Article

Bearing Fault Diagnosis Based on a Hybrid Classifier Ensemble Approach and the Improved Dempster-Shafer Theory

Yanxue Wang [1],* [ID], Fang Liu [1,2] and Aihua Zhu [1]

[1] Beijing Key Laboratory of Performance Guarantee on Urban Rail Transit Vehicles, Beijing University of Civil Engineering and Architecture, Beijing 100044, China; liudhuanglf@gmail.com (F.L.); zhuaihua@bucea.edu.cn (A.Z.)

[2] School of Mechanical and Electrical Engineering, Guilin University of Electronic Technology, Guilin 541004, China

* Correspondence: Yan.xue.wang@gmail.com; Tel.: +86-010-92602515; Fax: +86 010 92602515

check for updates

Abstract: Bearing fault diagnosis of a rotating machine plays an important role in reliable operation. A novel intelligent fault diagnosis method for roller bearings has been developed based on a proposed hybrid classifier ensemble approach and the improved Dempster-Shafer theory. The improved Dempster-Shafer theory well considered the combination of unreliable evidence sources, the uncertainty information of basic probability assignment, and the relative credibility of the evidence on the weights in the process of decision making under the framework of fuzzy preference relations, which can effectively deal with conflicts of the evidences and then well improve the diagnostic accuracy for the hybrid classifier ensemble. The effectiveness of the improved Dempster-Shafer theory has been verified via a numerical example. In addition, deep neural networks, a support vector machine, and extreme learning machine techniques have been utilized in the single-stage classification based on singular spectrum entropy, power spectrum entropy, time-frequency entropy, and wavelet packet energy spectrum entropy in this work. Performances of the proposed hybrid ensemble classifier has been demonstrated on a bearing test-rig, compared with the original Dempster-Shafer theory. It can be found that the overall error rate can be greatly reduced with the hybrid ensemble classifier and the improved Dempster-Shafer theory.

Keywords: rolling element bearing; hybrid classifier ensemble; Dempster-Shafer evidence theory; fuzzy preference relations

1. Introduction

Rolling element bearings are the key components widely used in rotating machines. A sudden breakdown of the mechanical system or even a severe catastrophe, may be caused due to an unexpected failure of the rolling element bearings. Therefore, many bearing fault diagnosis methods have been developed based on vibration signal analysis and feature extraction [1–3]. However, some of them are performed manually with low efficiency by means of knowledge and experiences of experts, which are not practical in real applications. Thus, there is still growing attention towards the development of bearing intelligent fault diagnosis techniques. For example, a novel intelligent fault diagnosis method has been proposed based on the affinity propagation clustering algorithm and the adaptive feature selection technique [4]. Qin et al. [5] proposed a model for fault diagnosis of gearboxes in wind turbines based on deep belief networks (DBNs), using improved logistic sigmoid units and the impulsive signatures. In addition, a three-stage intelligent fault diagnosis clustering technique has been proposed for the industrial process monitoring [6]. Generally, the diagnosis results achieved by

using a single-stage classifier may still be precarious [7–10]. According to Wolpert's theorem, there is not a single classifier approach that can be successfully applied for all pattern recognition tasks since each has its own domain of competence [11].

Nowadays, many different combinations of several different learning algorithms, such as the hybrid or ensemble systems, have been highlighted as a hot topic and promising trend in the fields of pattern recognition. The hybrid intelligent systems offer many alternatives for unorthodox handling of realistic increasingly complex problems, involving ambiguity, uncertainty, and high-dimensionality of data [12]. Nevertheless, the accuracy of the existing techniques needs to be further improved, since the structure of rotating machinery becomes increasingly complicated. Therefore, a novel hybrid classifier ensemble (HCE) algorithm has been developed in this work, which can perform fault diagnosis under an improved framework of information fusion.

Actually, there are various strategies for information fusion, such as the simple voting procedure [13]. The Dempster-Shafer theory (DST) has been also widely used as a combining decision method due to its uncertainty processing ability [14]. In recent years, DST has attracted lots of attention and has been used in fault diagnosis for different industrial equipment. For example, a fusion approach was proposed for fault diagnosis of roller bearing in the aeroengine based on n-dimensional characteristic parameter distance [15]. Since a hybrid technique can substantially increase the accuracy of fault detection, DST combined with Support Vector Machine (SVM) has been applied for bearing multi-fault diagnosis [16]. A fault diagnosis method was proposed for the reactor coolant system of a nuclear power plant based on DST and fuzzy function in reference [17]. DST is well suitable for information fusion, but it may generate counter-intuitive results for highly conflicting and unreliable pieces of evidence [18,19]. Thus, conflict management has always been an unavoidable problem in information fusion using DST, which is also the main limitation of DST. To solve this issue, many improved versions of DST have been proposed, such as the average approach in reference [20], the weighted average based on the evidence distance in reference [21], and the vector space introduced in reference [22]. Most of the available methods employed distance of the evidences as a critical factor to determine the weights, such as the Jousselme distance [23] and the MaxDiff distance [24]. Then, the support degrees of the evidences can be adjusted and be used to generate the appropriate weights with regard to the evidences. It can be found that a bigger weight is set to the reliable evidence and a smaller weight is set to the unreliable evidence. Although these techniques can reduce the influence of the unreliable evidence, they rarely consider the effects of the uncertain information of the evidences.

Many fuzzy modeling approaches have been successfully utilized in various applications in the past decades, since fuzzy sets technique also plays an important role in the decision-making process and can deal well with uncertain information. Qian etc. [25] successfully utilized the advantages of group decision making via fuzzy preference. The fuzzy preference relations (FPR) has been constructed for multiple pieces of evidence based on the variance of information entropy. However, according to reference [23], there are three drawbacks of this approach. (i) It does not satisfy the property of the additive consistency and the order consistency; (ii) It cannot calculate the preference values in some situations; (iii) The preference values in the consistency matrix are not always between zero and one. Therefore, a new improved DST approach is proposed in this paper inspired by reference [26], which well considers the combination of unreliable evidence in the group decision making under the framework of FPR.

Two major contributions have been made in this work. First, a new hybrid classifier ensemble (HCE) method is proposed based on entropy features to improve the performance and accuracy of fault diagnosis. Second, an improved DST has been proposed to perform information fusion of classification decisions obtained by HCE, which considers the combination of unreliable and conflictive evidence sources, the uncertainty information of basic probability assignment (BPA) and the relative credibility of the evidence on the weights under the framework of FPR. The novel HCE model combined with the improved DST technique has been utilized to automatically identify bearing faults in a rotating machine. Results have demonstrated well the effectiveness of the proposed method.

This work is organized as follows. Theories of entropy feature extraction and single-stage classifier have been briefly reviewed in Section 2. The improved DST for dealing with conflicting evidence has been given in Section 3, where the performance of the proposed approaches has also been demonstrated using two examples. The HCE approach combined with the improved DST is adopted to identify bearing fault automatically, whose effectiveness was demonstrated on a test-rig in Section 4. Conclusions are drawn in Section 5.

2. Methodologies

The techniques of entropy feature extraction and the classifiers mentioned in HCE have been briefly introduced in this section.

2.1. Entropy Feature Extraction

Feature extraction is crucial in pattern recognition and mechanical fault diagnosis. However, traditional signal processing methods, like Fourier transform, are not suitable for analyzing the non-linear and non-stationary bearing vibration signals. It seems that time-frequency analysis techniques are much more suitable for extracting bearing fault features. Several advanced time-frequency signal processing techniques have been adopted in feature extraction. For example, variational mode decomposition (VMD) [27] is as a self-adaptive decomposition method lately proposed with a solid theory [28].

Moreover, traditional statistical properties and frequency-domain signatures cannot meet the requirements because of the non-linear and non-stationary characteristics of the decomposed components [29]. Many non-linear parameter estimation methods have been proved to get the feature information, such as entropy theory introduced in reference [30] to estimate the complexity and stationarity of the signal. Entropy features can be also applied to quantify the malfunction and reflect the uncertainty of vibration signals. In addition, different entropy features obtained in different domains can be used to fully describe a vibration signal. Thus, singular spectrum entropy (SSE) [31], power spectrum entropy (PSE) [32], time-frequency entropy (TFE) [33], and wavelet packet energy spectrum entropy (WPESE) [34] have been used to calculate the feature sets in this work, which are associated with singular spectrum in time domain, power spectrum in frequency domain, time-frequency spectrum, and wavelet packet energy spectrum in time-frequency domain, respectively. These four entropy features will be indicated as follows.

2.1.1. Singular Spectrum Entropy

SSE indicates the uncertainty degree of the signal energy divided by singular spectrum analysis, which can effectively represent the signal energy change in the time domain [31]. Based on the delay embedding technique, an arbitrary signal $\{x_i\}(i = 1, 2, \ldots, N)$ was mapped to an embedded space represented by the $M \times N$ matrix U, i.e., As explained in reference [31], the calculation of U is shown as

$$U = \begin{bmatrix} x_1 & x_2 & \ldots & x_M \\ x_2 & x_3 & \ldots & x_{M+1} \\ \vdots & \vdots & \ldots & \vdots \\ x_{N-M} & x_{N-M+1} & \ldots & x_N \end{bmatrix} \tag{1}$$

where M is the length of the embedded space, N is the number of samples. The singular values $\{\lambda_i\}$ of the matrix U are achieved based on the singular value decomposition (SVD). Thus, the SSE of the signal via information entropy theory is defined as

$$H_S = -\sum_{i=1}^{M} p_i \log p_i \tag{2}$$

in which

$$p_i = \lambda_i / (\sum_{i=1}^{M} \lambda_i) \tag{3}$$

and p_i is the ratio of the ith singular spectrum to the whole spectrum.

2.1.2. Power Spectrum Entropy

PSE can reflect the complexity and stability of a signal, which is also used to indicate the distribution of signal energy in frequency domain [32]. The proportional distribution of different frequencies is defined as a probability distribution. When $X(\omega)$ is obtained by using the discrete Fourier transform for a signal $\{x_t\}$, as explained in reference [32], the calculation of the power spectrum is shown as

$$S(\omega) = \frac{1}{N}|X_i(\omega)|^2. \tag{4}$$

where $S = \{S_1, S_2, \dots, S_N\}$ can be regarded as the partition of a signal $\{x_t\}$. Hence the PSE can be defined as follows:

$$H_P = -\sum_{i=1}^{N} q_i \log q_i. \tag{5}$$

where $q_i = S_i / (\sum_{i=1}^{N} S_i)$, and q_i is the ratio of the ith power spectrum to the whole spectrum.

2.1.3. Time-Frequency Entropy

TFE is used to quantitatively measure the time-frequency representation [33]. Let a time-frequency plot have L equal blocks, where the information source for the entire plane is η and for each block is $\gamma_i (i = 1, 2, \dots, L)$. As explained in reference [33], the calculation of the time-frequency entropy is shown as

$$H_T = -\sum_{i=1}^{N} \delta_i \log \delta_i. \tag{6}$$

where $\delta_i = \gamma_i / \eta$, δ_i the ratio of the i-th energy to the whole energy.

2.1.4. Wavelet Packet Energy Spectrum Entropy

A sequence $\left\{ J_k^j, \ k = 0, 1, 2, \dots, 2^j - 1 \right\}$ represents the decomposition result using j-layer wavelet package transform. The sum of squares of signals in each frequency band after wavelet packet transform (WPT) is selected as wavelet packet energy. As explained in reference [34], the calculation of energy value corresponding to the i-th band is given below

$$E_i = \sum_{k=1}^{2^j} |W_i(k)|^2. \tag{7}$$

where $W_i(k)$ is the reconstructed coefficients for each node. Thus, WPESE can be defined by

$$H_W = -\sum_{i=1}^{2^j} r_i \log r_i. \tag{8}$$

2.2. Classification Models

The difference between classifiers in HCE should be increased to enhance the complementarity between classification methods, which can comprehensively describe the diagnostic object. Three supervised classification models are selected, that is, the traditional Deep Neural Networks

(DNN), the shallow learning algorithm Support Vector Machine (SVM), and Extreme Learning Machine (ELM).

DNN is one of the most widely used intelligent methods in pattern recognition, fault diagnosis and classification. DNN is a kind of deep learning technique, which is comprised of unsupervised layer-by-layer greedy training and global parameter tuning using the back propagation algorithm. DNN can not only solve complex nonlinear problems but also extract features in a high-dimensional space. Presently, many different models of DNN have been developed. For example, a DNN-based model was used to identify the fault condition of roller bearing [35]. The Deep Boltzmann machine combined with multi-grained scanning forest ensemble was developed for the fault diagnosis of industrial big data [7]. Thus, DNN will be adopted as single-stage classifier in HCE in this work.

SVM is a well-known shallow learning method in classification and regression applications. SVM has good generalization capability for classification of a small sample [36], which have been widely used in fault diagnosis and prognostics. To improve the performance of SVM, PSO is adopted to optimize the parameters in SVM.

ELM is considered as a single hidden layer feed forward neural networks [37,38]. The input weights are set randomly, then the network is expressed as a linear system, and the output weights can be calculated analytically [38]. The weight between the hidden layer and the output layer of ELM does not need to be adjusted iteratively, which is obtained by generalized inverse of a matrix. The performance of ELM depends on the randomly input weights and thresholds. In this work, the fruit fly optimization algorithm (FOA) is used to improve the performance of traditional ELM. Both SVM and ELM are utilized in HCE in this work.

2.3. Dempster-Shafer Theory

DST is one of the most powerful tools for the ensemble of multiple classifiers system, which can deal with incomplete, uncertain, and unclear information in the multi-sensor information fusion [39]. DST was initially developed by Shafer in 1976. Assume $\Theta = \{D_1, D_2, \ldots, D_n\}$ is a set of mutually exclusive and collectively exhaustive events, which is called the frame of discernment (FOD). A basic probability assignment (BPA) is a map of m from 2^Θ to [0, 1], as explained in reference [40], the calculation of the BPA function is shown below,

$$
\begin{cases}
\sum\limits_{A \subset 2^\Theta} m(A) = 1 \\
m(\varnothing) = 0
\end{cases}
. \tag{9}
$$

Based on the belief function theory, two independent BPAs can be combined by Dempster's rule, denoted as $m = m_1 \oplus m_2$, which is defined as follows.

$$
m(A) =
\begin{cases}
\frac{1}{1-K} \sum\limits_{B \cap C = A} m_1(B)m_2(C), & A \neq \varnothing \\
0, & A = \varnothing
\end{cases}
\tag{10}
$$

where $K = \sum_{B \cap C = \varnothing} m_1(B)m_2(C)$. The conflict coefficient K is used to measure the conflict between two pieces of evidences. The larger the value of K is, the larger conflict between evidences gets.

It should be noted that there may exist conflict between the evidence in the fusion of HCE. To solve this issue, a new weighted average approach is proposed, which considers not only the support degree between the pieces of evidence but also the uncertainty information of BPA. This improved version of DST is given in the following subsection.

3. The Improved Dempster-Shafer Theory Approach

It is crucial to detect the relatively reliable evidence in the process of information fusion. In the multiple classifier systems, the conflict problem caused by the result of the classifier cannot be ignored. Thus, an improved DST approach is developed in this work and will be introduced in

detail subsequently. First, since cosine similarity reflects the confidence degree of the evidence itself, the cosine similarity is employed to indicate the support degree between the pieces of evidence. In addition, DST can be considered as a generalized probability theory, entropy can be used to measure the quantitative uncertainty in BPA. Therefore, entropy based on FPR is applied to indicate the relative reliability preference between the bodies of evidence (BOE). Considering the above two aspects, it can be found that the improved DST will be much more reasonable in dealing with conflicts compared with the original DST. The proposed technique includes three parts: The measurement of the degree of support between evidence using the cosine similarity, the calculation of the weight of BPA, and the improved fusion for BPAs, as shown in Figure 1.
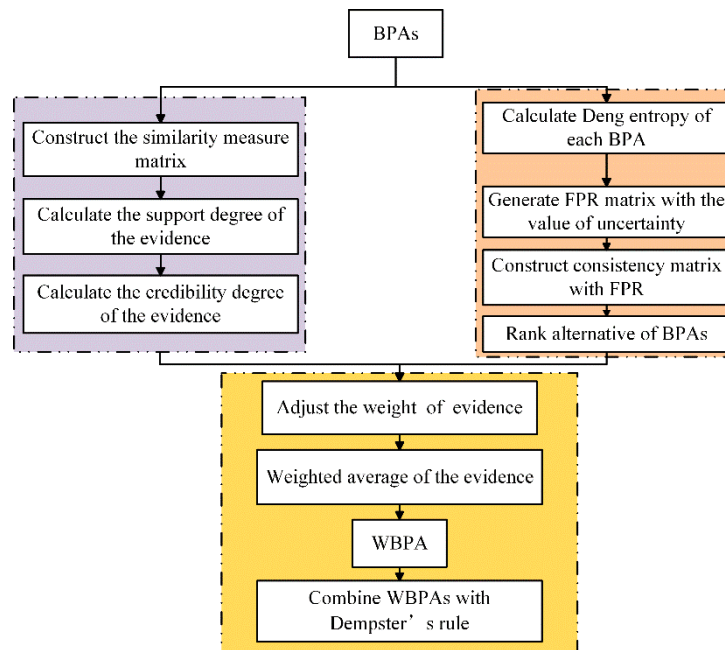

Figure 1. The flowchart of the proposed Dempster-Shafer theory (DST).

3.1. The Cosine Similarity

The cosine similarity is used to measure the confidence degree of evidence [41]. Let Θ be a frame of discernment and $\Theta = \{\theta_1, \theta_2, \ldots, \theta_n\}$. Employ the cosine similarity function, as explained in reference [41], the calculation of similarity degree between evidence m_i, m_j is given below,

$$S_{ij} = \frac{m_i \cdot m_j^T}{\| m_i \| \cdot \| m_j \|}.$$

(11)

where $m_i \cdot m_j$ is inner product of m_i and m_j. And $\| \cdot \|$ represents the norm of vector. For the n-sources fusion system, the similarity measure matrix is defined as follow.

$$S = \begin{bmatrix} 1 & \cdots & S_{1i} & \cdots & S_{1k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ S_{i1} & \cdots & 1 & \cdots & S_{ik} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ S_{k1} & \cdots & S_{ki} & \cdots & 1 \end{bmatrix}.$$

(12)

The Support degree of the evidence m_i can be defined as follows.

$$sup(m_i) = \sum_{j=1}^{n} S_{ij}. \tag{13}$$

Thus, the credibility degree of the evidence m_i is denoted below.

$$crd_i = \frac{sup(m_i)}{\max(sup(m_i))}. \tag{14}$$

3.2. The Uncertainty Measurement of the Weights

Deng entropy [42], which is used to measure the quantitative uncertainty of BPA in this work. Assume $m(\cdot)$ is a mass function defined on the frame of discernment, as explained in reference [42], the calculation of Deng entropy $E_d(m)$ of the BPA is shown as

$$E_d(m) = -\sum_{A \subseteq \Theta} m(A) \log_2 \frac{m(A)}{2^{|A|} - 1}. \tag{15}$$

where A is the focal element of m, $|A|$ is the cardinality of A.

The FPR analysis based on the Deng entropy is adopted to denote the relative reliability preference between bodies of evidence. Fuzzy sets have been widely used in various applications and play an important role in the decision-making process [43]. The concepts of FPR and the additive consistency of FPR are introduced briefly.

The fuzzy preference matrix is construct by the variance of entropy. If the system has more than two pieces of evidence, as explained in reference [25], the calculation of variance of entropy is shown as

$$V_i = e^{E_d(m_i)},\ 1 \leq i \leq k \tag{16}$$

$$Var_i = \mathrm{Var}\left(\left\{\overline{V}_1, \overline{V}_2, \ldots, \overline{V}_{i-1}, \overline{V}_{i+1}, \ldots, \overline{V}_k\right\}\right). \tag{17}$$

where $\overline{V}_i = V_i / \sum_{i=1}^{k} V_i$, and Var_i denotes the variance. Then, the off-diagonal elements ρ_{ij} and ρ_{ji} of the fuzzy preference matrix can be computed by.

$$\rho_{ij} = \frac{Var_i}{Var_i + Var_j},\quad \rho_{ji} = \frac{Var_j}{Var_i + Var_j}. \tag{18}$$

Let P be a fuzzy preference matrix for the set $\mathfrak{M}$ of alternatives $\mathfrak{M} = \{\mathcal{M}_1, \mathcal{M}_2, \ldots \mathcal{M}_n\}$, as explained in reference [43], the defined of P is shown as

$$P = \left(\rho_{ij}\right)_{n \times n} = \begin{bmatrix} 0.5 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & 0.5 & \cdots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \cdots & 0.5 \end{bmatrix}. \tag{19}$$

where ρ_{ij} denotes the degree of preference of alternative $\mathcal{M}_i$ over alternative $\mathcal{M}_j$. Let P be a fuzzy preference relation $P = \left(\rho_{ij}\right)_{n \times n}$, if P is a complete FPR as explained in reference [44], which satisfies the following additive consistency properties for all i, j and k.

$$\begin{cases} \rho_{ij} + \rho_{ji} = 1, \\ \rho_{ii} = 0.5, \\ P_{ik} = P_{ij} + P_{jk} - 0.5 \end{cases}. \tag{20}$$

where $1 \leq i \leq n$, $1 \leq j \leq n$ and $1 \leq k \leq n$, then P is called an additive consistent FPR. Based on the complete fuzzy preference relation P, as explained in reference [26], a consistency matrix $\overline{P}$ which satisfies the additive consistency is shown as

$$\overline{P} = \left(\overline{\rho}_{ik}\right)_{n \times n} = \left(\frac{1}{2n} \sum_{j=1}^{n}\left(\rho_{ij} - \rho_{ji} + \rho_{jk} - \rho_{kj}\right) + 0.5\right)_{n \times n}. \tag{21}$$

And then, as explained in reference [26], the calculation of the boundary constant ξ and the consistency degree ς are shown as

$$\begin{cases} \chi_i = \frac{1}{n} \sum\limits_{j=1}^{n} \overline{\rho}_{ij} \\ \varepsilon = \max(\chi_i | 1 \leq i \leq n) \\ \mu = \min(\chi_i | 1 \leq i \leq n) \\ \xi = \frac{1}{2 \cdot \max(0.5, \ (\varepsilon - \mu))} \\ \varsigma = 1 - \frac{2}{n(n-1)} \sum\limits_{i=1}^{n} \sum\limits_{k=1, k \neq i}^{n} \left|\rho_{ik} - \overline{\rho}_{ik}\right| \end{cases}. \tag{22}$$

where χ_i is the average value of preference values of alternative, ε is the maximum value of all χ_i, μ is the minimum value of all χ_i, ξ is the boundary constant to let the preference values in the consistency matrix $\overline{P}$ is between zero and one, ς represents the consistency degree between P and $\overline{P}$. The larger the value of ς, the more the consistency of the fuzzy preference relation. If the value of ς is close to one, then the information of fuzzy preference relation is more consistent $\xi \in [0, 1]$, $\varsigma \in [0, 1]$, $1 \leq i \leq n$, $1 \leq k \leq n$. As explained in reference [26], the calculation of the modified consistency matrix $\widetilde{P}$ is shown as

$$\widetilde{P} = (\widetilde{\rho}_{ik})_{n \times n} = \left(\overline{\rho}_{ik} \times \kappa + \frac{1}{2}(1 - \kappa)\right)_{n \times n}. \tag{23}$$

where κ denotes the modified constant. And $\kappa = \xi \times \zeta$, $\kappa \in [0, 1]$. The ranking value R_i of the alternative $\mathcal{M}_i$ in the set $\mathfrak{M}$ is calculation as follows

$$R_i = \frac{2}{n^2 - n} \sum_{j=1, j \neq i}^{n} \widetilde{\rho}_{ij}. \tag{24}$$

where $1 \leq i \leq n$, $1 \leq j \leq n$ and $\sum_{i=1}^{n} R_i = 1$.

3.3. The Improved Fusion Algorithm

With the credibility degree crd_i and the ranking value of alternative BPAs R_i, the support degree of the BPA is denoted as P_{Sup_i},

$$P_{Sup_i} = crd_i \times R_i. \tag{25}$$

Based on the weight P_{Sup_i}, the weighted average of the evidence (WAE) is given as follow.

$$\text{WAE}(m) = \sum_{i=1}^{k} \left(\overline{P}_{Sup_i} \times m_i\right). \tag{26}$$

where $\overline{P}_{Sup_i} = P_{Sup_i} / \sum_{i=1}^{k} P_{Sup_i}$. Therefore, the modified mass function obtained by Equation (26) will be fused with Dempster's rule of combination n-1 times when there are n pieces of evidence.

3.4. Numerical Verification

A numerical example obtained from reference [21] is illustrated to verify the effectiveness of the improved method in dealing with conflict evidences. Suppose the recognition target is *A* based on multiple sensor data given in Table 1. It showed five different types of sensors, and the FOD is given by $\Theta = \{A, B, C\}$. The results using different combination rules are shown in Table 2.

Table 1. Basic probability assignment (BPA) of the sensor data.

BPA	{*A*}	{*B*}	{*C*}	{*A,C*}
m_1	0.41	0.29	0.30	0.00
m_2	0.00	0.90	0.10	0.00
m_3	0.58	0.07	0.00	0.35
m_4	0.55	0.10	0.00	0.35
m_5	0.60	0.10	0.00	0.30

Table 2. Results of the evidence using different fusion methods.

Evidence	Method	{*A*}	{*B*}	{*C*}	{*A,C*}
m_1, m_2, m_3	Dempster	0	0.6350	0.3650	0
	Murphy [20]	0.4939	0.4180	0.0792	0.0090
	Deng et al. [21]	0.4974	0.4054	0.0888	0.0084
	Zhang et al. [22]	0.5681	0.3319	0.0929	0.0084
	The proposed method	0.8308	0.0532	0.1046	0.0115
m_1, m_2, m_3, m_4	Dempster	0	0.3321	0.6679	0
	Murphy [20]	0.8362	0.1147	0.0410	0.0081
	Deng et al. [21]	0.9089	0.0444	0.0379	0.0089
	Zhang et al. [22]	0.9142	0.0395	0.0399	0.0083
	The proposed method	0.9535	0.0046	0.0334	0.0085
m_1, m_2, m_3, m_4, m_5	Dempster	0	0.1422	0.8578	0
	Murphy [20]	0.9620	0.0210	0.0138	0.0032
	Deng et al. [21]	0.9820	0.0039	0.0107	0.0034
	Zhang et al. [22]	0.9820	0.0034	0.0115	0.0032
	The proposed method	0.9886	0.0004	0.0091	0.0032

As can be seen in Table 2, although more evidence supports target *A*, a wrong decision was still achieved with Dempster's method. When the number of evidence were not adequate, the performance of Murphy's method was not satisfactory. Obviously, the simple averaging and other weight averaging can provide reasonable results, but the proposed method in this work is much better in dealing with conflicting evidence.

3.5. An Example of Fault Diagnosis Application

Another example given in reference [45] has been utilized to further demonstrate the effectiveness of the improved DST in fault diagnosis. The BPAs of the sensor data are directly adopted from reference [46]. Suppose the frame of discernment is *F*, which have three types of fault in a motor rotor, denoted as $F_1 = \{Rotor\ unbalance\}$, $F_2 = \{Rotor\ misalignment\}$, and $F_3 = \{Pedestal\ looseness\}$, respectively. Three vibration accelerometer sensors are installed in different positions to collect the vibration signals, denoted by $S = \{S_1, S_2, S_3\}$. The frequency of vibration signal locating at 1×, 2× and 3× (× denotes rotor rotating frequency) are considered as the fault features, as are shown in Table 3.

Table 3. The obtained BPAs.

	*Freq*1				*Freq*2		*Freq*3			
	$\{F_2\}$	$\{F_3\}$	$\{F_1,F_2\}$	$\{F_1,F_2,F_3\}$	$\{F_2\}$	$\{F_1,F_2,F_3\}$	$\{F_1\}$	$\{F_2\}$	$\{F_1,F_2\}$	$\{F_1,F_2,F_3\}$
$S_1 : m_1$	0.8176	0.0003	0.1553	0.0268	0.6229	0.3771	0.3666	0.4563	0.1185	0.0586
$S_2 : m_2$	0.5658	0.0009	0.0646	0.3687	0.7660	0.2341	0.2793	0.4151	0.2652	0.0404
$S_3 : m_3$	0.2403	0.0004	0.0141	0.7452	0.8598	0.1402	0.2897	0.4331	0.2470	0.0302

The modified mass function could also be calculated with the proposed method. The weighted average of the evidence shown in the Table 4 can be obtained by Equation (26). It can be seen that the probability of F_2 is the largest, which can be preliminarily judged as the fault type. The modified mass function will be fused with Dempster's rule of combination. The fusion results given in reference [46] were obtained by Equation (10) using the Dempster's rule 2 times, which is also shown in Tables 5–7. The corresponding Target column represents the fault type for fusion diagnosis.

Table 4. The modified BPAs.

	*Freq*1				*Freq*2		*Freq*3			
	$\{F_2\}$	$\{F_3\}$	$\{F_1,F_2\}$	$\{F_1,F_2,F_3\}$	$\{F_2\}$	$\{F_1,F_2,F_3\}$	$\{F_1\}$	$\{F_2\}$	$\{F_1,F_2\}$	$\{F_1,F_2,F_3\}$
m_W	0.5836	0.0006	0.0870	0.3288	0.7576	0.2424	0.3109	0.4345	0.2118	0.0428

Table 5. Fusion results of different methods for motor rotor fault diagnosis at 1X frequency.

Method	$\{F_2\}$	$\{F_3\}$	$\{F_1,F_2\}$	$\{F_1,F_2,F_3\}$	Target
Jiang et al. [46]	0.8861	0.0002	0.0582	0.0555	F_2
The proposed method	0.9277	0.0002	0.0364	0.0356	F_2

Table 6. Fusion results of different methods for motor rotor fault diagnosis at 2X frequency.

Method	$\{F_2\}$	$\{F_1,F_2,F_3\}$	Target
Jiang et al. [46]	0.9621	0.0371	F_2
The proposed method	0.9858	0.0142	F_2

Table 7. Fusion results of different methods for motor rotor fault diagnosis at 3X frequency.

Method	$\{F_1\}$	$\{F_2\}$	$\{F_1,F_2\}$	$\{F_1,F_2,F_3\}$	Target
Jiang et al. [46]	0.3384	0.5904	0.0651	0.0061	F_2
The proposed method	0.3343	0.6321	0.0334	0.0002	F_2

The improved DST is used to solve the fusion issue in the fault diagnosis mentioned above. According to the results shown in Tables 5–7, the conflict of sensor reports has been solved with the proposed method. We can notice that the proposed method can successfully detect the fault type F_2, which is consistent with those given in reference [46]. Thus, both the two methods can conduct the conflictive pieces of evidence and identify the fault type F_2 well. Moreover, it can be seen in Figures 2–4 that the proposed method can deal well with the conflictive pieces of evidence. The belief degrees assigned to the target F_2 at 1× frequency, 2× frequency and 3× frequency using the proposed method were separately 0.9277, 0.9858, and 0.6321, which are all higher than the method in reference [46].
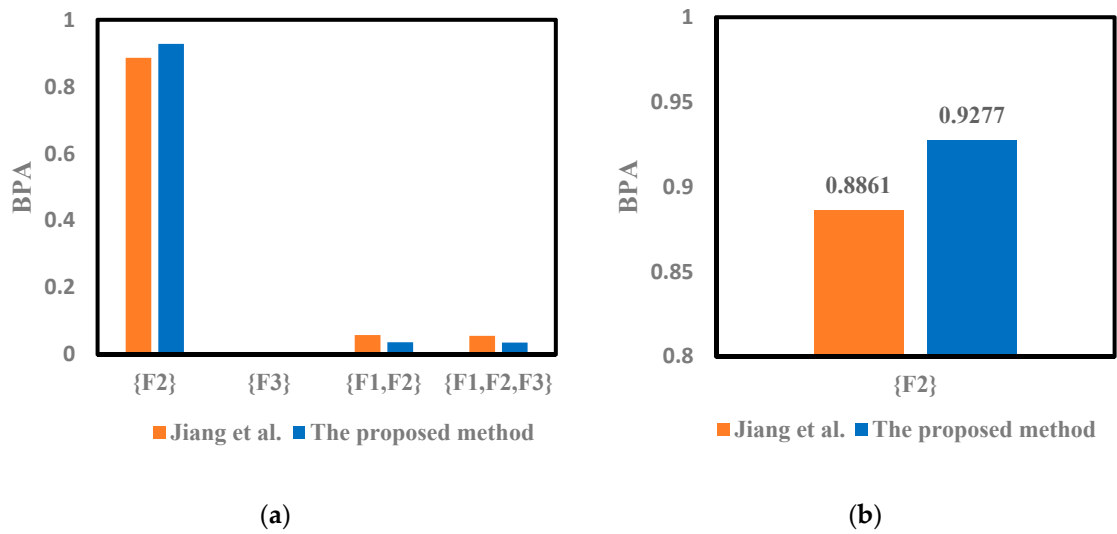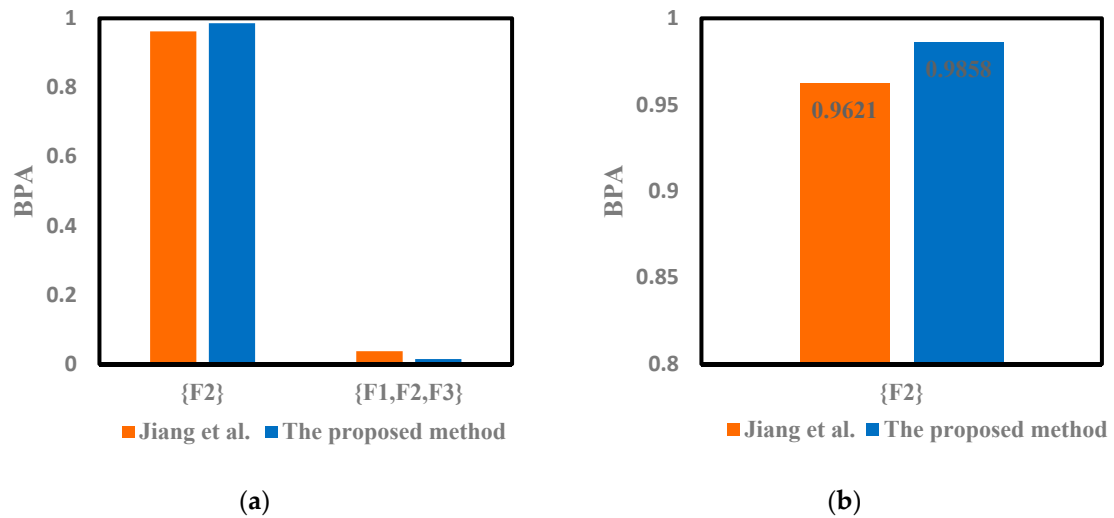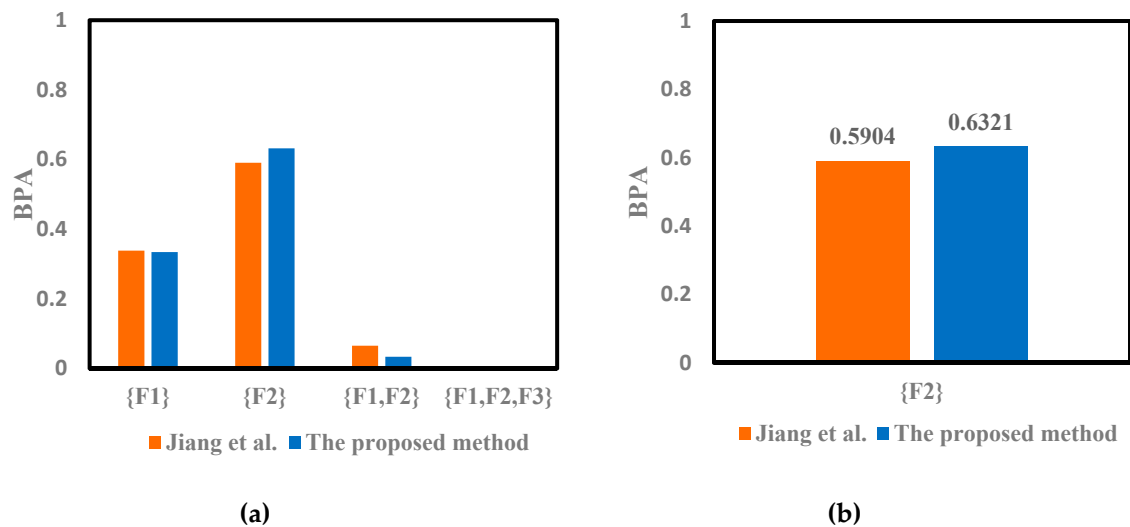
(**a**)

(**b**)

Figure 2. The comparison of different methods for motor rotor fault diagnosis at 1*X* frequency. (**a**) Fusion results of different methods. (**b**) The result of $\{F_2\}$ for 1*X*.

(**a**)

(**b**)

Figure 3. The comparison of different methods for motor rotor fault diagnosis at 2*X* frequency. (**a**) Fusion results of different methods. (**b**) The result of $\{F_2\}$ for 2*X*.

(a)   **(b)**

Figure 4. The comparison of different methods for motor rotor fault diagnosis at 3*X* frequency. (**a**) Fusion results of different methods. (**b**) The result of $\{F_2\}$ for 3*X*.

4. Experimental Analysis

The effectiveness of the improved Dempster-Shafer (D-S) evidence theory in dealing with conflicting evidence has been verified in the previous section. The proposed HCE framework in roller bearing fault diagnosis and the robustness of improved DST in information fusion will be illustrated in this section. The present technique is then applied for the rolling bearing fault diagnosis experiments on the Machinery Fault Simulator Magnum (MFS-MG) test-rig. The flowchart of the fault diagnosis using the proposed procedure is shown as Figure 5.
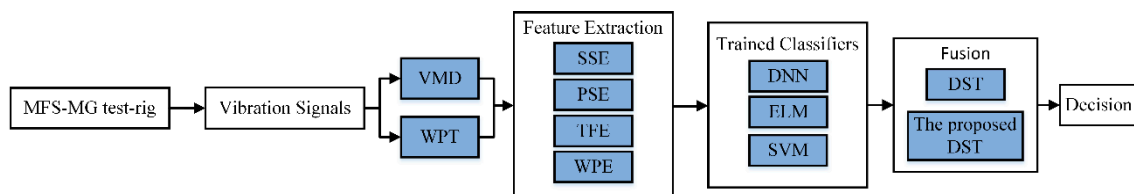
Figure 5. Flowchart of the proposed procedure.

4.1. The Experimental Set-Up

As shown in Figure 6, the vibration data set were acquired on the MFS-MG test rig, and the defective bearing of the type ER-12K was installed on the left side of the shaft. Accelerometer sensors were installed in vertical and horizontal on bearing seats. Sampling frequency was set to 25,600 Hz, and the rotating frequency of the motor was 29.87 Hz (about 1792 rpm). The fault types: Ball (B), cage (C), inner race (IR) and outer race (OR), as well as a normal (N) condition were used in the experiments. Each segment of the collected original vibration signal had 10,240 data points. The original vibration data and their frequency spectra are shown in Figure 7.
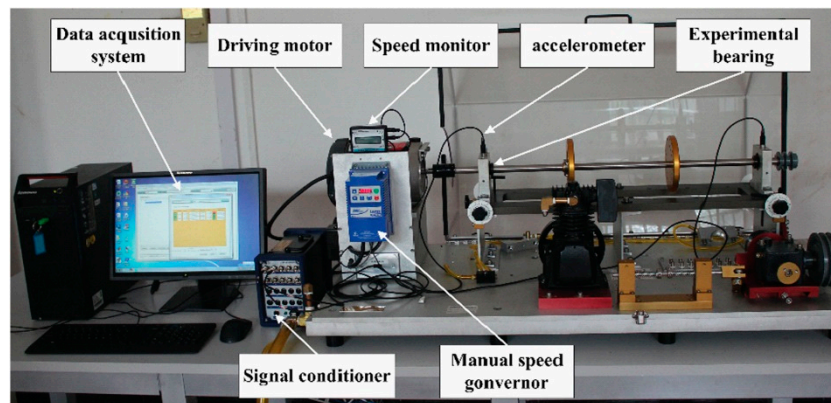
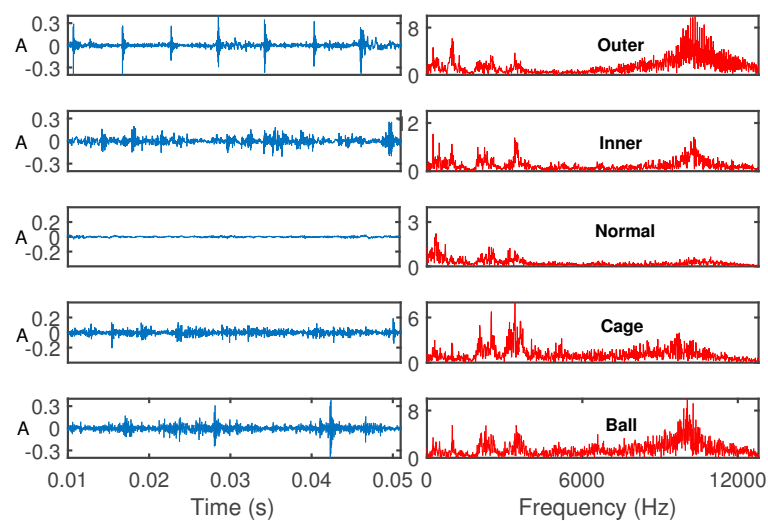**Figure 6.** Machinery Fault Simulator Magnum (MFS-MG) fault simulation test bench.

Figure 7. Original vibration signals and their spectra.

4.2. Entropy Feature Sets

We could obtain four entropy features, the features of vibration signals. The original vibration signal was decomposed with the VMD method, and the decomposed intrinsic mode function (IMF) were achieved. The key parameters used in VMD should be selected based on the empirical value, interested readers can refer to reference [47]. Assume $IMF_i = \{x_1, x_2, \ldots, x_K\}$, where K is the number of data points of IMF_i. The SSE, PSE, and TFE of each IMF_i were extracted using Equations (2), (5), and (6), respectively. Moreover, the WPESE of each original segment was also obtained using Equation (8). Here, a 3-level decomposition was used in WPT with the selected mother wavelet Db10. Since there were 112 samples for each experimental condition, the numbers of rows and columns in the feature matrix were 560 and 4, respectively. Figure 8 shows the entropy feature sets. The datasets were divided into two parts, and the former 75% of each class of data was randomly selected as training data, while the remaining 25% was testing data. The training data and the testing data was defined as a 420(row)–5(column) matrix and a 140(row)–5(column) matrix, respectively. The desirable classes were labeled with 1, 2, 3, 4, and 5. For example, outputs 1 and 3 were separately related to the first and the third class. In this way, three supervised classifiers could be used to identify the bearing faults.
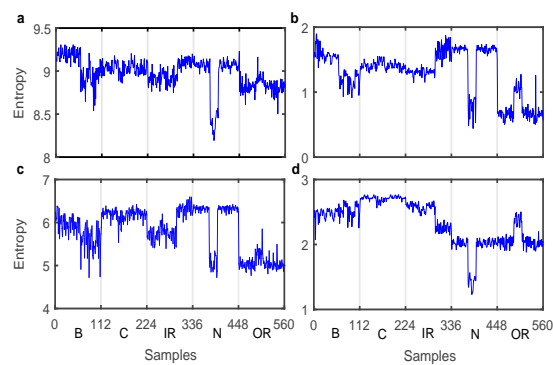
Figure 8. Four kinds of entropy features. (**a**) SSE (**b**) PSE (**c**) TFE (**d**) WPESE.

4.3. Classification Using Single-Stage Classifier

DNN, SVM, and ELM were separately adopted in the single-stage classification based on the above achieved entropy signatures. In this work, a large number of neurons were tested to find an optimal structure of DNN. The number of hidden layer neurons which resulted in the highest classification accuracy was selected as the optimum number. Then, the optimum DNN structure was constructed based on the obtained number of hidden layer neurons. Figure 9 shows the classification accuracies of DNN based on the different numbers of hidden layer neurons and mini-batch gradient descent (MBGD) algorithm. It can be seen in Figure 10 that the determined optimal number of hidden layer neurons is set to 13.
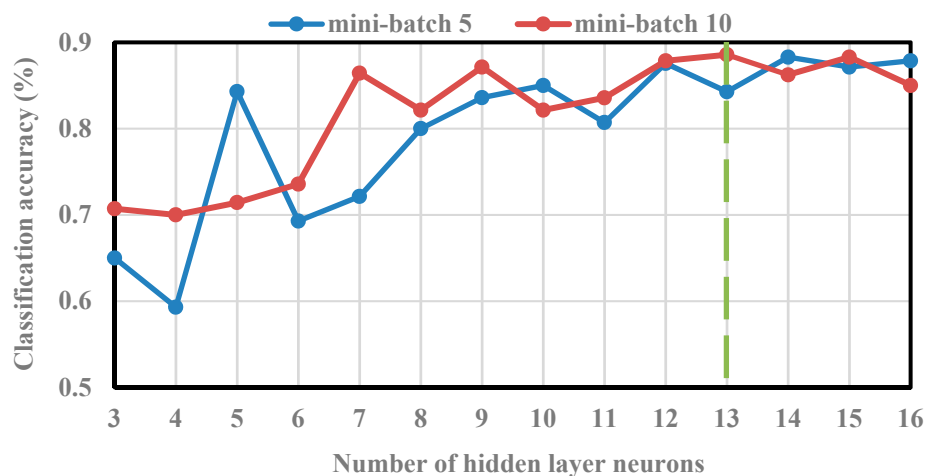
Figure 9. Classification accuracy using deep neural networks (DNN).

In the SVM technique, the Gaussian radial basis function (RBF) was selected as the kernel function, and the particle swarm optimization (PSO) was used to determine the optimized parameters in the SVM. The population size (pop), maximum number of iterations (maxgen), two acceleration constants (c_1, c_2), and the inertia weight (ψ) were set to $c_1 = 1.5$, $c_2 = 1.7$ and $\psi = 1$, pop = 20, maxgen = 100, respectively. In addition, the parameters of FOA used in ELM, such as the population size (pop) and maximum number of iterations (maxgen) were set to 20, 100, while the initial positions were set randomly.
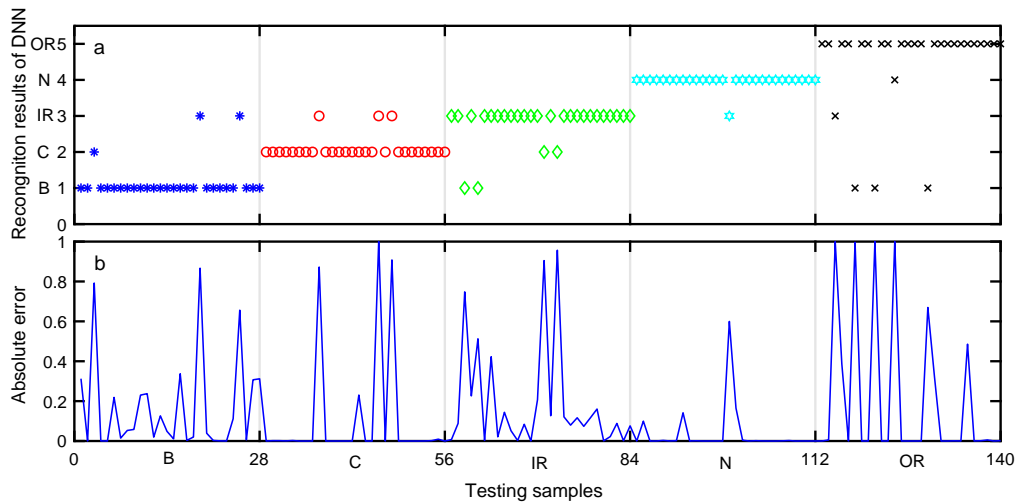
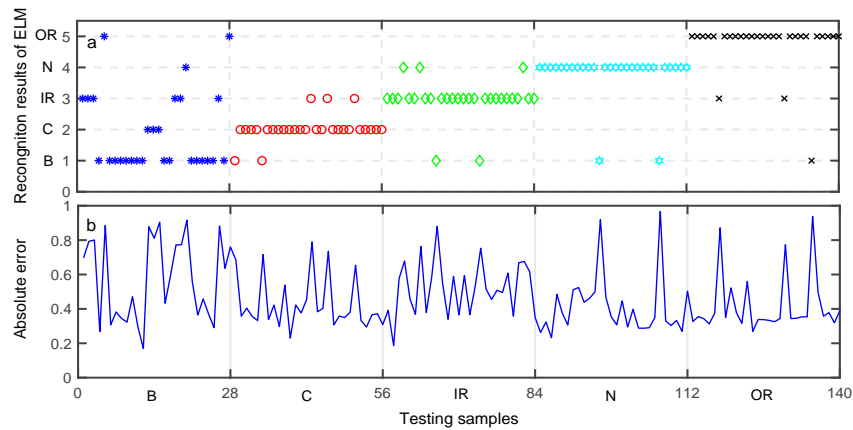**Figure 10.** Preliminary diagnosis of DNN (**a**) Recognition results. (**b**) Absolute error of the proposed approach output with respect to the desired output.

After data training using each classifier, the testing data set was used to validate the accuracy of each classifier model for bearing fault diagnosis. The aim of classification was to assign an input pattern to one of the 5 classes concerned in the present study and represented by the classification labels. The classification results of the testing data set obtained by preliminary diagnosis are shown in Figures 10–12. The performances of DNN, ELM, and SVM are illustrated in Tables 8–10, respectively. The meaning of Y-axis in Figure 10a, Figure 11a, and Figure 12a represents five bearing conditions, denoted by four fault types B, C, IR, OR as well as a normal condition (N).

Figure 11. Preliminary diagnosis of extreme learning machine (ELM). (**a**) Recognition results. (**b**) Absolute error of the proposed approach output with respect to the desired output.
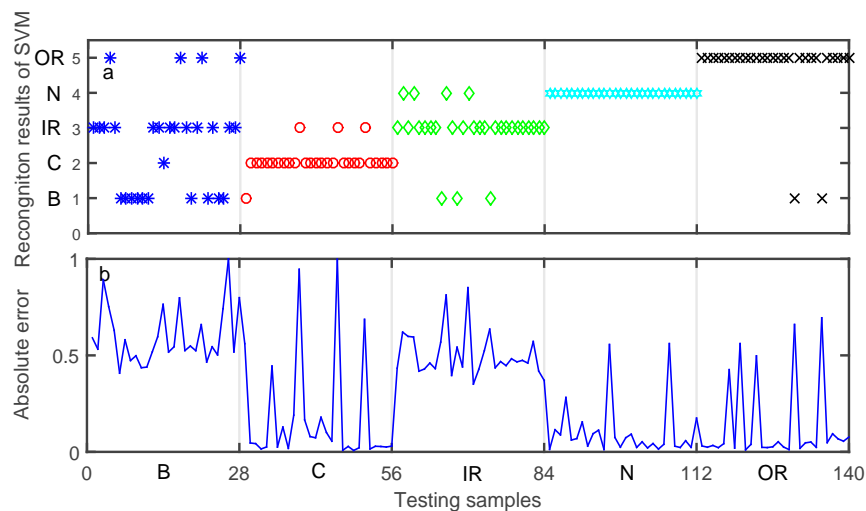
Figure 12. Preliminary diagnosis of support vector machine (SVM). (**a**) Recognition results. (**b**) Absolute error of the proposed approach output with respect to the desired output.

Table 8. Classification accuracy of DNN (%).

Bearing Condition	B	C	IR	N	OR	Average
B	89.29	3.57	7.14	0	0	
C	0	89.29	10.71	0	0	
IR	7.14	7.14	85.71	0	0	88.57
N	0	0	3.57	96.43	0	
OR	10.71	0	3.57	3.57	82.14	

Table 9. Classification accuracy of ELM (%).

Bearing Condition	B	C	IR	N	OR	Average
B	57.14	10.71	41.43	3.57	7.14	
C	7.14	82.14	10.71	0	0	
IR	7.14	0	82.14	10.71	0	80.81
N	7.14	0	0	92.86	0	
OR	3.57	0	7.14	0	89.29	

Table 10. Classification accuracy of SVM (%).

Bearing Condition	B	C	IR	N	OR	Average
B	25	3.57	53.57	0	14.29	
C	3.57	85.71	10.71	0	0	
IR	10.71	0	75	14.29	0	77.14
N	0	0	0	100	0	
OR	7.14	0	0	0	92.86	

Figure 10a shows the desired output and the output of the trained DNN. Figure 10b shows the absolute error of the DNN output with respect to the desired output, where a sample is misclassified when the absolute error is large. As can be seen from Table 8, the average classification accuracy of DNN is 88.57%. Figure 11a illustrates the desired output and the output of the trained ELM, while Figure 11b shows the absolute error of the ELM output with respect to the desired output. As can be seen from Table 9, the average classification accuracy of the testing data set using the ELM approach is about 80.81%. Similarly, Figure 12a shows the desired output and the output of the trained SVM, and Figure 12b shows the absolute error of the SVM output with respect to the desired output. As can be

seen from Table 10, the average classification accuracy of the testing data set using the SVM approach is only 77.14%.

It can be found that the classification rates separately using these three techniques were not good enough. Among them, DNN achieved the best classification results based on the deep learning technique as well as its optimal structures, compared with SVM and the ELM. The accuracy using single-stage classifier was still not good enough. Therefore, the data fusion method is necessary to be employed to increase the classification accuracy.

4.4. Results Using the HCE Algorithm and the Improved DST

Since the classification results were separately obtained using a single classifier, their results can be syncretized further. In this work, the fusion of the primary classification results was carried out using the improved DST method. First, three types of evidence were introduced as follows. E_1, E_2, and E_3 were the classification results using the supervised classifiers DNN, ELM, and SVM, respectively. The original Dempster's rule and the proposed method were both used to achieve the fusion results. In fact, the counter-intuitive results are often obtained when Dempster's rule of combination is utilized in some cases, especially, when the BOEs to be combined are highly conflicting.

In order to improve the diagnostic accuracy, DST and the proposed DST were used to fuse the preliminary diagnosis of HCE. The results of different methods are given in Table 11. In the fusion stage, each testing sample corresponded to a probabilistic output, which was the body of evidence. The meaning of X-axis in Figures 13–15 represents 140 bodies of evidence, while the meaning of Y-axis in Figures 13–15 represents fusion results of evidence using different methods. The fusion result of HCE by the proposed DST is shown in Figure 13, while the fusion result using HCE and the original DST is shown in Figure 14. A sample is misclassified when its fusion result is smaller than or equal to 0.5. It can be seen in Figures 13 and 14 that the classification accuracy using the proposed HCE and the improved DST is the highest, about 97.86%. In addition, the accuracy using the original DST is about 92.86%, which is also better than those using a single-stage classifier. Figure 15 illustrates the results using the technique given in reference [25]. We can find the result is better than those achieved using original DST, but it is still worse compared with our proposed methods. This well demonstrated that the proposed HCE approach combined with the improved DST can reliably be automatically used for roller bearing fault detection. It means that the fault detection accuracy can significantly be improved by applying HCE approach.

Table 11. Results of classification methods.

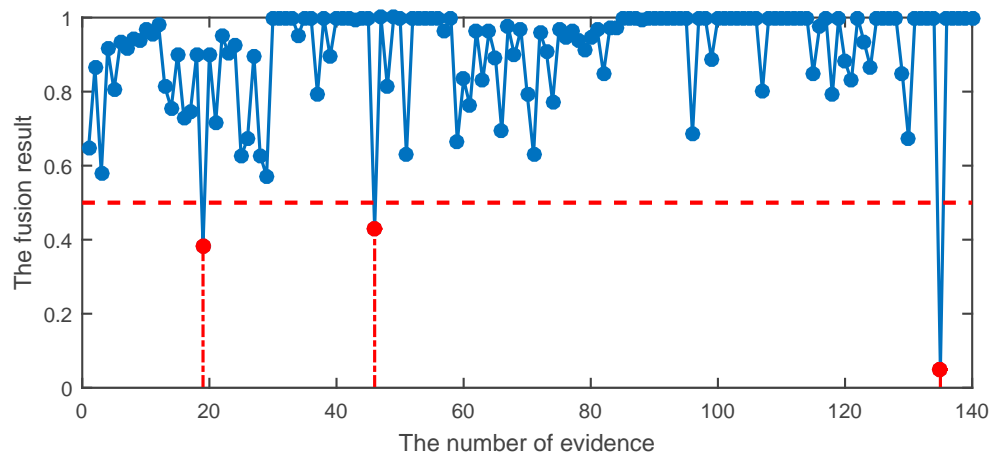| Method	Classification Rate (%)
HCE with improved DST	97.86
HCE with DST in [25]	96.43
HCE with DST	92.86
DNN	88.57
SVM	77.14
ELM	80.81

Figure 13. The fusion result using the proposed hybrid classifier ensemble (HCE) combined with the improved DST.
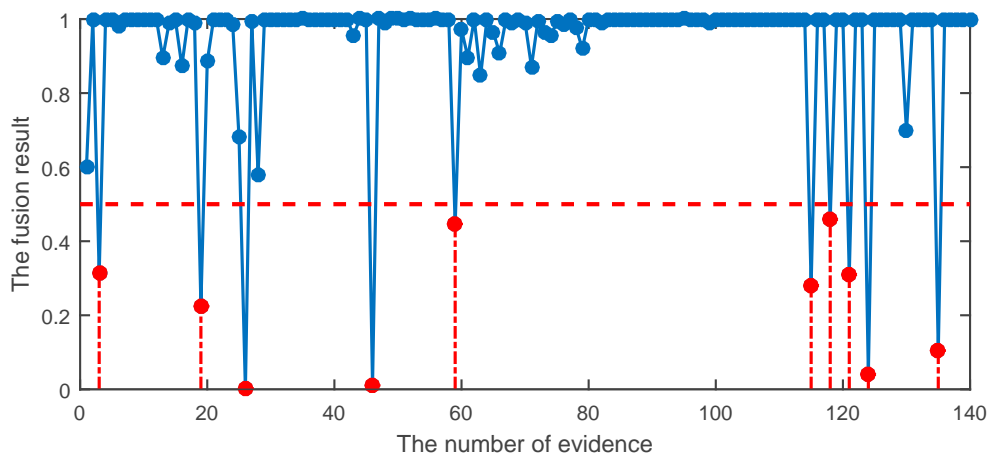

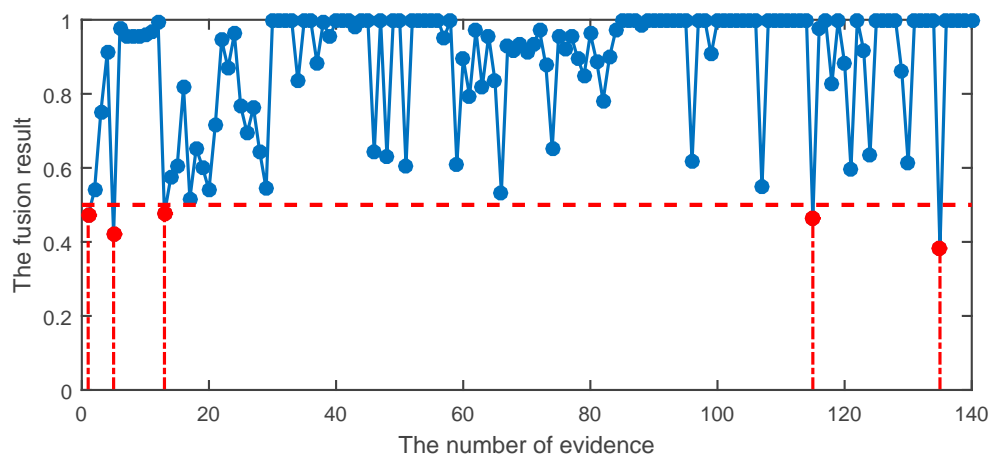**Figure 14.** The fusion result using HCE with original DST.

Figure 15. The fusion result using HCE with DST given in reference [25].

5. Conclusions

It is crucial to detect the relatively reliable evidence with the collected multi-source evidence in the process of information fusion. The HCE approach combined with the improved DST has been proposed for the fault diagnosis of roller bearings. The effects of support degree among the pieces of evidence, the uncertainty information of BPA, and the relative credibility of the evidence on the

weights are all considered in this improved DST. The improved DST can effectively deal with conflicts between the evidences and then improve the diagnostic accuracy. The cosine similarity is employed to indicate the confidence degree between the pieces of evidence. Entropy features are used to measure the quantitative uncertainty of BPA in the improved DST. In addition, entropy based FPR is employed to indicate the relative reliability preference between BOEs. Thus, the improved DST is much more reasonable in dealing with conflicts compared with the original DST. The effectiveness of the improved Dempster-Shafer theory has been verified via two examples.

In addition, SSE, PSE, TFE, and WPESE features have been utilized in the single-stage classification with DNN, SVM, and ELM in this work. Performances of the proposed HCE approach combined with the improved DST has been demonstrated on a bearing test-rig, compared with the original DST. It can be found that the overall error rate of the HCE approach can be greatly reduced using the improved DST, while the accuracy of the rolling element bearings diagnosis is successfully raised. Since there is not enough (complete) fault data for a rotating machine in practice, it is usually difficult dealing with a small sample and incomplete data in the process of decision-making. The proposed technique will be further investigated under these cases in the future.

References

1. Chen, Z.; Li, W.H. Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network. *IEEE Trans. Instrum. Meas.* **2017**, *66*, 1603–1702. [CrossRef]
2. Cui, L.L.; Huang, J.F.; Zhang, F.B.; Chu, F.L. HVSRMS localization formula and localization law: Localization diagnosis of a ball bearing outer ring fault. *Mech. Syst. Signal Process.* **2019**, *120*, 608–629. [CrossRef]
3. Zhu, Y.H.; Fu, Z.Y.; Fu, Z.; Chen, X.; Wu, Q. Multi-Features Fusion for Fault Diagnosis of Pedal Robot Using Time-Speed Signals. *Sensors* **2019**, *19*, 163. [CrossRef]
4. Wei, Z.X.; Wang, Y.X.; He, S.L.; Bao, J. A novel intelligent method for bearing fault diagnosis based on affinity propagation clustering and adaptive feature selection. *Knowl.-Based Syst.* **2017**, *116*, 1–12. [CrossRef]
5. Qin, Y. A new family of model-based impulsive wavelets and their sparse representation for rolling bearing fault diagnosis. *IEEE Trans. Ind. Electron.* **2018**, *65*, 2716–2726. [CrossRef]
6. Wang, Y.X.; Wei, Z.X.; Yang, J. Feature trend extraction and adaptive density peaks search for intelligent fault diagnosis of machine. *IEEE Trans. Ind. Inform.* **2019**, *15*, 105–115. [CrossRef]
7. Hu, G.; Li, H.; Xia, Y.; Luo, L. A deep Boltzmann machine and multi-grained scanning forest ensemble collaborative method and its application to industrial fault diagnosis. *Comput. Ind.* **2018**, *100*, 287–296. [CrossRef]
8. Pashazadeh, V.; Salmasi, F.R.; Araabi, B.N. Data driven sensor and actuator fault detection and isolation in wind turbine using classifier fusion. *Renew. Energ.* **2018**, *116*, 99–106. [CrossRef]
9. Zhong, J.H.; Wong, P.K.; Yang, Z.X. Fault diagnosis of rotating machinery based on multiple probabilistic classifiers. *Mech. Syst. Signal Process.* **2018**, *108*, 99–114. [CrossRef]
10. Kaltungo, A.Y.; Sinha, J.K.; Elbhbah, K. An improved data fusion technique for faults diagnosis in rotating machines. *Measurement* **2018**, *58*, 27–32. [CrossRef]
11. Wolpert, D. The supervised learning no-free-lunch theorems. In Proceedings of the 6th Online World Conference on Soft Computing in Industrial Applications, WSC6, 10–24 September 2001; pp. 25–42.
12. Wozniak, M.; Grana, M.; Corchado, E. A survey of multiple classifier systems as hybrid systems. *Inf. Fusion* **2014**, *16*, 3–17. [CrossRef]
13. Aburomman, A.A.; Reaz, M.B.I. A survey of intrusion detection systems based on ensemble and hybrid classifiers. *Comput. Secur.* **2017**, *65*, 135–152. [CrossRef]

14. Hall, D.L.; Llinas, J. An introduction to multisensor data fusion. *Proc. IEEE* **2002**, *85*, 6–23. [CrossRef]

15. Ai, Y.T.; Guan, J.Y.; Fei, C.W.; Tian, J.; Zhang, F.L. Fusion information entropy method of rolling bearing fault diagnosis based on n-dimensional characteristic parameter distance. *Mech. Syst. Signal Process.* **2017**, *88*, 123–136. [CrossRef]

16. Hui, K.H.; Lim, M.H.; Leong, M.S.; Al-Obaidi, S.M. Dempster-Shafer evidence theory for multi-bearing faults diagnosis. *Eng. Appl. Artif. Intel.* **2017**, *57*, 160–170. [CrossRef]

17. Gong, Y.J.; Su, X.Y.; Qian, H.; Yang, N. Research on fault diagnosis methods for the reactor coolant system of nuclear power plant based on D-S evidence theory. *Ann. Nucl. Energ.* **2018**, *112*, 395–399. [CrossRef]

18. Zadeh, L.A. A simple view of the Dempster–Shafer theory of evidence and its implication for the rule of combination. *AI Mag.* **1986**, *2*, 85–90.

19. Haenni, R. Shedding new light on Zadeh's criticism of Dempster's rule of combination. In Proceedings of the 7th International Conference on Information Fusion, Philadelphia, PA, USA, 25–28 July 2005; pp. 25–28.

20. Murphy, C.K. Combining belief functions when evidence conflicts. *Decis. Support Syst.* **2000**, *29*, 1–9. [CrossRef]

21. Deng, Y.; Shi, W.K.; Zhu, Z.F.; Liu, Q. Combining belief functions based on distance of evidence. *Decis. Support Syst.* **2004**, *38*, 9–493.

22. Zhang, Z.; Liu, T.; Chen, D.; Zhang, W. Novel algorithm for identifying and fusing conflicting data in wireless sensor networks. *Sensors* **2014**, *14*, 9562–9581. [CrossRef]

23. Jousselme, A.L.; Grenier, D.; Bosse, E. A new distance between two bodies of evidence. *Inf. Fusion* **2001**, *2*, 91–101. [CrossRef]

24. Tessem, B. Approximations for efficient computation in the theory of evidence. *Artif. Intell.* **1993**, *61*, 315–329. [CrossRef]

25. Qian, J.; Guo, X.F.; Deng, Y. A novel method for combining conflicting evidences based on information entropy. *Appl. Intell.* **2017**, *46*, 876–888.

26. Chen, S.M.; Lin, T.E.; Lee, L.W. Group decision making using incomplete fuzzy preference elations based on the additive consistency and the order consistency. *Inf. Sci.* **2014**, *259*, 1–15.

27. Dragomiretskiy, K.; Zosso, D. Variational mode decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544. [CrossRef]

28. Wang, Y.X.; Yang, L.; Xiang, J.W.; Yang, J.; He, S. A hybrid approach to fault diagnosis of roller bearings under variable speed conditions. *Meas. Sci. Technol.* **2017**, *28*, 125104. [CrossRef]

29. Cheng, G.; Chen, X.H.; Li, H.Y.; Li, P.; Liu, H. Study on planetary gear fault diagnosis based on entropy feature fusion of ensemble empirical mode decomposition. *Measurment* **2016**, *91*, 140–154. [CrossRef]

30. Xing, X.S. Physical entropy, information entropy and their evolution equations. *Sci. China A Math.* **2001**, *44*, 1331–1339. [CrossRef]

31. Pasi, L. Feature selection using fuzzy entropy measures with similarity classifier. *Expert Syst. Appl.* **2011**, *38*, 4600–4607.

32. Fei, C.W.; Bai, G.C.; Tang, W.Z. Quantitative diagnosis of rotor vibration fault using process power spectrum entropy and support vector machine method. *Shock Vib.* **2014**, *2014*, 957531. [CrossRef]

33. Yu, D.; Yang, Y.; Cheng, J. Application of time–frequency entropy method based on Hilbert–Huang transform to gear fault diagnosis. *Measurement* **2007**, *40*, 823–830. [CrossRef]

34. Wei, Z.; Gao, J.; Zhong, X.; Jiang, Z.; Ma, B. Incipient fault diagnosis of rolling element bearing based on wavelet packet transform and energy operator. *WSEAS Trans. Syst.* **2011**, *10*, 81–90.

35. Chen, Z.Q.; Deng, S.C.; Chen, X.D.; Li, C.; Sanchez, R.V.; Qin, H. Deep neural networks-based rolling bearing fault diagnosis. *Microelectron. Reliab.* **2017**, *75*, 327–333. [CrossRef]

36. Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297.

37. Huang, G.B.; Zhu, Q.Y.; Siew, C.K. Extreme learning machine: Theory and applications. *Neurocomputing* **2006**, *70*, 489–501. [CrossRef]

38. Huang, G.B.; Chen, L.; Siew, C.K. Universal approximation using incremental constructive feedforward networks with random hidden nodes. *IEEE Trans. Neural Netw.* **2006**, *17*, 879–892.

39. Dempster, A.P. Upper and lower probabilities induced by a multi-valued mapping. *Ann. Math. Stat.* **1967**, *38*, 325–339. [CrossRef]

40. Klir, G.J. Generalized information theory. *Fuzzy Sets Syst.* **1991**, *40*, 127–142. [CrossRef]

41. Wen, C.L.; Wang, Y.C.; Xu, X.B. Fuzzy information fusion algorithm of fault diagnosis based on similarity measure of evidence. *Lect. Notes Comput. Sci.* **2008**, *5264*, 506–515.
42. Deng, Y. Deng entropy. *Chaos Solitons Fractals* **2016**, *91*, 549–553. [CrossRef]
43. Ning, X.; Yuan, J.; Yue, X.; Ramirez-Serrano, A. Induced generalized choquet aggregating operators with linguistic information and their application to multiple attribute decision making based on the intelligent computing. *Intell. Fuzzy Syst.* **2014**, *27*, 1077–1085.
44. Tanino, T. Fuzzy preference orderings in group decision making. *Fuzzy Sets Syst.* **1984**, *12*, 117–131. [CrossRef]
45. Jiang, W.; Xie, C.; Zhuang, M.; Shou, Y.; Tang, Y. Sensor data fusion with Z-numbers and its application in fault diagnosis. *Sensors* **2016**, *16*, 1509. [CrossRef]
46. Wen, C.; Xu, X. *Theories and Applications in Multi-Source Uncertain Information Fusion—Fault Diagnosis and Reliability Evaluation*; Beijing Science Press: Beijing, China, 2012.
47. Wang, Y.X.; Markert, R. Filter bank property of variational mode decomposition and its applications. *Signal Process.* **2016**, *120*, 509–521. [CrossRef]