dataset D is outstanding. The mean accuracy of ten trials is 99.90%, maximum accuracy is 99.97%, minimum accuracy is 99.80%, and its standard deviation is 0.0570%. The standard deviation of ten trials shows a more reliable performance for the proposed method. Figure 5 presents the confusion matrix of the best of ten trials for dataset D. In the confusion matrix, rows stand for the actual label, and columns stand for predicted label for each condition. Seen from Figure 5, the overall diagnosis accuracy of ten condition is 99.97%, error rate is 0.03%, and ORF007, ORF014, ORF021, IRF007, IRF021, BF0.007, BF014, BF021 and Normal is 100%, while IRF014 is the worst one but its accuracy still reaches to 99.63% and it only has one error diagnosis. For the worst in the ten trials, there are overall five-error diagnosis and the fault diagnosis is satisfactory for 3000 testing samples. From Figure 5, the F_1 score [40] also could be obtained, and the mean F_1 score of all the class label is 99.97%. In addition, the mean F_1 score of all the class label of the worst trial is 99.92%.

Table 2. The parameters of the constructed CNN.

No.	Layer Type	No. of Filters	Kernel Size	Stride	Output Size	Padding
1	Convolution 1	32	2×2	2×2	64×128	No
2	Max-pooling 1	N/A	2×2	2	32×64	No
3	Convolution 2	64	2×2	2×2	16×32	No
4	Max-pooling 2	N/A	2×2	2	8×16	No
5	Convolution 3	128	2×2	2×2	4×8	No
6	Convolution 4	256	2×2	2×2	2×4	No
7	Max-pooling 3	N/A	2×2	2	1×2	No

Table 3. Result of the ST-CNN in different load conditions in case one (%).

Dataset	A	B	C	D
Max	100	100	100	99.97
Min	99.90	99.83	99.90	99.80
Mean	99.977	99.939	99.974	99.900
Std	0.0356	0.0642	0.0347	0.0570

Predicted condition	Normal	300 10.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	IRF007	0 0.0%	300 10.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	IFR014	0 0.0%	0 0.0%	299 9.97%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	IFR021	0 0.0%	0 0.0%	0 0.0%	300 10.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	ORF007	0 0.0%	0 0.0%	0 0.0%	0 0.0%	300 10.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	ORF014	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	300 10.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	ORF021	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	300 10.0%	0 0.0%	0 0.0%	100% 0.0%
	BF007	0 0.0%	0 0.0%	1 0.03%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	300 10.0%	0 0.0%	99.7 0.3%
	BF014	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	300 10.0%	100% 0.0%
	BF021	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	300 10.0%
		100% 0.0%	100% 0.0%	99.7% 0.3%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	99.97% 0.03%
		Normal	IRF007	IFR014	IFR021	ORF007	ORF014	ORF021	BA007	BA014	BA021
		Actual condition									

Figure 5. Condition classification confusion matrix in case one.

In the proposed ST-CNN, the size of the feature map from MP 3 has a great impact on the fully-connected layer and classification layer. The output size of these feature maps could be changed by modifying the stride of convolutional layer or pooling layer. The result of ten trials of these ST-CNNs with different output sizes is presented in Table 4. In these ST-CNNs, the output size 1×3 , output size 2×6 and output size 6×14 is better than output size 1×2 , and the mean accuracy of ten trials of output size 6×14 is 99.965%, the minimum mean accuracy of ten condition is 99.90%, the maximum mean accuracy of ten condition is 100%, and the standard deviation of ten trials is of 0.0299%. From the tuning result, the fault diagnosis of the proposed method is outstanding, and it is close to 100%.

Table 4. Result of the ST-CNN with the different output size (%).

Output Size	1×2	1×3	2×6	6×14	14×30	30×62
Max	99.97	99.97	100	100	100	100
Min	99.80	99.87	99.70	99.90	99.83	99.73
Mean	99.900	99.917	99.924	99.965	99.887	99.869
Std	0.0570	0.0356	0.1201	0.0299	0.0546	0.1026

4.1.3. Compared with Other Methods

Three time-frequency analysis methods based on the similar CNN compared with the proposed method. Three time-frequency analysis methods, that is, STFT, HHT and continuous wavelet transform (CWT), are used as well as ST. The comparison result is presented in Table 5 after ten trials. From the comparison result, the ST-CNN performs the higher and more reliable diagnosis performance in all ten trials than the three time-frequency analysis methods integrated with the similar CNN. Furthermore, the other methods including support vector machine (SVM) [41], k-nearest neighbor (KNN) [42], bagged trees (BT) [43] and linear discriminant (LD) [44] are compared with the proposed method. The comparison result is presented in Table 6 after ten trials. From the comparison results, the proposed ST-CNN achieves the higher diagnosis performance.

In this case study, all experiment methods are performed on a computer (Intel Core (TM) 3.6 GHz processor with 8GB of RAM) and a windows version of the tensorflow platform. All mentioned methods are trained by using the same training set. The training time and testing time of mentioned

methods are presented in Table 7. In this table, the training time of ST-CNN, CWT+CNN, HHT+CNN and STFT+CNN of 30 epochs and the testing time of one testing sample are calculated. The training time of SVM, KNN, BT and LD and the testing time of one testing sample are also counted, respectively. Seen from this table, the proposed ST-CNN consumes more time than STFT+CNN, SVM, KNN, BT and LD. It is because the process of converting sensor data with the 1D time series format into time-frequency complex matrix will consume some time. However, the capability of the computer has a great impact on the training and testing performance. In addition, the testing time of the proposed ST-CNN for one testing sample is only 8.1 ms, which is smaller the human reaction speed (100–400 ms). Therefore, the proposed method is suitable for real-time diagnosis of bearing.

Table 5. Comparison of bearing fault diagnosis using other time-frequency analysis methods (%).

Methods	ST	CWT	HHT	STFT
Max	100	99.17	98.83	97.50
Min	99.90	98.87	98.23	95.83
Mean	99.965	99.000	98.486	96.982
Std	0.0299	0.1141	0.1872	0.4447

Table 6. Comparison of bearing fault diagnosis result using different methods (%).

Methods	Mean Accuracy
ST-CNN	99.96
SVM	94.65
KNN	98.65
BT	71.70
LD	79.80

Table 7. Cost time for the proposed method and other method.

Methods	Training Time (s)	Testing Time (ms)
ST-CNN	4860.1	81
CWT + CNN	10981.4	179
HHT + CNN	18293.8	495
STFT + CNN	2902.9	24.6
SVM	256.7	1.0
KNN	85.1	1.3
BT	606.1	0.08
LD	8.76	0.05

4.2. Case Two: Bearing Fault Diagnosis in Different Fault and Load Conditions

4.2.1. Experimental Setup and Data Description

In this case, a new set of bearing fault datasets from the Mechanical Failures Prevention Group (MFPT) [45] are used to evaluate the diagnosis performance of the proposed method. These fault datasets consist of monitoring signals of NICE bearings, whose roller diameter is 0.235 inch, pitch diameter is 1.245 inch, number of elements is eight and their contact angle is zero. Among these datasets, inner ring fault and outer ring fault datasets are obtained under various loads. The speed of input shaft of the test rig is 25 Hz. The sampling rate of accelerometers mounted on the bearing is 48,828 Hz. Figure 6 shows two bearings with inner race fault (IRF) and outer race fault (ORF), respectively. Two bearings fault sensor data are collected by the accelerometer.

In this experiment, there are a total of six load conditions, which are 50, 100, 150, 200, 250 and 300 lbs. The fault dataset for each load condition consists of IRF and ORF dataset. Each type of bearing fault provides 2000 samples with 256 data point. Six fault datasets are available, which are dataset A,

dataset B, dataset C, dataset D, dataset E and dataset F. For each fault dataset, 2800 samples are used as training set, 600 samples as validation set, and 600 samples as testing set. Furthermore, dataset G contains 2800 training samples, 600 validation samples and 600 testing samples, which are uniformly extracted from six fault datasets. The details of these datasets are presented in Table 8.

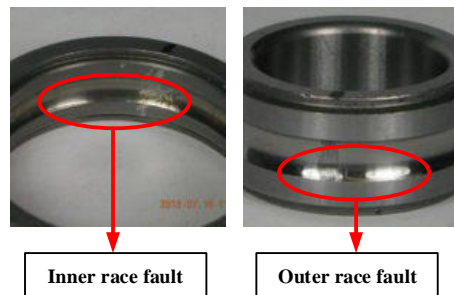


Figure 6. Experimental setup for bearing fault diagnosis.

Table 8. Description of the dataset in case two.

Fault Type	Dataset Division (Training/Validation/Testing)						
	Dataset A	Dataset B	Dataset C	Dataset D	Dataset E	Dataset F	Dataset G
IRF	1400/300/300	1400/300/300	1400/300/300	1400/300/300	1400/300/300	1400/300/300	1400/300/300
ORF	1400/300/300	1400/300/300	1400/300/300	1400/300/300	1400/300/300	1400/300/300	1400/300/300

4.2.2. ST-CNN Testing Result

The testing result of the ST-CNN is given in Table 9. From the testing result, the diagnosis performance of the ST-CNN is satisfactory in different datasets. The mean diagnosis accuracy of ten trials is 98.80% and the standard deviation of ten trials is 0.4270 on dataset G. Figure 7 shows the confusion matrix of the best diagnosis performance of the proposed method in ten trials. Seen from the confusion matrix, only two samples of 300 samples of each failure type are diagnosed incorrectly, and its F_1 score is 99.33%. For the worst of ten trials, 11 testing samples are diagnosed incorrectly which is also satisfactory.

To improve the diagnosis performance of the ST-CNN, the size of the feature maps of MP 3 is changed by modifying the stride of the convolutional layer or pooling layer. Table 10 shows the fault diagnosis performance of different ST-CNNs. Seen from the result, output size 1×3 , output size 2×6 , output size 6×14 and output size 30×62 are better than output size 1×2 . Additionally, the mean accuracy of output size 2×6 in ten trials is 99.50% and its standard deviation is 0.4073%.

Table 9. Result of the ST-CNN in different load conditions in case two (%).

Dataset	A	B	C	D	E	F	G
Max	99.67	99.83	99.50	97.67	98.17	98.33	99.33
Min	99.50	98.83	98.67	97.17	97.00	96.67	98.17
Mean	99.584	99.647	99.231	97.485	97.548	97.684	98.80
Std	0.1425	0.3373	0.2617	0.2000	0.4113	0.5047	0.4270

Table 10. Result of the ST-CNN with different output size (%).

Output Size	1×2	1×3	2×6	6×14	14×30	30×62
Max	99.33	99.50	99.83	99.33	99.83	99.00
Min	98.17	99.17	98.50	99.00	99.00	98.50
Mean	98.80	99.349	99.500	99.249	99.424	98.783
Std	0.4270	0.1222	0.4073	0.1155	0.2220	0.1925

Predicted condition	IRF	298 49.7%	2 0.3%	99.33% 0.67%
	ORF	2 0.3%	298 49.7%	99.33% 0.67%
		99.33% 0.67%	99.33% 0.67%	99.33% 0.67%
		IRF	ORF	
		Actual condition		

Figure 7. Condition classification confusion matrix in case two.

4.2.3. Compared with Other Methods

In the section, the diagnosis performance of the proposed ST-CNN is compared with other methods based on three different time-frequency analysis methods and the similar CNN. Table 11 shows the comparison result after ten trials. Seen from the result, the mean accuracy of CWT is 97.951%, HHT is 91.151%, STFT is 94.551%, and the ST-CNN preforms the higher diagnosis accuracy. In addition, the proposed method compared with other methods such as SVM, KNN, BT and LD. The comparison result is presented in Table 12. From the comparison result, the proposed method also preforms the higher performance than other methods. The training time and testing time of all experiment methods are shown in Table 13. In this table, the testing time is satisfactory.

Table 11. Comparison of bearing fault diagnosis using other time-frequency analysis methods (%).

Methods	ST	CWT	HHT	STFT
Max	99.83	98.33	92.33	95.50
Min	98.50	97.50	90.17	94.00
Mean	99.500	97.951	91.151	94.551
Std	0.4073	0.2604	0.7704	0.4373

Table 12. Comparison of bearing fault diagnosis result using different methods (%).

Methods	Mean Accuracy
ST-CNN	99.50
SVM	93.60
KNN	99.00
BT	93.30
LD	59.40

Table 13. Cost time for the proposed method and other method.

Methods	Training Time (s)	Testing Time (ms)
ST-CNN	1060.2	27
CWT + CNN	2439.2	35
HHT + CNN	683.1	6.7
STFT + CNN	3634.4	52
SVM	5.8	0.053
KNN	3.7	0.26
BT	10.6	0.091
LD	1.9	0.032

5. Conclusions

In this paper, a novel sensor data-driven fault diagnosis method is proposed based on the ST-CNN. The ST is introduced into the ST-CNN as the input processing layer named ST layer. The time-frequency

coefficients matrix with 2D format is obtained directly from original sensor data by the ST layer where no manual conversion work is needed. In this regard, the representative fault features are automatically extracted from original sensor data by training the constructed ST-CNN. Two public available bearing datasets are used to evaluate diagnosis performance. The testing and comparison results show that the proposed method could achieve higher and more reliable diagnosis performance for bearing faults than other existing methods.

In our future work, the fault feature extraction based on ST-CNN will be used to obtain the correlation features between multi-type faults and single faults, and the feasibility of the proposed method applied to the diagnosis of multi-type fault will be discussed.

Author Contributions: Methodology, G.Q. and J.W.; Algorithm Design, G.Q., C.D. and J.W.; Validation, J.W., X.B. and G.Q.; Formal Analysis, G.Q., C.D. and X.S.; Writing—Original Draft Preparation, G.Q. and C.D.; Writing—Review and Editing, J.W., C.D. G.Q. and X.S.; Visualization, J.W., G.Q. and C.D.; Project Administration, J.W. and X.S.; Funding Acquisition, J.W., C.D. and Y.W.

Funding: This work is funded in part by the National Natural Science Foundation of China under the Grant No. 51875225 and 51605095, in part by the National Key Research and Development Program of China under the Grant No. 2018YFB1702302, and in part by the Key Research and Development Program of Guangdong Province under the Grant No. 2019B090916001.

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

References

1. Cao, H.G.; Niu, L.K.; Xi, S.T.; Chen, X.F. Mechanical model development of rolling bearing-rotor systems: A review. *Mech. Syst. Signal Process.* **2018**, *102*, 37–58. [\[CrossRef\]](#)
2. Wu, J.; Wu, C.Y.; Cao, S.; Or, S.W.; Deng, C.; Shao, X.Y. Degradation data-driven time-to-failure prognostics approach for rolling element bearings in electrical machines. *IEEE Trans. Ind. Electron.* **2019**, *66*, 529–539. [\[CrossRef\]](#)
3. Loparo, K.A.; Adams, M.L.; Lin, W.; Abdel-Magied, F.M.; Afshari, N. Fault detection and diagnosis of rotating machinery. *IEEE Trans. Ind. Electron.* **2000**, *47*, 1005–1014. [\[CrossRef\]](#)
4. Jardine, A.K.; Lin, D.; Banjevic, D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mech. Syst. Signal Process.* **2006**, *20*, 1483–1510. [\[CrossRef\]](#)
5. Cerrada, M.; Sánchez, R.V.; Li, C.; Pacheco, F.; Cabrera, D.; Oliveira, J.V.; Vásquez, R.E. A review on data-driven fault severity assessment in rolling bearings. *Mech. Syst. Signal Process.* **2018**, *99*, 169–196. [\[CrossRef\]](#)
6. Feng, Z.P.; Zhou, Y.K.; Zuo, M.J.; Chu, F.L.; Chen, X.W. Atomic decomposition and sparse representation for complex signal analysis in machinery fault diagnosis: A review with examples. *Measurement* **2016**, *103*, 106–132. [\[CrossRef\]](#)
7. Obuchowski, J.; Zimroz, R.; Wylomańska, A. Blind equalization using combined skewness–kurtosis criterion for gearbox vibration enhancement. *Measurement* **2016**, *88*, 34–44. [\[CrossRef\]](#)
8. Zhao, H.Y.; Wang, J.D.; Han, H.; Gao, Y.Q. A feature extraction method based on HLMD and MFE for bearing clearance fault of reciprocating compressor. *Measurement* **2016**, *89*, 34–43. [\[CrossRef\]](#)
9. Wu, J.; Wu, C.; Lv, Y.; Deng, C.; Shao, X. Design a degradation condition monitoring system scheme for rolling bearing using EMD and PCA. *Ind. Manag. Data Syst.* **2017**, *117*, 713–728. [\[CrossRef\]](#)
10. Cheng, Y.W.; Zhu, H.P.; Wu, J.; Shao, X.Y. Machine health monitoring using adaptive kernel spectral clustering and deep long short-term memory recurrent neural networks. *IEEE Trans. Ind. Inf.* **2019**, *15*, 987–997. [\[CrossRef\]](#)
11. Wu, J.; Su, Y.H.; Cheng, Y.W.; Shao, X.Y.; Deng, C.; Liu, C. Multi-sensor information fusion for remaining useful life prediction of machining tools by adaptive network based fuzzy inference system. *Appl. Soft Comput.* **2018**, *68*, 12–23. [\[CrossRef\]](#)
12. Frank, P.M.; Ding, X. Frequency domain approach to optimally robust residual generation and evaluation for model based fault diagnosis. *Automatica* **1994**, *30*, 789–804. [\[CrossRef\]](#)
13. Kinnaert, M.; Peng, Y. Residual generator for sensor and actuator fault detection and isolation: A frequency domain approach. *Int. J. Control* **1995**, *61*, 1423–1435. [\[CrossRef\]](#)

14. Almeida, L.B. The fractional Fourier transform and time-frequency representations. *IEEE Trans. Signal Process.* **1994**, *42*, 3084–3091.
15. Ozaktas, H.M.; Arikan, O.; Kutay, A.; Bozdagi, C. Digital Computation of the Fractional Fourier Transform. *IEEE Trans. Signal Process.* **1996**, *44*, 2141–2150. [\[CrossRef\]](#)
16. He, M.; He, D. Deep learning based approach for bearing fault diagnosis. *IEEE Trans. Ind. Appl.* **2017**, *53*, 3057–3065. [\[CrossRef\]](#)
17. Wang, L.H.; Zhao, X.P.; Wu, J.X.; Xie, Y.Y.; Zhang, Y.H. Motor fault diagnosis based on short-time fourier transform and convolutional neural network. *Chin. J. Mech. Eng.* **2017**, *30*, 1–12. [\[CrossRef\]](#)
18. Ding, X.; He, Q. Energy-fluctuated multiscale feature learning with deep ConvNet for intelligent spindle bearing fault diagnosis. *IEEE Trans. Instrum. Meas.* **2017**, *66*, 1926–1935. [\[CrossRef\]](#)
19. Zhao, M.; Kang, M.; Tang, B.; Pecht, M. Deep residual networks with dynamically weighted wavelet coefficients for fault diagnosis of planetary gearboxes. *IEEE Trans. Ind. Electron.* **2018**, *65*, 4290–4300. [\[CrossRef\]](#)
20. Lou, X.; Loparo, K.A. Bearing fault diagnosis based on wavelet transform and fuzzy inference. *Mech. Syst. Signal Process.* **2004**, *18*, 1077–1095. [\[CrossRef\]](#)
21. Samanta, B.; Al-Balushi, K.R. Artificial neural network based fault diagnostics of rolling element bearings using time-domain features. *Mech. Syst. Signal Process.* **2003**, *17*, 317–328. [\[CrossRef\]](#)
22. Malhi, A.; Gao, R.X. PCA-based feature selection scheme for machine defect classification. *IEEE Trans. Ins. Meas.* **2004**, *53*, 1517–1525. [\[CrossRef\]](#)
23. 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. [\[CrossRef\]](#)
24. Zhao, X.; Jia, M. Fault diagnosis of rolling bearing based on feature reduction with global-local margin Fisher analysis. *Neurocomputing* **2018**, *315*, 447–464. [\[CrossRef\]](#)
25. Zhang, J.; Ma, W.; Lin, J.; Ma, L.; Jia, X. Fault diagnosis approach for rotating machinery based on dynamic model and computational intelligence. *Measurement* **2015**, *59*, 73–87. [\[CrossRef\]](#)
26. Singh, D.S. QingZhao Pseudo-fault signal assisted EMD for fault detection and isolation in rotating machines. *Mech. Syst. Signal Process.* **2016**, *81*, 202–218. [\[CrossRef\]](#)
27. Jia, F.; Lei, Y.; Lin, J.; Zhou, X.; Lu, N. Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. *Mech. Syst. Signal Process.* **2015**, *73*, 303–315. [\[CrossRef\]](#)
28. Sun, W.; Shao, S.; Zhao, R.; Yan, R.; Zhang, X.; Chen, X. A Sparse Auto-encoder-Based Deep Neural Network Approach for Induction Motor Faults Classification. *Measurement* **2016**, *89*, 171–178. [\[CrossRef\]](#)
29. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Netw.* **2015**, *61*, 85–117. [\[CrossRef\]](#)
30. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [\[CrossRef\]](#)
31. Ince, T.; Kiranyaz, S.; Eren, L.; Askar, M.; Gabbouj, M. Real-time motor fault detection by 1-D convolutional neural networks. *IEEE Trans. Ind. Electron.* **2016**, *63*, 7067–7075. [\[CrossRef\]](#)
32. Han, Y.; Tang, B.; Deng, L. Multi-level wavelet packet fusion in dynamic ensemble convolutional neural network for fault diagnosis. *Measurement* **2018**, *127*, 246–255. [\[CrossRef\]](#)
33. Zhu, D.; Gao, Q.W.; Sun, D.; Lu, Y.X.; Peng, S.L. A detection method for bearing faults using null space pursuit and s transform. *Signal Process.* **2014**, *96*, 80–89. [\[CrossRef\]](#)
34. Li, B.; Zhang, P.L.; Liu, D.S.; Mi, S.S.; Ren, G.Q.; Tian, H. Feature extraction for rolling element bearing fault diagnosis utilizing generalized s transform and two-dimensional non-negative matrix factorization. *J. Sound Vib.* **2011**, *330*, 2388–2399. [\[CrossRef\]](#)
35. Djurovi, I.; Sejdi, E.; Jiang, J. Frequency-based window width optimization for S-transform. *AEU Int. J. Electron. Commun.* **2008**, *62*, 245–250. [\[CrossRef\]](#)
36. Stockwell, R.G.; Mansinha, L.; Lowe, R.P. Localization of the complex spectrum: The S transform. *IEEE Trans. Signal Process.* **1996**, *44*, 998–1001. [\[CrossRef\]](#)
37. Simon, C.; Ventosa, S.; Schimmel, M.; Heldring, A. The S-transform and its inverse: Side effects of discretizing and filtering. *IEEE Trans. Signal Process.* **2007**, *55*, 4928–4937. [\[CrossRef\]](#)
38. Liu, W.Y.; Wen, Y.D.; Wen, Z.D. Large-Margin Softmax Loss for Convolutional Neural Networks. In Proceedings of the International Conference Machine Learning (ICML), New York, NY, USA, 19–24 June 2016; pp. 507–516.

39. Loparo, K. Case Western Reserve University Bearing Data Centre Website. Available online: <http://csegroups.case.edu/bearingdatacenter/pages/download-data-file> (accessed on 20 September 2017).
40. He, H.; Garcia, E.A. Learning from imbalanced data. *IEEE Trans. Knowl. Data Eng.* **2009**, *21*, 1263–1284.
41. Aydmj, T.; Duin, R.P.W. Pump Failure Determination Using Support Vector Data Description. In *Advances in Intelligent Data Analysis (Lecture Notes in Computer Science)*; Springer: Berlin, Germany, 1999; pp. 415–425.
42. Gao, R.X.; Chen, X. Wavelets for fault diagnosis of rotary machines: A review with applications. *Signal Process.* **2014**, *96*, 1–15.
43. Mishra, P.K.; Yadav, A.; Pazoki, M. A Novel Fault Classification Scheme for Series Capacitor Compensated Transmission Line Based on Bagged Tree Ensemble Classifier. *IEEE Access* **2018**, *6*, 27373–27382. [[CrossRef](#)]
44. Jin, X.; Zhao, M.; Chow, T.W.S.; Pecht, M. Motor Bearing Fault Diagnosis Using Trace Ratio Linear Discriminant Analysis. *IEEE Trans. Ind. Electron.* **2014**, *61*, 2441–2451. [[CrossRef](#)]
45. Machinery Failure Prevention Technology (MFPT) Datasets. Available online: <http://www.mfpt.org/FaultData/FaultData.htm> (accessed on 17 January 2013).



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).