



Article TF Entropy and RFE Based Diagnosis for Centrifugal Pumps Subject to the Limitation of Failure Samples

Xuanyuan Su^{1,2}, Hongmei Liu^{1,2} and Laifa Tao^{1,2,*}

- ¹ School of Reliability and Systems Engineering, Beihang University, Beijing 100191, China; Suxuanyuan@buaa.edu.cn (X.S.); liuhongmei@buaa.edu.cn (H.L.)
- ² Science & Technology Laboratory on Reliability & Environmental Engineering, Beihang University, Beijing 100191, China
- * Correspondence: taolaifa@buaa.edu.cn

Received: 31 March 2020; Accepted: 20 April 2020; Published: 23 April 2020



Abstract: In practical engineering, the vibration-based fault diagnosis with few failure samples is gaining more and more attention from researchers, since it is generally hard to collect sufficient failure records of centrifugal pumps. In such circumstances, effective feature extraction becomes quite vital, since there may not be enough failure data to train an end-to-end classifier, like the deep neural network (DNN). Among the feature extraction, the entropy combined with signal decomposition algorithms is a powerful choice for fault diagnosis of rotating machinery, where the latter decomposes the non-stationary signal into multiple sequences and the former further measures their nonlinear characteristics. However, the existing entropy generally aims at processing the 1D sequence, which means that it cannot simultaneously extract the fault-related information from both the time and frequency domains. Once the sequence is not strictly stationary (hard to achieve in practices), the useful information will be inevitably lost due to the ignored domain, thus limiting its performance. To solve the above issue, a novel entropy method called time-frequency entropy (TfEn) is proposed to jointly measure the complexity and dynamic changes, by taking into account nonlinear behaviors of sequences from both dimensions of time and frequency, which can still fully extract the intrinsic fault features even if the sequence is not strictly stationary. Successively, in order to eliminate the redundant components and further improve the diagnostic accuracy, recursive feature elimination (RFE) is applied to select the optimal features, which has better interpretability and performance, with the help of the supervised embedding mechanism. To sum up, we propose a novel two-stage method to construct the fault representation for centrifugal pumps, which develops from the TfEn-based feature extraction and RFE-based feature selection. The experimental results using the real vibration data of centrifugal pumps show that, with extremely few failure samples, the proposed method respectively improves the average classification accuracy by 12.95% and 33.27%, compared with the mainstream entropy-based methods and the DNN-based ones, which reveals the advantage of our methodology.

Keywords: fault diagnosis; time-frequency entropy-based feature extraction; recursive feature elimination-based feature selection; centrifugal pumps

1. Introduction

The centrifugal pump is one of the most critical elements in hydraulic systems [1], which has been widely applied to the modern industry. Due to long-term running in the harsh environment, there is an increasing need to develop and improve the technique of fault diagnosis for centrifugal pumps, to avoid the consequent loss of manpower and economy [2]. As the typical rotating equipment,

the vibration-based fault diagnosis for centrifugal pumps with few fault samples has been attracting more and more attention from scholars, due to the significant meaning of its practical application [3].

By investigating the literature published in recent years that focused on vibration-based fault diagnosis, we divide these fault diagnosis methods into the following two categories: deep model-based approaches and non-deep model-based approaches.

The first category of approaches utilize the deep neural networks [4] to automatically learn fault representations from the raw data, which can generate quite admirable performance and rarely need the feature extraction designed manually, when the training data are sufficient [5]. However, there are usually only a few failure samples of the complex rotating machineries like the centrifugal pump in practical engineering, and it thus limits the further application of these deep model-based approaches, due to their great demand for failure samples and uninterpretable processes of fault diagnosis [6].

The second category of approaches usually do not depend on an overly complex classifier model, where the design of the feature engineering is emphasized to obtain the effective fault representations for the mode of objects [7]. Therefore, these approaches can bring more satisfying and interpretable results in scenarios with relatively few failure samples [8]. The above characteristics of these non-deep model-based approaches have led to their continuous development. In general, these methods consist of two main stages [9]: the feature extraction and the feature selection.

In the feature extraction stage, the raw signal will be converted into a feature vector that can effectively indicate the fault characteristics of equipment. As a statistical measure, entropy can effectively measure the complexity and dynamic change of nonlinear signals and has thus been widely applied to fault diagnosis for rotating machinery [10]. In general, entropy, along with signal processing algorithms such as empirical model decomposition (EMD) [11], ensemble empirical model decomposition (EEMD) [12] and wavelet transform (WT) [13], etc. are the common combinations to extract the fault features from the nonlinear and non-stationary signals, where the signal is first decomposed into multiple approximatively stationary time series and the entropy is further transformed by these series into the feature vector. Up to now, the entropy has evolved into different variants, such as spectrum entropy (SpectEn) [14], approximate entropy (ApEn) [15], sample entropy (SampEn) [16], etc. to adapt to different scenarios. The aforementioned variants provide some new ideas for feature extraction, but there are still two key problems to be solved.

The first one is the loss of useful information. Regardless of the variants, the existing entropy is generally aimed at processing the 1D sequence, which can only measure the complexity and irregularity of one domain in time or frequency. Even with the signal processing algorithms, the decomposed 1D sequence is almost impossible to be strictly stationary. In this case, parts of useful information will inevitably be lost due to a certain dimension (time or frequency) being ignored. The second one is the low computation efficiency. Due to the operation of state space reconstruction and lots of loop comparisons, the mainstream entropy methods, such as the ApEn and SampEn usually need considerable calculation time [10], which limits their applications in practices.

In the feature selection stage, the redundant components will be eliminated from the fault features and the optimal feature subset with lower dimensions will be obtained. In order to overcome the curse of dimension [17] and further improve the performance of fault diagnosis, there are large amounts of feature selection methods applied to select optimal feature subset from the high-dimensional features. As an unsupervised dimensionality reduction method, principal component analysis (PCA) obtains the most widely applications on the feature selection, because of its simple calculation and stable effects [18,19]. However, the PCA still has two shortcomings that need to be solved. The first one is that the feature subset selected by PCA is usually difficult to achieve the satisfied diagnostic accuracy, due to its unsupervised mechanism. Another one is that the feature subset is fused from the raw feature set, and it thus damages the original physical meaning. In scenarios where the diagnostic accuracy and feature interpretability are well emphasized, the PCA is limited to further applications, due to the above-mentioned shortcomings [20].

To solve the above-mentioned issues, a novel fault representation method is designed and proposed in this paper, which is mainly composed of the time-frequency entropy (TfEn)-based feature extraction and recursive feature elimination (RFE)-based feature selection. Compared with the existing mainstream entropy-based methods aimed at processing the 1D sequence, our proposed method can jointly highlight and measure the intrinsic characteristics of nonlinear and non-stationary signals from both dimensions of time and frequency, and it can achieve more accurate fault diagnosis, even with limited failure samples. In the feature extraction stage, TfEn is proposed to comprehensively quantify the complexity and dynamic change by taking into account nonlinear behaviors of signals in the form of a 2D time-frequency matrix. In the feature selection stage, RFE is applied to select the optimal feature subset and eliminate the redundant components. Abandoning the feature fusion of PCA, RFE retains the original physical meanings of selected features, and thus has the better interpretability [21]. In addition, it is worth noting that our proposed method has no rigorous requirements for calculation resources and failure samples, which means that it more suitable for practical scenes with few failure samples. To sum up, the main contributions of this paper can be listed as follows:

- 1. A novel feature extraction called TfEn is proposed to jointly quantify the robust nonlinear characteristics from both time and frequency dimensions, which overcomes the limitations of existing entropy based on 1D sequence and thus being more suitable for non-stationary signals;
- 2. RFE-based feature selection is applied to select the optimal feature subset form the high-dimensional feature vector, which achieves better performance on improving the classification accuracy and retaining the feature interpretability compared with mainstream methods;
- 3. The proposed two-stage fault representation has the capability of highlighting and depicting the intrinsic differences among vibration signals, which presents the stability against the variable sizes of training samples, and obtains the satisfied results even with extremely few failure data;
- 4. Compared with the existing entropy variants and mainstream deep models, the experiments with the real vibration signals of centrifugal pumps illustrate the validity and superiority of our proposed fault diagnosis method.

The remaining of the paper is organized as follows. Section 2 introduces the necessary background. Section 3 describes the proposed fault diagnosis. Section 4 implements the experiments, then presents and discusses the results. Finally, the conclusion is shown in Section 5.

2. Preliminaries

This section presents the necessary background of our proposed method. Complementary ensemble empirical mode decomposition (CEEMD), is applied to process the raw signals of multiple components in this paper. Its principle is introduced as follows.

As an adaptive decomposition algorithm, CEEMD [22] can decompose non-stationary signals into several approximately stationary intrinsic mode functions (IMFs) and non-stationary residual terms. The brief process of CEEMD is described as follows:

1. Given a raw signal x(t). First, add the particular white noise into the signal x(t) and decompose it. Repeat the decomposition under different noise realizations and average these results as the first sub-signal IMF_1 :

$$IMF_{1} = \frac{1}{N} \int_{i=1}^{N} E_{1}[x(t) + \sigma_{1}\omega_{i}], \qquad (1)$$

where *N* is the average time, $E_1[]$ is the defined operator which produces the 1th mode obtained from the given signal using the EMD, σ_1 is the ratio coefficient and ω_i is the particular white noise in each time.

2. Compute the first-order residual as follows:

$$r_1 = x(t) - IMF_1, \tag{2}$$

3. Set $r_1 + \sigma_1 E_1[\omega_i]$, i = 1, ..., N as the new signal and compute the *IMF*₂:

$$IMF_{2} = \frac{1}{N} \int_{i=1}^{N} E_{1}[x(t) + \sigma_{1}E_{1}[\omega_{i}]], \qquad (3)$$

4. Repeating step (1) to step (3) until the IMF_{n+1} is obtained, and the raw signal is finally decomposed into *S* sub-signals IMF_s and a residual R(t):

$$IMF_{n+1} = \frac{1}{N} \int_{i=1}^{N} E_1[x(t) + \sigma_n E_n[\omega_i],$$
(4)

$$x(t) = \sum_{s=1}^{S} IMF_s + R(t).$$
 (5)

Through the CEEMD, the non-stationary signal is decomposed into several approximately stationary sub-signals, and these sub-signals will be support for the subsequent feature extraction.

3. Proposed Fault Diagnosis Method

In this paper, we propose a novel fault diagnosis method for centrifugal pumps based on TfEn and RFE. The section is composed of four parts: procedure of the fault diagnosis, CEEMD-based signal processing, proposed TfEn-based feature extraction and RFE-based feature selection. In each subsection, the specific process, innovations and related logic of the methodology will be introduced respectively.

3.1. Procedure of Fault Diagnosis

In this subsection, a novel fault diagnosis method is proposed for centrifugal pumps with few failure samples, which is mainly composed of the CEEMD-based signal processing, TfEn-based feature extraction and RFE-based feature selection. The procedure of our propose fault diagnosis method is shown in Figure 1.

As shown in Figure 1, the main steps of our proposed method can be represented as follows:

- 1. Decompose the non-stationary signal into several sub-signals using CEEMD to obtain the multiplescale components;
- 2. Using our proposed TfEn, transform each non-stationary or approximately stationary component into the 2D time-frequency matrix, and construct their fault representation by taking into account nonlinear behaviors from both the time and frequency domain;
- 3. Select the optimal TfEn feature subset from the high-dimensional feature set using the RFE, where the redundant components are eliminated automatically;
- 4. Choose classifier models to implement the fault diagnosis for centrifugal pumps, based on the optimal TFIE feature subset.

3.2. CEEMD-Based Signal Processing

The signal processing is the first step of our proposed method.

Typically, the vibration signal is composed of multiple narrowband time series. In order to extract the multi-scale intrinsic feature of signals, CEEMD is utilized to adaptively decompose the non-stationary raw signal into several sub-signals.



Figure 1. The procedure of the proposed fault diagnosis method.

Fault diagnosis

Among all the adaptive decomposition algorithms, EMD is the one that is earliest proposed and obtains the most applications in fault diagnosis. However, the EMD has the problem called mode mixing [23]. Although, EEMD, a variant of EMD had been proposed to deal with this problem through the ensemble of original signal with Gaussian noise. Considering all the above issues, CEEMD, a more complementary variant of EMD, is applied in this paper to obtain the better separated modes, which combines the improved noise addition and ensemble strategies [22].

As described in Section 2 given a raw vibration signal x(t), it will be adaptively decomposed into *S* sub-signals IMF_s , s = 1, 2, ..., S. Taking a vibration signal under the inner-race fault mode as the example, the 10 sub-signals decomposed by CEEMD is shown in Figure 2.

As shown in Figure 2, 10 Imfs are sequentially decomposed in order of the frequency using CEEMD. We can see that the first 5 Imfs are approximately stationary, where their means and variances are approximately independent of time. However, as shown in the red dashed boxes in the figure, the last 5 Imfs present relatively more obvious non-stationary characteristics, namely, the means and variances of the series present large differences at different time instants.

To sum up, multiple narrowband sub-signals are decomposed from the non-stationary vibration signal using CEEMD, which depict the intrinsic characteristic of raw signals from the different scale. However, it should be noted that these components are not strictly stationary, which means that both time and frequency dimensions need to be considered jointly, so as to fully extract effective fault-related information.



Figure 2. 10 sub-signals decomposed by complementary ensemble empirical mode decomposition (CEEMD).

3.3. Proposed TfEn-Based Feature Extraction

Entropy is the quantitative tool to describe the complexity and irregularity [24] of systems, which varies with the changing status of systems. For the centrifugal pumps, if the signal is approximately stationary (such as the first 5 Imfs shown in Figure 2), we usually think that the distribution of signals in the frequency domain better reflect the status of objects, where the existing entropy aiming at 1D sequence may be enough to measure their complexity. While, if the signal is close to non-stationary (such as the last 5 Imfs shown in Figure 2), both the time domain and frequency domain contain the useful information reflecting the status of objects. In this case, the existing entropy cannot comprehensively quantify the irregularity of signals in two dimensions, which inevitably bring about the loss of useful information.

Aiming at addressing the above-mentioned issues, we propose TfEn to jointly measure the complexity and dynamic change though taking into account the nonlinear behaviors from both dimensions of the time and frequency. The process of TfEn is introduced as follows.

1. Given a 1D time series x(t). For the purpose of measuring its complexity both in time and frequency domain, x(t) should be first extended into the 2D time-frequency matrix. In this paper, short-time Fourier transform (STFT) [25], a classical time-frequency analysis method, is applied to achieve the above target considering its balance of calculation efficiency and performance.

$$X(t,f) = \int_{-\infty}^{\infty} x(\tau)w(\tau-t)e^{-j2\pi f\tau}d\tau,$$
(6)

where w(t) is the window function, t and τ are both the time variables and f is the frequency variable. X(t, f) can be illustrated as the correlation between $x(\tau)$ and $w(\tau - t)e^{-j2\pi f\tau}$.

2. Further, through the square operation, the X(t, f) will be transformed into the energy matrix.

$$M = |X(t,f)|^{2} = \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1Y} \\ e_{21} & e_{22} & \cdots & e_{2Y} \\ \vdots & \vdots & e_{ij} & \vdots \\ e_{X1} & e_{X2} & \cdots & e_{XY} \end{bmatrix}, i = 1, \dots, X, j = 1, \dots, Y,$$
(7)

where the *M* is a 2D matrix consisting of $X \times Y$ elements e_{ij} , and e_{ij} can be interpreted as the energy located at frequency *i* and time *j*.

3. According to the Shannon methodology, the average degree of uncertainty depicts the irregularity of probability system. In this case, to obtain the robust representation of the time-frequency energy matrix M, we utilize a sliding window of size $F \times T$ to divide the M into $I \times J$ blocks $W_{i,j}$, i = 1, ..., I, j = 1, ..., J. Then a normalization operator will process these blocks one by one and transform them into a probability density matrix Q. The above operation can be illustrated in Figure 3.



Figure 3. Matrix division and normalization operations using time-frequency entropy (TfEn).

As shown in Figure 3, the F, T, L_f and L_t are 4 hyper-parameters of our proposed TfEn, which respectively represent the span of frequency-axis, span of time-axis, translation span of frequency-axis and translation span of time-axis of each time-frequency block.

4. As shown in Figure 3, through the normalization operators, each time-frequency energy block $W_{i,j}$ will be transformed into a probability density reflecting its irregularity. The calculation of the normalization operators is described as follows:

$$w_{i,j} = \sum_{y=1+L_t \times j}^{T+L_t \times j} \sum_{x=1+L_f \times i}^{F+L_f \times i} e_{x,y},$$
(8)

$$A = \sum_{i=1}^{I} \sum_{j=1}^{J} w_{i,j},$$
(9)

$$q_{i,j} = \frac{w_{i,j}}{A},\tag{10}$$

where $e_{x,y}$ is an energy element of the matrix M at coordinate (x, y), $w_{i,j}$ is defined as the sum of energy values in the block $W_{i,j}$, and A is defined as the total sum of energy elements of all blocks. $q_{i,j}$ is the normalized probability density for the block $W_{i,j}$, which quantifies the concentration of energy in each local time-frequency block.

5. According to the normalized probability density of all these blocks, the TfEn feature of the given time series x(t) can be extracted as follows:

$$s(q) = -c \sum_{i=1}^{I} \sum_{j=1}^{J} q_{i,j} \ln q_{i,j},$$
(11)

where *c* is an arbitrary constant and it is generally set as one, s(q) is the TfEn feature of 1D time series, which jointly represent its complexity and irregularity from both time and frequency domains.

6. Repeating the above operations for all the *S* sequences decomposed by CEEMD, we will finally obtain the TfEn feature vector TfEns of size $1 \times S$:

$$TfEns = \{s(q)_1, s(q)_2, \dots, s(q)_s, \dots, s(q)_S\},$$
(12)

with the help of TfEns, we obtain the multi-scale representations reflecting the complexity and dynamic change of nonlinear and non-stationary vibration signals. Besides that, it is worth noting that TfEn requires no state space reconstruction and a large amount of loop comparisons, so it has higher computational efficiency than the methods such as SampleEn and ApEn.

3.4. RFE-Based Feature Selection

After the TfEn-based feature extraction, we obtain a feature set TfEns representing the status of centrifugal pumps. The high-dimensional feature set may contain the sufficient status-related information, while it also has a price: increasing dimensions of the feature set leads to the sparseness of the feature space, which will cause the overfitting problem when the training samples are scare. Considering that there is usually not enough failure data in the practical engineering, the feature selection is a necessary operation to ensure the accuracy of fault diagnosis.

In this subsection, a supervised feature selection method called the recursive feature elimination (RFE) [26] is applied to solve the above-mentioned issues, which can automatically select the optimal feature subset from the high-dimensional feature set. The key steps of RFE-based feature selection are illustrated in Figure 4.



Figure 4. The flow chart of recursive feature elimination (RFE)-based feature selection.

As shown in Figure 4, the RFE-based feature selection is an iterative process, which utilize the criteria developed from the coefficients in a support vector machine (SVM) model to assess features and recursively removes the features with the small criteria.

1. Given a set of training samples $\{x_i, y_i\}, i = 1, ..., N$, where x_i is the feature vector sample with *S* features, y_i is the status label of centrifugal pumps. A linear SVM model is trained using these training samples, and its decision function is:

$$f(X) = \omega \cdot X + b, \tag{13}$$

2. According to the trained SVM model, the ranking criterion of features can be calculated as follows:

$$\omega = \sum_{i=1}^{N} \alpha_i y_i x_i,\tag{14}$$

$$J(s) = \omega_s^2, \tag{15}$$

where α_i is the Lagrange multiplier, ω is the weight vector of the trained SVM model with *S* elements ω_s , J(s) is the ranking criterion for the feature $s(q)_s$.

3. Eliminate the feature $s(q)_k$ with the lowest criterion from the feature set:

$$s(q)_k = \operatorname{argmin}(J), \tag{16}$$

$$TfEns^{1} = \{s(q)_{1}, s(q)_{2}, \dots, s(q)_{k-1}, s(q)_{k+1}, \dots, s(q)_{s}\},$$
(17)

where $T f E n^1$ is the feature subset with S - 1 features in the first iteration of feature selection.

4. Repeating the above operations until all the features are eliminated from the feature set, we will obtain the *S* feature subsets:

$$TfEns = \{s(q)_1, s(q)_2, \dots, s(q)_s, \dots, s(q)_s\},$$
(18)

using the cross-validation technique [27], the feature subset $TfEns^*$ with the best accuracy will be selected as the optimal feature subset for the subsequent fault diagnosis.

4. Experiments

4.1. Description of Dataset

In this experiment, the real vibration data of the centrifugal pumps is utilized to verify the effectiveness of out proposed fault diagnosis method. The vibration signals are acquired from an installed accelerometer, with a sampling rate of 10.24k Hz, and their engineering unit is millivolt (mV), which is shown in Figure 5.



Figure 5. Data acquisition system of the centrifugal pump.

There are 5 typical modes of centrifugal pumps in our dataset, which are respectively the inner-race bearing fault, outer-race bearing fault, ball fault, impeller fault and normal status. For each mode of the centrifugal pumps, we divide and obtain 60 samples according to the standard of 5120 points per sample, which means that there are a total of $5 \times 60 = 300$ samples in our dataset.

In order to test the performance of the proposed method under the scenarios of few failure samples, we divide the raw dataset into the training set and the testing set according to the ratio of 1: 5, which means that there are 10 samples under each fault mode to be used to train the fault classification model. In addition to that, in the Section 4.3, we set up additional several sets of experiments with training samples of smaller sizes, to further compare the adaptability on the extremely few failure samples between the proposed method and other existing mainstream ones.

4.2. Fault Diagnosis Using the Proposed Method

In this subsection, using the dataset divided by 10 training samples and 50 testing samples under each mode, we validate the proposed method, and the corresponding results of each key step will be presented and analyzed in the following.

A. CEEMD-based signal processing

As the first step of our proposed fault diagnosis method, CEEMD is utilized to adaptively decompose the raw signal into several sub-signals called Imfs. The related hyper-parameters are set as follows: the noise standard deviation (Nstd) is set as 0.2, the number of ensemble (Ne) is set as 60 and the shifting maximum number of iterations (MaxIter) is set as 500. Taking the modes of normal and outer-race fault as the examples, the corresponding decomposition results are illustrated in Figures 6 and 7.



Figure 6. Raw signal and decomposed intrinsic mode functions (Imfs) under the normal mode.



Figure 7. Raw signal and decomposed Imfs under the outer-race fault.

As shown in Figures 6 and 7, using the CEEMD, each vibration signal is adaptively decomposed into 10 Imfs, which provide the multiple-scale representations for the status of centrifugal pumps. It is worth noting that some decomposed components still exhibit quite obvious non-stationarity characteristics, even when processed by CEEMD, which means that both their time domain and frequency domain need to be comprehensively considered during the subsequent feature extraction.

B. TfEn-based feature extraction

To more fully take advantage of the useful information, in the proposed TfEn, each decomposed Imf is first transformed into the time-frequency matrix, and then each 2D matrix is further divided and calculated to the energy density matrix using the normalization operator. The energy density matrix exhibits the irregularity and dynamic change from both the time and frequency domain, and it can be finally quantified as a TfEn feature.

Specifically, the related hyper-parameters of STFT are set as follows: the window length (window) is set as 512, the cover length (noverlap) is set as 510, the Fourier transform length (nfft) is set as 512 that is equal to the window length. And the hyper-parameters of normalization operator are defined: the span of frequency-axis (F) is set as 64, the span of time-axis (T) is set as 4, the translation span of frequency-axis (L_f) is set as 32 and the translation span of time-axis (L_t) is set as 2.

Repeating the above operations for all 10 Imfs decomposed by CEEMD, we will get the TfEn feature vector of size 10×1 . Limited to the paper space, we only present the corresponding results of the Imf_1 under the modes of normal and outer-race fault, which are shown in Figures 8 and 9.



Figure 8. Two-dimensional time-frequency matrix (left); energy density matrix (right) under the normal mode.



Figure 9. Two-dimensional time-frequency matrix (**left**); energy density matrix (**right**) under the outerrace fault.

As shown in Figures 8 and 9, 2D time-frequency matrices exhibit the difference of energy distribution among the Imf₁s under the modes of normal and outer-race fault. Through the calculation of normalization operators block by block, these 2D matrices are further extracted and compressed into the energy density matrices of smaller size. The normalization operator is similar to the convolution filters in convolutional neural networks, which condenses each low-level local information into the high-level features. As shown by these energy density matrices, the irregularity of time-frequency distribution under each mode are further highlighted and depicted.

As described in Section 3.3, each density matrix is finally quantified as a TfEn feature, it measures the complexity and dynamic change of each Imf sub-signal from both time and frequency domains. Repeating the above operations for all the other Imfs, we obtain the TfEn feature vector including 10 elements.

C. RFE-based feature selection

For the purpose of further improving the performance of the fault diagnosis, the RFE algorithm is utilized to eliminate the redundant components from the feature set and then select the optimal

feature subset. With the help of cross-validation technique, the above process of feature selection is totally automatic.

As described in Section 3.4, we finally obtain the ranking of features by training the linear SVM using the $5 \times 10 = 50$ training samples. The ranking of our 10 TfEn features is presented in Table 1.

Feature No.	Ranking
$s(q)_1$	1
$s(q)_2$	1
$s(q)_3$	1
$s(q)_4$	1
$s(q)_5$	6
$s(q)_6$	5
$s(q)_7$	4
$s(q)_8$	3
$s(q)_9$	2
$s(q)_{10}$	1

Table 1. Feature ranking obtained by RFE.

As shown in Table 1, a 5-dimensional feature vector including $s(q)_1$, $s(q)_2$, $s(q)_3$, $s(q)_4$ and $s(q)_{10}$ is finally selected as the optimal feature subset $TfEns^*$. In the optimal feature subset, the first 4 Imfs occupy the dominant, which is consistent with the general experience and intuition of fault diagnosis. While contrary to common sense, the last 3 non-stationary components also obtain quite a high ranking, and the last Imf is also selected as the optimal feature. The above-mentioned phenomenon indicates that it is not reasonable to directly abandon the non-stationary components without analysis, and it also demonstrates that it is necessary to take into account both time and frequency domains during the process of feature extraction.

For the comparisons, another two mainstream feature selection methods, PCA [28] and kernel PCA (KPCA) [29], are also applied to select the optimal TfEn features. In order to understand the performance of these feature selection methods more intuitively, t-distributed stochastic neighbor embedding (t-SNE), a popular technique of dimensionality reduction, is applied to respectively transform the total raw features and the optimal features selected by PCA, KPCA, RFE into the 2D feature space, their corresponding scatter plots with 250 testing samples, are shown in Figure 10.

As shown in Figure 10, we can see that the scatters of features selected by RFE achieve the best resolution under 5 modes. Using these RFE-based optimal features, the difference of TfEn between different fault modes are further highlighted.

D. Fault classification

Considering the performance of our fault representations and the limitation of training samples, a classical classifier, SVM model, is employed to implement the fault classification.

To obtain the SVM model with the best accuracy, we utilize the grid search with cross validation (GSCV) technique [30] to train the SVM models with several hyper-parameters sets. Through the process of the GSCV, the SVM model with the hyper-parameter set {C = 1.5, 'kernel' = 'rbf', gamma = 'auto'} is selected as the final classifier to implement the fault classification for centrifugal pumps. In addition to that, the formula for accuracy is defined as Accuracy = Rights/(Rights + Errors) × 100%, and the corresponding diagnostic results are illustrated in Table 2.



Figure 10. Scatters of total TfEn features (**a**); scatters of optimal TfEn features selected by principal component analysis (PCA) (**b**); scatters of optimal TfEn features selected by kernel PCA (KPCA) (**b**); scatters of optimal TfEn features selected by RFE (**d**).

Method	Mode	Accuracy	Errors	Rights
	Normal	100.00%	0	50
TfEn	Inner-race fault	100.00%	0	50
	Outer-race fault	100.00%	0	50
	Ball fault	100.00%	0	47
	Impeller fault	94.34%	3	50
	Average/Total	98.80%	3	247
	Normal	100.00%	0	50
	Inner-race fault	100.00%	0	50
TEP DCA	Outer-race fault	75.00%	13	39
ITEN-PCA	Ball fault	73.91%	12	34
	Impeller fault	94.23%	3	49
	Average/Total	88.80%	28	222
	Normal	91.11%	4	41
	Inner-race fault	100.00%	0	46
TfEn-KPCA	Outer-race fault	60.24%	33	50
	Ball fault	86.49%	5	32
	Impeller fault	94.87%	2	37
	Average/Total	82.40%	44	206
TfEn-RFE (proposed)	Normal	100.00%	0	50
	Inner-race fault	100.00%	0	50
	Outer-race fault	100.00%	0	50
	Ball fault	100.00%	0	50
	Impeller fault	100.00%	0	50
	Average/Total	100.00%	0	250

 Table 2. Results of proposed fault diagnosis.

As shown in Table 2, defining the size of training samples as 50 and using the SVM as the classifier model, the proposed method based on TfEn and RFE brings about the best results among all 4 methods, where its average accuracy is 100.00%, and no testing sample is misclassified.

Compared with total raw TfEn features, our proposed method improves the average accuracy by 1.20%. Compared with TfEn-PCA and TfEn-KPCA, our proposed method improves the average accuracy by 11.20% and 17.60%. The results preliminarily demonstrate the validity of TfEn, and the superiority of RFE over the mainstream PCA and KPCA-based feature selection methods.

4.3. Comparisons with Variable Taining Sample Sizes

In this subsection, two sets of representative methods are applied for comparisons, namely the entropy-based methods and the deep neural network (DNN)-based methods. In addition, in order to further test the adaptability of these methods to extremely few failure samples, a variety of training sample sizes are designed and applied for fault diagnosis

Category I entropy-based methods

In this category, still applying CEEMD for the signal processing, 4 existing mainstream entropy-based feature extraction methods are utilized to implement the comparisons, which consist of the singular value decomposition entropy (SvdEn) [31], sample entropy (SampleEn) [32], approximately entropy (ApEn) [31] and spectral entropy (SpecEn) [33].

Additionally, in the feature selection stage, the PCA and KPCA are still utilized to compare with the RFE, which means that there are 3 classes of representations for each feature extraction method applied to achieve the fault diagnosis, namely, features selected by PCA, features selected by KPCA and features selected by RFE. In the fault classification stage, as in Section 4.2, there are still 10 samples under each mode applied to train an SVM classifier model. The corresponding results are shown in Figure 11.



Figure 11. Comparisons with 4 mainstream entropy-based feature extraction methods.

As shown in Figure 11, under the condition of 10 training samples per fault mode, the best testing accuracy (100%) is achieved by the proposed method mainly based on TfEn and RFE. For each entropy

method, the RFE-based feature selection outperforms another two based on PCA and KPCA, and the advantages are quite obvious, especially for SvdEn and SpectEn, which illustrates the superiority of the RFE-based feature selection. In addition, using the same feature selection based on RFE, the proposed TfEn obtains a better classification accuracy that the other 4 existing entropy methods, where the accuracy is respectively improved by 12.80%, 5.20%, 5.60% and 6.40%. It further proves that our proposed TfEn can extract the more intrinsic fault representations than the other 4 mainstream entropy methods aimed at processing the 1D sequence, which is thus more suitable for the fault classification especially with few failure samples.

In addition to the performance on feature extraction, we compare the calculation efficiency among these entropy methods by the quantified running time, which is illustrated in Table 3.

Method	SvdEn	SampleEn	ApEn	SpectEn	TfEn(Proposed)
Running time(s) per 50 samples	0.24	2862.31	2740.10	0.14	1.61

As shown in Table 3, due to the state space reconstruction and lots of loops, SampleEn and ApEn take 2862.31 s and 2740.10 s, respectively, to complete the feature extraction for each 50 samples. Although these two entropy methods obtain a slightly higher classification accuracy than SvdEn and SpectEn, the overly low calculation efficiency still makes them almost intolerable in practices.

To sum up, the proposed TfEn achieves the best balance between the performance and computation efficiency. Although its running time (1.61 s per 50 samples) is slightly longer than SvdEn (0.24 s per 50 samples) and SpectEn (0.14 s per 50 samples), it is still acceptable, taking into account its considerable performance on fault classification.

• Category II DNN-based methods

In this category, 3 representative DNN models, namely, multilayer perceptron (MLP) [34], stacked autoencoder (SAE) [35] and convolutional neural network (CNN) [36], are applied to compare with our proposed method. Referring to the related literatures [37,38], the preprocessing and model architectures are respectively set as follows.

Besides that, in order to further test the adaptability of methods to limiting few failure samples, the training sample size under each fault mode is respectively set as 4, 6, 8, 10, 12 and 14 in this subsection. In addition to the above 3 deep models, the 5 entropy-based methods in Category I are also applied to the comparisons, which are respectively SvdEn-RFE, SampleEn-RFE, ApEn-RFE, SpecEn-RFE and our proposed TfEn-RFE. The corresponding results are illustrated in Figure 12 and Table 4.

Method	Preprocessing	Type of Layer Hidden Size		Layer Num
MLP	Fast Fourier transform (FFT)	Fully connection layer 512-128; 128-64; 64-5		3
		Encoder layer	512-128; 128-64	2
SAE	FFT	Decoder layer 64-128; 128-512		2
		Fully connection layer	64-16; 16-5	2
CNN	STFT	Conv layer	1-2-5; 2-3-5; 3-3-5	3
		Max pool layer	2-2	1
		Fully connection layer	1980-32; 32-5	2
Dropout	oout Optimizer Training Epochs		Learning Rate	Batch Size
0.5	Adam	100	0.02	15

Table 4. The details of hyper-parameters and model architectures in the deep neural network (DNN) methods.



Figure 12. Comparisons among entropy and DNN-based methods with various training samples.

According to the comparative results, we can obtain the following conclusions of the proposed TfEn-RFE method.

(1) Stability for variable sizes of training samples

As shown in Figure 12 and Table 5, with variable sizes of training samples (4/6/8/10/12/14):

For the three DNN methods (FFT-MLP, FFT-SAE and STFT-CNN), the ranges between the maximum accuracy and minimum accuracy are 20.37%, 39.29% and 30.00%, respectively. The DNN-based methods do have extraordinary performance when the training samples are relatively sufficient, with the help of their powerful capability in feature mining. However, their accuracy presents the obvious degradation with the extremely few training samples (size of 4 per fault mode), which results by their worst stability for variable training sizes among all the methods.

Training Size/Methods	4	6	8	10	12	14
SvdEn-RFE	78.93%	87.69%	82.31%	87.20%	83.78%	93.04%
SampleEn-RFE	91.43%	92.22%	92.69%	94.80%	93.75%	95.22%
ApEn-RFE	93.57%	93.71%	93.08%	94.40%	96.25%	96.52%
SpectEn-RFE	81.43%	94.07%	93.85%	93.60%	97.92%	96.09%
FFT-MLP	70.36%	70.00%	74.40%	75.00%	80.00%	90.37%
FFT-SAE	60.71%	80.00%	80.00%	85.20%	100.00%	100.00%
STFT-CNN	70.00%	74.82%	80.00%	100.00%	100.00%	100.00%
TfEn-RFE (proposed)	99.29%	99.26%	100.00%	100.00%	100.00%	100.00%

Table 5. Detailed classification accuracy of each method with various training samples.

For the four existing entropy methods (SvdEn-RFE, SampleEn-RFE, ApEn-RFE and SpectEn-RFE), their accuracy ranges are 14.11%, 3.79%, 3.44% and 16.49%, respectively. These entropy-based methods achieve the better stability than the DNN-based ones. However, their performance does not achieve significant improvements like the DNN-based ones with increasing training sizes, because they can

only extract the features from a single domain (time or frequency), and thus lose useful information of those non-stationary Imfs.

For the proposed TfEn-RFE method, its accuracy range is only 0.74%. To sum up, the proposed TfEn-RFE method always maintains the best classification accuracy, and shows no obvious degradations, even with extremely few training samples, which exhibits the best stability for the variable training samples among all the methods.

(2) Significant improvement on classification accuracy with extremely few training samples

As shown in Figure 12 and Table 5, under the sceneries of extremely few training samples (size of 4 per fault mode), the proposed TfEn improve the average classification accuracy by 33.27% compared with three mainstream DNN-based methods, and the average classification accuracy is improved by 12.95% compared with four mainstream entropy-based methods. Among all the methods, the DNN-based methods obtain the lowest average accuracy. This is because there are many parameters that need to be learned and updated in their complicated model architectures, and the scarce training samples are not enough to support this, and thus limit their performance. Compared to these DNN-based methods, the existing entropy-based methods present the relatively better adaptiveness to the scarce failure samples, while there is still room for improved performance due to the information loss caused by the ignored domain of time or frequency.

To sum up, through jointly taking into account of the nonlinear behaviors from both time and frequency domains, the proposed TfEn-RFE method can extract more intrinsic fault representation and thus achieve the best adaptiveness for the extremely few failure samples. With the help of the proposed method, the classification accuracy is significantly improved compared with the other comparative methods.

4.4. Discussion

Considering all the above-mentioned experimental results, the five main conclusions are listed as follows:

1. TfEn can extract more intrinsic fault features of non-stationary series from both time and frequency domains.

As shown in Figures 11 and 12, TfEn achieves the higher classification accuracy than all the other existing entropy methods. Through jointly taking into account nonlinear behaviors from both dimensions of time and frequency, TfEn can extract more intrinsic fault features, even if the decomposed components are non-stationary or approximately stationary (shown in Figures 8 and 9), which avoids the risk of losing useful information caused by existing entropy methods aimed at processing the 1D sequence.

2. TfEn achieves the good balance between the performance and calculation efficiency.

As shown in Table 3, compared with the existing SampleEn (2862.31 s) and ApEn (2740.10 s), t TfEn presents the significantly shorter running time (1.61 s) but the higher classification, since it does not rely on the state space reconstruction and lots of loop comparisons that need considerable computation resources. Additionally, compared with another two entropy methods, SvdEn (0.24 s) and SpectEn (0.14 s), the running time of TfEn is slightly longer (1.61 s), while it brings about the considerable improvement on classification accuracy with variable training sizes (shown in Table 5). The relatively better balance between performance and calculation efficiency makes TfEn more suitable for practices.

3. RFE exhibits superior performance and interpretability than the mainstream PCA and KPCA.

As shown in Figure 10 and Table 2, using the RFE, the difference of TfEn features among each fault mode is further highlighted, and its average classification accuracy is improved by 11.20% and 17.60%, compared with mainstream PCA and KPCA. For the other four entropy methods, the RFE also brings about the better results than the PCA and KPCA-based ones (shown in Figure 11 and Table 5), which

reveals its superiority on the feature selection. Besides that, the optimal features selected by the RFE retain their original physical meanings and thus having the better interpretability (shown in Table 1).

4. Effective feature extraction designed manually is still necessary under the scenery of few failure samples.

As described in Section 4.3, various sizes of training samples (4/6/8/10/12/14) are defined to test the stability and adaptiveness for methods. According to Figure 12 and Table 5, we can see that, with the help of strong capability on the feature self-learning, the DNN-based methods do achieve superior performance (accuracy of 100%) when training sizes are relatively large (size of 12/14 per fault mode). However, their classification accuracy is overtaken by the entropy-based methods designed manually when the training samples are gradually decreased, which means that the effective manual feature extraction is still necessary for the practices where the failure data is usually hard to collect.

5. The proposed method provides a novel concept for fault diagnosis subject to the limitation of few failure samples.

With the help of TfEn and RFE, robust fault representations are constructed for centrifugal pumps, which presents the superior stability and adaptiveness for the variable sizes of training samples. When the training samples are extremely scarce (size of four per fault mode), the average classification accuracy is significantly improved by 12.95% and 33.27%, compared with the mainstream entropy-based methods and the DNN-based ones. Thanks to the remarkable fault representation, the proposed method has no requirements for massive data, and is thus more suitable for practical engineering.

5. Conclusions

In this paper, we proposed a TfEn and RFE based diagnosis method for centrifugal pumps, subject to the limitation of failure samples. Through jointly taking into account the complexity and dynamic change from both the domains of time and frequency, TfEn can highlight and extract more intrinsic fault features of non-stationary vibration signals for centrifugal pumps compared with the mainstream entropy methods. Using the RFE-based feature selection, the optimal fault features are automatically selected, and the diagnostic performance is further improved, and it should be noted that the selected features still retain their original physical meanings, and thus have better interpretability than the PCA. The experiments show that our proposed two-stage fault representation method presents stability against various training sizes, and superior performance with extremely few failure samples, which exhibits remarkable potential for further applications in various aspects of practical engineering.

In the future, we will further study the fault diagnosis method, considering the fusion of multi-type sensor signals, which may be more suitable for real industrial scenarios.

Author Contributions: Conceptualization, X.S.; Data curation, X.S.; Funding acquisition, L.T.; Methodology, X.S., H.L. and L.T.; Writing original draft, X.S.; Writing review and editing, L.T. All authors have read and agreed to the published version of the manuscript.

Funding: This study is funded by Key-Area Research and Development Program of Guangdong Province (Grant No.2018B010108005), the National Natural Science Foundation of China (Grant Nos. 61973011 and 61803013), the Fundamental Research Funds for the Central Universities (Grant No. YWF-20-BJ-J-723), National key Laboratory of Science and Technology on Reliability and Environmental Engineering (Grant No. 6142004180501), the Research Fund (Grant No. 61400020401), the Capital Science & Technology Leading Talent Program (Grant No. Z191100006119029).

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

References

- Irfan, M.; Alwadie, A.; Glowacz, A. Design of a Novel Electric Diagnostic Technique for Fault Analysis of Centrifugal Pumps. *Appl. Sci.* 2019, *9*, 5093. [CrossRef]
- 2. Azadeh, A.; Saberi, M.; Kazem, A.; Ebrahimipour, V.; Nourmohammadzadeh, A.; Saberi, Z. A flexible algorithm for fault diagnosis in a centrifugal pump with corrupted data and noise based on ANN and support vector machine with hyper-parameters optimization. *Appl. Soft Comput.* **2013**, *13*, 1478–1485. [CrossRef]

- 3. Cao, P.; Zhang, S.; Tang, J. Preprocessing-Free Gear Fault Diagnosis Using Small Datasets with Deep Convolutional Neural Network-Based Transfer Learning. *IEEE Access* **2018**, *6*, 26241–26253. [CrossRef]
- 4. Pathirage, C.S.N.; Li, J.; Li, L.; Hao, H.; Liu, W.; Ni, P. Structural damage identification based on autoencoder neural networks and deep learning. *Eng. Struct.* **2018**, *172*, 13–28. [CrossRef]
- Jia, F.; Lei, Y.; Lin, J.; Zhou, X.; Lu, N. Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. *Mech. Syst. Signal Process.* 2016, 72, 303–315. [CrossRef]
- Liu, R.; Yang, B.; Zio, E.; Chen, X. Artificial intelligence for fault diagnosis of rotating machinery: A review. *Mech. Syst. Signal Process.* 2018, 108, 33–47. [CrossRef]
- 7. Feng, Z.; Liang, M.; Chu, F. Recent advances in time–frequency analysis methods for machinery fault diagnosis: A review with application examples. *Mech. Syst. Signal Process.* **2013**, *38*, 165–205. [CrossRef]
- 8. Xu, Y.; Sun, Y.; Wan, J.; Liu, X.; Song, Z. Industrial Big Data for Fault Diagnosis: Taxonomy, Review, and Applications. *IEEE Access* 2017, *5*, 17368–17380. [CrossRef]
- 9. Hui, K.H.; Ooi, C.S.; Lim, M.H.; Leong, M.S.; Al-Obaidi, S.M. An improved wrapper-based feature selection method for machinery fault diagnosis. *PLoS ONE* **2017**, *12*, e0189143. [CrossRef]
- 10. Li, Y.; Wang, X.; Liu, Z.; Liang, X.; Si, S. The Entropy Algorithm and Its Variants in the Fault Diagnosis of Rotating Machinery: A Review. *IEEE Access* **2018**, *6*, 66723–66741. [CrossRef]
- 11. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.-C.; Tung, C.C.; Liu, H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **1998**, *454*, 903–995. [CrossRef]
- 12. Wu, Z.; Huang, N.E. ENSEMBLE EMPIRICAL MODE DECOMPOSITION: A NOISE-ASSISTED DATA ANALYSIS METHOD. *Adv. Adapt. Data Anal.* **2009**, *1*, 1–41. [CrossRef]
- 13. Li, G.; Deng, C.; Wu, J.; Chen, Z.; Xu, X. Rolling Bearing Fault Diagnosis Based on Wavelet Packet Transform and Convolutional Neural Network. *Appl. Sci.* **2020**, *10*, 770. [CrossRef]
- 14. Mu, Z.; Hu, J.; Min, J.-L. Driver Fatigue Detection System Using Electroencephalography Signals Based on Combined Entropy Features. *Appl. Sci.* **2017**, *7*, 150. [CrossRef]
- 15. An, X.; Pan, L. Wind turbine bearing fault diagnosis based on adaptive local iterative filtering and approximate entropy. *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.* **2016**, *231*, 3228–3237. [CrossRef]
- 16. Kumar, M.; Pachori, R.B.; Acharya, U.R. Automated Diagnosis of Myocardial Infarction ECG Signals Using Sample Entropy in Flexible Analytic Wavelet Transform Framework. *Entropy* **2017**, *19*, 488. [CrossRef]
- 17. Lechner, M.; Miquel, R.; Wunsch, C. The Curse and Blessing of Training the Unemployed in a Changing Economy: The Case of East Germany After Unification. *Ger. Econ. Rev.* **2007**, *8*, 468–509. [CrossRef]
- Sun, W.; Chen, J.; Li, J. Decision tree and PCA-based fault diagnosis of rotating machinery. *Mech. Syst. Signal Process.* 2007, 21, 1300–1317. [CrossRef]
- Lipu, M.S.H.; Hannan, M.A.; Hussain, A.; Saad, M.H.M. Optimal BP neural network algorithm for state of charge estimation of lithium-ion battery using PSO with PCA feature selection. *J. Renew. Sustain. Energy* 2017, 9, 064102. [CrossRef]
- 20. Bolón-Canedo, V.; Sánchez-Maroño, N.; Alonso-Betanzos, A. Recent advances and emerging challenges of feature selection in the context of big data. *Knowl. Based Syst.* **2015**, *86*, 33–45. [CrossRef]
- 21. Yan, K.; Zhang, L. Feature selection and analysis on correlated gas sensor data with recursive feature elimination. *Sens. Actuators B Chem.* **2015**, *212*, 353–363. [CrossRef]
- 22. Yeh, J.-R.; Shieh, J.-S.; Huang, N.E. complementary ensemble empirical mode decomposition: A novel noise enhanced data analysis method. *Adv. Adapt. Data Anal.* **2010**, *2*, 135–156. [CrossRef]
- 23. Yang, J.; Li, P.; Yang, Y.; Xu, D. An improved EMD method for modal identification and a combined static-dynamic method for damage detection. *J. Sound Vib.* **2018**, *420*, 242–260. [CrossRef]
- 24. Deng, W.; Zhang, S.; Zhao, H.; Yang, X. A Novel Fault Diagnosis Method Based on Integrating Empirical Wavelet Transform and Fuzzy Entropy for Motor Bearing. *IEEE Access* **2018**, *6*, 35042–35056. [CrossRef]
- 25. Kwok, H.; Jones, D. Improved instantaneous frequency estimation using an adaptive short-time Fourier transform. *IEEE Trans. Signal Process.* **2000**, *48*, 2964–2972. [CrossRef]
- Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V. Gene Selection for Cancer Classification using Support Vector Machines. *Mach. Learn.* 2002, 46, 389–422. [CrossRef]
- 27. Wang, C.; Xiao, Z.; Wu, J. Functional connectivity-based classification of autism and control using SVM-RFECV on rs-fMRI data. *Phys. Medica* 2019, *65*, 99–105. [CrossRef] [PubMed]

- Tipping, M.E.; Bishop, C.M. Probabilistic Principal Component Analysis. J. R. Stat. Soc. Ser. B (Stat. Methodol. 1999, 61, 611–622. [CrossRef]
- 29. Schölkopf, B.; Smola, A.; Müller, K.-R. Kernel principal component analysis. In *International Conference on Artificial Neural Networks*; Springer: Berlin/Heidelberg, Germany, 1997; pp. 583–588.
- Lameski, P.; Zdravevski, E.; Mingov, R.; Kulakov, A. SVM Parameter Tuning with Grid Search and Its Impact on Reduction of Model Over-fitting. In *Lecture Notes in Computer Science*; Springer Science and Business Media LLC: Berlin, Germany, 2015; Volume 9437, pp. 464–474.
- 31. Yentes, J.M.; Hunt, N.; Schmid, K.K.; Kaipust, J.P.; McGrath, D.; Stergiou, N. The appropriate use of approximate entropy and sample entropy with short data sets. *Ann. Biomed. Eng.* **2012**, *41*, 349–365. [CrossRef]
- 32. Costa, M.; Goldberger, A.L.; Peng, C.-K. Multiscale entropy analysis of biological signals. *Phys. Rev. E* 2005, 71, 021906. [CrossRef]
- 33. Su, H.; Shi, T.; Chen, F.; Huang, S. New method of fault diagnosis of rotating machinery based on distance of information entropy. *Front. Mech. Eng.* **2011**, *6*, 249. [CrossRef]
- 34. Gardner, M.; Dorling, S. Artificial neural networks (the multilayer perceptron)—A review of applications in the atmospheric sciences. *Atmospheric Environ.* **1998**, *32*, 2627–2636. [CrossRef]
- 35. Lu, C.; Wang, Z.-Y.; Qin, W.-L.; Ma, J. Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification. *Signal Process.* **2017**, *130*, 377–388. [CrossRef]
- 36. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Pdf ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
- 37. Chen, Z.; Deng, S.; Chen, X.-D.; Li, C.; Sánchez, R.-V.; Qin, H. Deep neural networks-based rolling bearing fault diagnosis. *Microelectron. Reliab.* **2017**, *75*, 327–333. [CrossRef]
- Verstraete, D.; Ferrada, A.; Droguett, E.L.; Meruane, V.; Modarres, M. Deep Learning Enabled Fault Diagnosis Using Time-Frequency Image Analysis of Rolling Element Bearings. *Shock. Vib.* 2017, 2017, 1–17. [CrossRef]



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