

Article

Misalignment Fault Diagnosis for Wind Turbines Based on Information Fusion

Yancai Xiao ^{1,*}, Jinyu Xue ¹, Long Zhang ² , Yujia Wang ¹ and Mengdi Li ¹

¹ School of Mechanical, Electronic and Control Engineering, Beijing Jiaotong University, Beijing 100044, China; 20126082@bjtu.edu.cn (J.X.); shiyanshi10071@163.com (Y.W.); shiyanshi10072@163.com (M.L.)

² Department of Electrical and Electronic Engineering, The University of Manchester, Manchester M139PL, UK; long.zhang@manchester.ac.uk

* Correspondence: ycxiao@bjtu.edu.cn

Abstract: Most conventional wind turbine fault diagnosis techniques only use a single type of signal as fault feature and their performance could be limited to such signal characteristics. In this paper, multiple types of signals including vibration, temperature, and stator current are used simultaneously for wind turbine misalignment diagnosis. The model is constructed by integrated methods based on Dempster–Shafer (D–S) evidence theory. First, the time domain, frequency domain, and time–frequency domain features of the collected vibration, temperature, and stator current signal are respectively taken as the inputs of the least square support vector machine (LSSVM). Then, the LSSVM outputs the posterior probabilities of the normal, parallel misalignment, angular misalignment, and integrated misalignment of the transmission systems. The posterior probabilities are used as the basic probabilities of the evidence fusion, and the fault diagnosis is completed according to the D–S synthesis and decision rules. Considering the correlation between the inputs, the vibration and current feature vectors’ dimensionalities are reduced by t-distributed stochastic neighbor embedding (t-SNE), and the improved artificial bee colony algorithm is used to optimize the parameters of the LSSVM. The results of the simulation and experimental platform demonstrate the accuracy of the proposed model and its superiority compared with other models.

Keywords: wind turbines; misalignment; fault diagnosis; information fusion; improved artificial bee colony algorithm; LSSVM; D–S evidence theory



Citation: Xiao, Y.; Xue, J.; Zhang, L.; Wang, Y.; Li, M. Misalignment Fault Diagnosis for Wind Turbines Based on Information Fusion. *Entropy* **2021**, *23*, 243. <https://doi.org/10.3390/e23020243>

Academic Editors:

Philip Broadbridge, Yongbo Li, Xihui (Larry) Liang and Fengshou Gu

Received: 13 January 2021

Accepted: 18 February 2021

Published: 20 February 2021

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In order to address global warming issues, many countries have reduced carbon emissions year by year as one of their targets for economic and social development. As one typical source of clean energy, wind power has significant advantages in terms of environmental and ecological impact compared with hydropower and nuclear power [1]. In recent years, wind power has been rapidly developed in many countries, and the installed capacity has been increasing year by year [2].

The working environment of wind turbines is often complex, so the failure rate of the components of wind turbines is relatively high [3]. If the key components of the wind turbine system fail, it will cause damage and even stop the whole turbine, resulting in huge economic losses. Therefore, in recent years, a large number of research work has been focused on fault diagnosis of wind turbines. The failures typically include blade failures, transmission system failures, generator failures, and tower failures. Among them, misalignment of the transmission system is one of the common failures [4]. Many reasons, such as bearing eccentricity, installation error, and coupling misalignment, can cause misalignment of the wind turbine transmission system that connects the gearbox and generator for a typical doubly-fed wind turbine [5]. The misalignment of the transmission system can inevitably lead to vibration of the unit, which will reduce the reliability of the power generation system. In addition, the misalignment failure can cause damage to gears

and bearings [6]. Therefore, it is necessary to monitor and diagnose the misalignment of the transmission system in doubly-fed wind turbines.

Although there is much work on the misalignment fault diagnosis for a conventional rotating system, there is little work for wind turbine misalignment diagnosis. In particular, a wind turbine presents additional and unique challenges as it operates under variable rotational conditions [7,8]. At present, the main research on detecting the misalignment of wind turbines includes the following work. Zhao et al. applied variational mode decomposition (VMD) to decompose the fault vibration signal to isolate features and diagnose the misalignment faults in a direct drive wind turbine [9]. Abdalla et al. diagnosed misalignment of planetary gearbox based on vibration measurements using spectrum analysis and modulation signal bispectrum (MSB) analysis [10]. Huang et al. applied the Hilbert–Huang transform (HHT) method for fault diagnosis of wind turbine rotors and discussed three typical faults by the HHT, including rotor mass imbalance, aerodynamic asymmetries, and yaw misalignment [11]. An and Kong proposed a modified empirical mode decomposition (EMD) method to extract characteristics from vibration signals and applied a back-propagation neural network to data from various sensors to diagnose faults of offshore wind turbines included stator imbalanced, rotor unbalanced, and bearing misalignment [12]. Villa et al. developed a statistical diagnosis algorithm based on the significance level of the modeled fault to detected unbalance fault and misalignment fault of wind turbine, and tested the algorithm on vibration from a test-bed [13]. He et al. analyzed the vibration characteristics of the transmission chain of a wind turbine based on double-elastic support with natural axial misalignment between the output shaft of gearbox and the shaft of generator causing vibration signals of normal gearbox blend with serious high-order gear mesh frequency and smooth modulation [14]. However, these methods mainly applied rely on single information, and their performance could be limited owing to the limited source of information.

Because the diagnosis based on single information often cannot reflect the overall condition, the information fusion methodology for multiple source information is needed for the diagnostic system. Information fusion is a synchronous and comprehensive processing of the information obtained from multiple sensors. It can ensure the integrity of the information from a different perspective and overcome the shortcomings of traditional single information to form a more objective and closer understanding of the system [15], which can greatly improve the accuracy of diagnosis.

Information fusion can be divided into three levels: data level, feature level, and decision level [16,17].

- Data level fusion. The direct fusion of signals collected by the same type of sensors retains the most information among the three levels.
- Feature level fusion. In this process, the signals from multiple sensors need to be preprocessed. Features are extracted to form the fusion vector and its attributes are used to judge the state of targets to be diagnosed.
- Decision level fusion. After initial state judgment of the target to be diagnosed, the final state is obtained based on the fusion of some decision rules. Decision level fusion is the highest among the three levels. Its real-time performance and fault tolerance are very good, but the information loss is very large, so more complex algorithms are needed.

At present, there are many research methods and achievements in decision level fusion, including Bayesian theory [18], Dempster–Shafer (D–S) evidence theory [19], fuzzy set theory [20,21], rough set theory [22], and so on. The classification principle of Bayesian theory is to calculate the posterior probability of an object (the probability that the object belongs to a certain class) using the prior probability and Bayes formula, and select the class with the largest posterior probability as the one to which the object belongs. In D–S evidence theory, trust function and likelihood function are obtained by calculating the orthogonal sum of basic probability distribution functions of different evidences. After fusing multiple evidences, the final decision is made according to decision rules. Among them, basic

probability distribution function is the probability distribution of all possible faults in each state, trust function is the lower bound of fault event probability, and likelihood function is the upper bound. Fuzzy set theory (FS) was founded by Zadeh. Membership $T(x)$ was used to describe fuzzy information. At this time, non-membership $F(x)$ did not appear. Then, intuitionistic fuzzy sets (IFSs) and interval intuitionistic fuzzy sets (IVIFSs) appeared successively. The fuzzy information processing technology developed from fuzzy set theory can provide a simple and effective means to explore uncertainty and simulate human recognition mechanism. Rough set theory, initially developed by Pawlak (1982), is a mathematical tool that deals with vague, uncertain, and incomplete information. Rough set theory has been successfully applied in many fields such as machine learning, pattern recognition, control systems, data mining, and image classification.

The advantages and limitations of the above four methods are listed in Table 1.

Table 1. Comparison of information fusion algorithms. D–S, Dempster–Shafer.

Approach	Advantages	Disadvantages
Bayes' theorem	Takes probability as the input data, has sufficient theoretical knowledge	Difficult to define prior probability function, lacks the ability to allocate the total uncertainty
D–S evidence theory	The premise is easier to meet, no need to know prior probability	Cannot solve the serious conflict or complete conflict of evidences
Fuzzy set theory	Based on local theory of classification, has strong adaptive ability	Determines the uncertainty according to the subjective judgement
Rough set theory	Deals with redundant information and inconsistent information effectively	Discretization of symptom attributes is needed

In this paper, based on the good theoretical basis and application effect of D–S evidence theory [23–27], it is used to complete decision fusion, which provides a sufficient fault diagnosis solution for wind turbine misalignment fault.

The aim of this paper is to use multiple sources of information to distinguish the misalignment-free (normal condition) and three different types of transmission misalignment. The main contributions are summarized as follows.

Multiple sources of information and integrated approach are used for wind turbine transmission misalignment detection. More specifically, the vibration, temperature, and stator current signal are taken as the original source, and their time domain features, frequency domain features, and time-frequency domain features are extracted as fault characteristics. t-distributed stochastic neighbor embedding (t-SNE) is used to reduce the vibration and current characteristics dimensionality, and then three posterior probability least squares support vector machine with parameters optimized by improved artificial bee colony algorithm are constructed. The probability outputs of the three LSSVM are taken as the basic probabilities of evidence fusion. The probability distribution after fusion is calculated according to the Dempster fusion rule. Compared with the non-fusion models, it is demonstrated that the model based on D–S evidence fusion has higher diagnostic accuracy for wind turbine misalignment faults.

The remainder of the paper is organized in the following way. In Section 2, the formulas of D–S evidence theory, posterior probability least squares support vector machines, and the improved artificial bee colony are presented in detail. Section 3 describes the specific steps for D–S fault diagnosis. Section 4 presents the fault diagnosis case study based on the simulation model. Section 5 presents the fault diagnosis case study based on the experimental platform. Section 6 concludes the current work.

2. Theoretical Background

2.1. D–S Evidence Theory

The D–S evidence theory is a method of uncertainty reasoning, proposed by Dempster in 1967 and later improved and developed by Shafer [28]. The D–S evidence method can produce a probability interval to an uncertain event by fusing multiple evidences with known probability distribution. As an indeterminate reasoning method, D–S evidence theory uses weaker conditions than Bayesian, and has the ability to quantify unknown and uncertainty [29]. The evidence theory contains three important functions: basic probability assignment function, belief function, and plausibility function. The basic probability assignment function is the probability distribution of all possible faults in each state, the belief function is the lower bound of the probability of the fault event, and the plausibility function is the upper bound of the probability of the fault event. The belief function and the plausibility function can be obtained by calculating the sum of the basic probability assignment function, and the final decision is made after combining multiple evidences from different sources.

The D–S evidence theory consists of the following parts [30].

- Frame of discernment:

A variety of possible mutually exclusive hypothesis $X_i (i = 1, 2, \dots, s)$ of a question constitute a finite and non-empty set, which is called the frame of discernment, denoted as $\Omega = \{X_1, X_2, \dots, X_s\}$.

- Basic probability assignment (BPA) function:

BPA function is also known as the mass function. Suppose H is a subset of Ω , if function $m(H)$ satisfying

$$\begin{aligned} & \text{(a) } m(\phi) = 0 \\ & \text{(b) } \sum m(H) = 1 \\ & \text{(c) } m(H) > 0 \end{aligned}$$

then, function $m(H)$ is called the basic probability assignment of H on Ω .

- Belief function:

In the frame of discernment, the belief function represents the sum of the basic probability assignment functions of all subsets of H . The expression of the belief function is as follows:

$$bel(H) = \sum_{Y/Y \subseteq H} m(Y) \quad (1)$$

- Plausibility function:

In the frame of discernment, the plausibility function represents the degrees of belief for not denying H , which is the sum of the basic probability assignments of all the subsets intersecting H . The expression of the plausibility function is as follows:

$$pl(H) = \sum_{S/S \cap H \neq \emptyset} m(S) \quad (2)$$

- Dempster's rule of combination:

Dempster's rule of combination is used to combine the BPA functions of multiple evidences. Although this rule is controversial at present, the authors of [31] have showed that it behaves perfectly when evidences do not conflict reciprocally. Only if we integrate conflicting evidences do we need to improve it. In this paper, there is no serious and complete conflict among the outputs from vibration signal, temperature signal, and stator current signal as evidences in this study. Therefore, Dempster's rule is still used here.

Suppose there are n independent evidences (sensors or expert opinions), H_1, H_2, \dots, H_n (are subsets of Ω), the BPA of them are m_1, m_2, \dots, m_n . Then, Dempster’s rule for the BPA functions on Ω is as follows:

$$m = m_1 \oplus m_2 \oplus \dots \oplus m_n \tag{3}$$

Specifically, it can be expressed as follows:

$$m(H) = \frac{\sum_{H_1 \cap \dots \cap H_n} m_1(H_1)m_2(H_2) \dots m_n(H_n)}{1 - K}, H \neq \emptyset \tag{4}$$

where the expression of K is as follows:

$$K = \sum_{H_1 \cap \dots \cap H_n = \emptyset} m_1(H_1)m_2(H_2) \dots m_n(H_n) \tag{5}$$

where K is the degree of conflict between evidences. When $K = 1$, the evidences are completely conflicted and cannot be synthesized by this formula; when K tends to 1, the evidences are highly conflicted, and synthesizing by this formula may lead to results contrary to fact [32].

- Decision rules:

The decision rule is to draw a diagnosis based on the uncertain interval $[bel(H), pl(H)]$ of the evidence. In the interval of $[0, 1]$, the uncertainty of a proposition is shown in Figure 1.

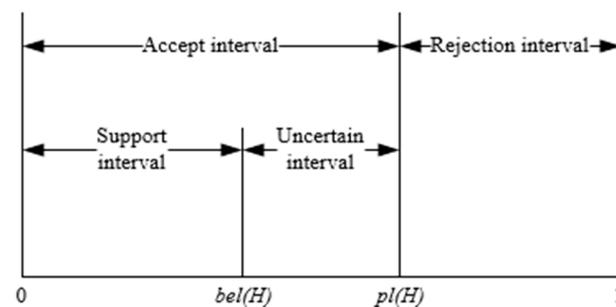


Figure 1. Uncertainty representation of a proposition.

In Figure 1, $[0, bel(H)]$ belongs to the support interval, $[0, pl(H)]$ is the accept interval, $[pl(H), 1]$ is the rejection interval, and $[bel(H), pl(H)]$ is the uncertain interval.

When making a decision, choose a value in the uncertain interval as the final trustworthiness of the proposition. If this value has the highest trustworthiness among the possible hypothesis, this assumption is the final decision result.

2.2. Posterior Probability Least Squares Support Vector Machine

In the study, the fault samples collected are limited, while support vector machine (SVM) and LSSVM can obtain high diagnosis accuracy based on small sample data. Moreover, the speed of LSSVM is faster than that of SVM, so LSSVM is selected to be the initial classifier to judge the state. As the input parameters of the D–S evidence fusion are basic probability assignments in all classification spaces, the hard output (whether or not) of the traditional classifier has to be converted to a soft one (probability) [33]; that is, the output of the classifier must be changed to the posterior probability output. For the two-class problem, the posterior probability can be calculated using the sigmoid function to map the output $f(x)$ ($+1, -1$) of the LSSVM to the $[0, 1]$ interval. Assuming that the probability is consistent with the sigmoid distribution, the posterior probability can be calculated [34]:

$$p(y = 1/f) = \frac{1}{1 + \exp(Af + B)} \tag{6}$$

$$p(y = -1/f) = 1 - p(y = 1/f) \quad (7)$$

where f is the classification result of the standard LSSVM, $p(y = 1/f)$ is the probability when the classification is correct under the condition that the output value is f , $p(y = -1/f)$ is the probability when the classification is wrong under the condition that the output value is f , and A and B are parameters. So, the key to calculating the posterior probability is to obtain parameters A and B . The posterior probability least squares support vector machine model is usually established by first establishing a standard LSSVM model, and then obtaining A and B on the training set (f_i, t_i) , where t_i is the target probability output of the standard LSSVM:

$$t_i = \begin{cases} \frac{N_+ + 1}{N_+ + 2}, f_i = +1 \\ \frac{1}{N_- + 2}, f_i = -1 \end{cases} \quad (8)$$

where N_+ is the number of positive samples; N_- is the number of negative samples; and the problem of obtaining parameters A and B is to solve the minimum likelihood optimization problem of the following, i.e.,

$$\min \left\{ - \sum_{i=1}^n [t_i \log(p_i) + (1 - t_i) \log(1 - p_i)] \right\} \quad (9)$$

where

$$p_i = \frac{1}{1 + \exp(Af_i + B)} \quad (10)$$

The Hessian matrix for solving (9) is as follows:

$$H(z) = \begin{bmatrix} \sum_i f_i^2 p_i (1 - p_i) & \sum_i f_i p_i (1 - p_i) \\ \sum_i f_i p_i (1 - p_i) & \sum_i p_i (1 - p_i) \end{bmatrix} \quad (11)$$

In order to get the minimum value of (9), the Hessian matrix must be positively determined. So, A and B are finally obtained by solving all the eigenvalues of the matrix that are greater than zero. The posterior probability can be obtained.

It is proved that the posterior probability least squares support vector machine with sigmoid function works well in practical applications [35], but this method can only be used for the two-class problem. The main methods for extending LSSVM from two-class to multi-class are the "one-versus-one" and "one-versus-all" methods. The Platt algorithm calculates the probability formula for each classifier as follows, where p_m is the probability that sample x belongs to the i -th class [35]:

$$p_m = (y = m/x) = \frac{1}{1 + \exp(A_m f(x) + B_m)} \quad (12)$$

2.3. The Improved Artificial Bee Colony

There are three kinds of kernel functions commonly used in LSSVM: linear kernel function, polynomial kernel function, and radial basis function (RBF) ($K(x_i, x_i) = \exp(-\|x_i - x_i\|^2 / \sigma^2)$, where σ is the kernel width). Many studies and experiments [36] show that, compared with other kernel functions, RBF can map the original space into an infinite dimensional space and find the hyperplane better. It is a better choice as the kernel function. Therefore, it is necessary to select the regularization parameter γ (necessary for LSSVM, determining the trade-off between the training error minimization and smoothness) and the kernel squared bandwidth σ^2 .

Choosing a better parameter value can greatly improve the performance of the LSSVM classifier and the accuracy of diagnosis. At present, the commonly used methods include trial and error, cross validation, grid search, and intelligent optimization algorithm [37].

Among them, the trial and error method not only consumes time and energy, but also the choice of parameters is greatly affected by subjective factors; the cross validation method divides the data set into training, validation, and testing, and different proportions will lead to different optimal models and optimal parameters; and the grid search method optimizes the model according to the set step size in the upper and lower limits of parameters, and then determines the optimal parameters, so the search speed is too slow and the precision is not high. Therefore, the advantages of the intelligent optimization algorithm are highlighted. It realizes the optimal distribution of food by simulating the behavior of animals in the population (interact information and cooperation among individuals). A swarm intelligence optimization algorithm is easy to implement and has high efficiency, so it is applied to the parameter optimization process of LSSVM.

Swarm intelligence optimization algorithms include genetic algorithm, particle swarm optimization, artificial fish swarm algorithm, artificial bee colony algorithm, and so on. Among them, artificial bee colony algorithm (ABC) is an optimization algorithm proposed in recent years, which not only has good optimization ability, but also controls less parameters in the process. Furthermore, it is simple, flexible, and easier to implement. The research [38] shows that the optimization performance of ABC is better than that of genetic algorithm and particle swarm algorithm, and the classification diagnosis accuracy of LSSVM optimized by ABC is higher than that of LSSVM optimized by genetic algorithm and particle swarm algorithm.

However, ABC has some shortcomings, such as slow convergence speed in the later stage of operation and the fact that it is easy to fall into local optimum. Therefore, in this paper, on the one hand, chaotic initialization is introduced in the artificial bee colony algorithm, which is used to initialize the population position to improve the diversity of the population and the ergodicity of the population search process. On the other hand, in the collecting bees stage of the artificial bee colony algorithm, the bees are divided into two parts: one part collects the optimal information of the region according to the original algorithm, and the other does Lévy flight around the global optimal solution to improve their global search capabilities. At the same time, in the observing bees stage, a search strategy based on the current local optimal solution (called pbest) is adopted to improve the local search ability of the algorithm.

(1) The logistic chaotic map is proposed to initialize the population. The equation for the logistic chaotic map is as follows:

$$y_{t+1} = \mu y_t(1 - y_t) \quad t = 0, 1, \dots, l \quad (13)$$

In the formula, $y_t \in (0, 1)$, t is the number of iterations of the chaotic sequence, μ is the control parameter of the chaotic sequence, and the value range is [3.75, 4] [39].

(2) Lévy flight was introduced in the evolution strategy to improve the performance of the algorithm and achieve good results [39]. The calculation method is based on

$$L(\alpha) = \left[\frac{\Gamma(1 + \alpha) \sin\left(\frac{\pi\alpha}{2}\right)}{\Gamma\left(\frac{1+\alpha}{2}\right) \alpha 2^{\left(\frac{\alpha-1}{2}\right)}} \right]^{1/\alpha} \quad (14)$$

where α is the characteristic index, which usually satisfies $0 < \alpha < 2$. $\Gamma(\cdot)$ is the Gamma function defined as

$$\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt.$$

Its update equation is as follows:

$$v_{ij} = x_{ij} + \alpha(x_{ij} - x_{best})L(\alpha) \quad (15)$$

where α is the step length, which usually meets the standard normal distribution, and $L(\cdot)$ is the random search path for Lévy flight.

(3) In the observing bees stage, for any current solution in each generation, the top $p\%$ solutions are randomly selected among all current solutions, and the best one (called pbest) can be used to balance global search capabilities and local development capabilities. The neighborhood search formula is as follows:

$$v_{ij} = x_{ij} + \varphi_{ij}(x_{ij} - x_{kj}) + \varnothing_{ij}(x_{randp,j}^{best} - x_{ij}) \quad (16)$$

where $k \in \{1, 2, \dots, S_N\}$, S_N is the number of solutions for the bee colony, $j \in \{1, 2, \dots, D\}$, D is the dimension of the optimization problem, $k \neq i$, $\varphi_{ij} \in [-1, 1]$, and $\varnothing_{ij} \in [0, 1.5]$.

3. Specific Steps for Misalignment Diagnosis

D–S evidence theory is used to carry out the fault diagnosis of wind turbines. The specific steps are as follows.

(1) Identify the frame of discernment of the fault diagnosis system

The frame of discernment is the common faults of the wind turbines misalignment in the study. At the same time, the normal working state of the unit is added. So, the frame of discernment is expressed as follows: {normal, parallel misalignment, angular misalignment and integrated misalignment}.

(2) Determination of evidence

The posterior probability least squares support vector machines are trained by the vibration signal, the temperature signal, and the stator current signal feature vectors separately. The hard outputs of the traditional LSSVM are mapped to the $[0, 1]$ interval using the sigmoid function. The soft outputs of the transformation are used as evidences for D–S evidence theory.

(3) Determination of basic probability assignment function, belief function, and plausibility function

The three least squares support vector machines give the probability vectors of all the classifications on the entire identification framework respectively, and the probability vectors to be directly used as the basic probability assignments, belief function, and plausibility function can be obtained by calculation.

(4) Evidence synthesis and diagnosis

According to Dempster's law, the probability vectors directly participate in the evidence fusion process. After the final probability vector is given, the final diagnosis result based on the probability vector after fusion can be obtained.

Figure 2 summarizes process of D–S evidence-based misalignment diagnosis.

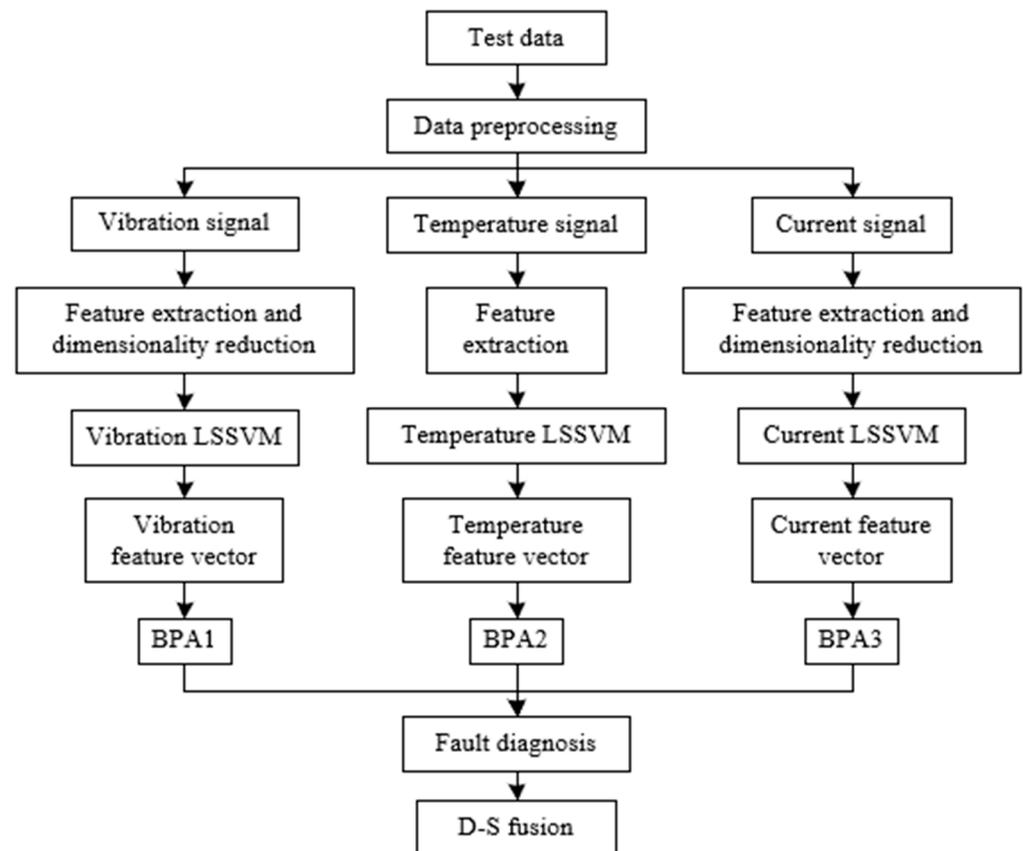


Figure 2. Process of Dempster–Shafer (D–S) evidence fusion. LSSVM, least square support vector machine; BPA, basic probability assignment.

4. The Simulation Case Studies of Misalignment Fault Diagnosis

The simulation wind turbine system is established by ADAMS 2013, MATLAB R2014a, and Ansys 17.0. The three-dimensional (3D) model of the 1.5 MW wind turbine is established using SolidWorks, and then it is imported into ADAMS 2013, where the Marker point is moved according to the type and degree of misalignment; that is, parallel misalignment is simulated by making the center of mass deviate from the center of rotation for a certain distance; angle misalignment is simulated by rotating the marker a certain angle around the y-axis, and placing the rotation axis of the coupling relative to the ground on the z-axis of the Marker point; and integrated misalignment is simulated by adding the parallel misalignment and angle misalignment in the local coordinate system (marker) of the left half coupling at the same time. The correctness of the models has been verified in the literature [40]. The vibration signals were extracted under the input speeds of $81.3^\circ/\text{s}$, using step function as the input of ADAMS, the simulation time is 1.5 s, and simulation steps are 6000 steps. The wind turbine models and its control system are established by SIMULINK/MATLAB, where the stator current was sampled at the same speed at which the vibration signal was sampled, and the sample frequency is 200 kHz. The correctness of the models has been verified in the literature [41]. After that, the high-speed gear shaft and the main shaft of the generator are introduced into HyperMesh to divide the grid. Then, the model is imported into Ansys Workbench to get the corresponding temperature signals (details in the literature [42]). In this paper, 100 samples are taken for each of the four types of diagnostic states (normal, parallel misalignment, angular misalignment, and integrated misalignment), of which 60 are for training and 40 are for testing. So, there are 240 (60×4) samples in the training set and there are 160 (40×4) samples in the testing set.

4.1. Data Processing

After the vibration signal, temperature signal, and stator current signal under four working conditions are collected, in order to make better use of them and get good diagnosis results, the feature indexes in the time, frequency, and time–frequency domain are extracted. Table 2 shows a 21-dimension mixed feature library of the vibration signal.

Table 2. Mixed feature library of vibration signals. IEMD, image extended empirical mode decomposition.

Feature Library	Feature	Index
Mixed-domain feature library	Time domain	Root mean square, square root amplitude, variance, standard deviation, kurtosis, waveform index, peak index, pulse index, margin index, kurtosis index
	Frequency domain	Center of gravity frequency, mean square frequency, frequency variance
	Time–frequency domain	The first eight energy entropy of the IMF (intrinsic mode function) component of IEMD decomposition

Suppose signal $x(x_0, x_1, x_2, \dots, x_{N-1})$ is a discrete time series with a finite length, the calculation formulas of time domain characteristic indexes are shown in Table 3, where \bar{x} is the mean value of the signal, x' is the average amplitude, and x_p is the peak value of the signal.

Table 3. Time domain characteristic index.

Time Domain Index		Calculation Formula
Dimensional indicators	Root mean square	$x_{rms} = (\frac{1}{N} \sum_{i=0}^{N-1} x_i^2)^{1/2}$
	Square root amplitude	$x_r = (\frac{1}{N} \sum_{i=0}^{N-1} x_i ^{1/2})^2$
	Variance	$\delta = \frac{1}{N} \sum_{i=0}^{N-1} (x_i - \bar{x})^2$
	Standard deviation	$x_{std} = (\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2)^{1/2}$
	Kurtosis	$\beta = \frac{1}{N} \sum_{i=1}^N x_i^4$
Dimensionless index	Waveform index	$K = x_{rms} / x' (x' = \frac{1}{N} \sum_{i=0}^{N-1} x_i)$
	Peak index	$C = x_p / x_{rms}$
	Pulse index	$I = x_p / x'$
	Margin index	$L = x_p / x_r$
	Kurtosis index	$K_r = \frac{\frac{1}{N} \sum_{i=1}^N x_i^4}{x_{rms}^4}$

In signal analysis, power spectrum analysis is usually used to extract the frequency domain index. Center of gravity frequency, mean square frequency, root mean square frequency, and frequency variance are commonly used. The sampling frequency is set as f_s , and the calculation formula of each index is shown in Table 4, where $S(\omega)$ is the power spectrum of discrete time series, $S(\omega) = X(\omega) \cdot \overline{X(\omega)}$, $X(\omega) = \sum_{i=0}^{N-1} x(i)e^{-j\pi\omega}$, ω is the angular frequency.

Table 4. Frequency domain characteristic index.

Frequency Domain Index	Calculation Formula
Center of gravity frequency	$FC = \frac{1}{2\pi f_s} \cdot \frac{\int_0^\pi \omega S(\omega) d\omega}{\int_0^\pi S(\omega) d\omega}$
Mean square frequency	$MSF = \frac{1}{4\pi^2 f_s^2} \cdot \frac{\int_0^\pi \omega^2 S(\omega) d\omega}{\int_0^\pi S(\omega) d\omega}$
Root mean square frequency	$RMSF = \left(\frac{1}{4\pi^2 f_s^2} \cdot \frac{\int_0^\pi \omega^2 S(\omega) d\omega}{\int_0^\pi S(\omega) d\omega} \right)^{\frac{1}{2}}$
Frequency variance	$VF = \frac{1}{4\pi^2 f_s^2} \cdot \frac{\int_0^\pi (\omega - 2\pi f_s FC) S(\omega) d\omega}{\int_0^\pi S(\omega) d\omega}$

Time–frequency analysis is a fault diagnosis method that combines the law and reason of frequency changing with time. In this paper, image extended empirical mode decomposition (IEMD) is used to process the vibration signal, and dual tree complex wavelet transform (DTCWT) is used to process the stator current signal (see the literature [43] for details).

The gearbox tooth temperature T_1 and the generator rotor shaft temperature T_2 are selected as the characteristic values of the temperature signal. Construct a two-dimensional vector of the temperature signal: $X = [T_1, T_2]$.

Table 5 is a mixed feature library with a total of 29 dimensions in the time domain, frequency domain, and time–frequency domain of the stator current signal (see the literature [41] for details).

Table 5. Mixed feature library of stator current signals.

Feature Library	Feature	Index
Mixed-domain feature library	Time domain	Root mean square, square root amplitude, variance, standard deviation, kurtosis, waveform index, peak index, pulse index, margin index, kurtosis index
	Frequency domain	Center of gravity frequency, mean square frequency, root mean square frequency, frequency variance
	Time–frequency domain	Sample entropy 1–5, energy entropy H_1, H_2, H_3, H_4, H_5 , spectral kurtosis a_1, a_2, a_3, a_4, a_5

In order to eliminate the influence of different input dataset dimensions and large numerical differences, the original dataset is normalized, i.e.,

$$y = \frac{y_{max} - y_{min}}{x_{max} - x_{min}} * (x - x_{min}) + y_{min} \tag{17}$$

where x is the value to be normalized, y_{min} is the lower bound of the normalized interval, and y_{max} is the upper bound of the normalized interval. In this paper, $y_{min} = 0, y_{max} = 1$, and the vector is normalized by column.

Because of the high dimensionality of the constructed vectors of the vibration signal and the stator current signal, not only does the amount of calculation increase, but also some difficulties are brought to fault diagnosis [44]. In order to make better use of various information and obtain good diagnostic results, the feature vectors are subjected to dimensionality reduction using t-SNE.

t-SNE based on conditional probability retains the similarity between high-dimensional and low dimensional space data and adopts symmetric objective function, and t distribu-

tion in low-dimensional space replaces Gaussian distribution, which solves the problem of crowding and clear visualization in low-dimensional space [45]. Its implementation steps are as follows:

- (1) Define a high-dimensional data set: $x = \{x_1, x_2, \dots, x_n\}$.
- (2) Compute the complexity parameter of the value equation c :

$$c = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} \quad (18)$$

$$\text{perp}(p_i) = 2^{H(p_i)} \quad (19)$$

$$H(p_i) = - \sum_j p_{j/i} \log_2^{p_{j/i}} \quad (20)$$

where p_i is the conditional probability of data points (other than x_i) with respect to x_i , $p_{j/i}$ is the conditional probability of high-dimensional data, p_{ij} is the joint probability density in the high-dimensional space, and q_{ij} is the joint probability density in the low-dimensional mapping space.

(3) Define the optimization parameters: the number of iterations T , the learning rate η , and the momentum factor at the t th ($t \leq T$) iteration $\alpha(t)$ ($0 < \alpha(t) < 1$). The value equation c is learned by the gradient descent method, and the low-dimensional mapping of the high-dimensional data is finally obtained:

$$\frac{\delta c}{\delta y_i} = 4 \sum_j (p_{ij} - q_{ij}) (y_i - y_j) (1 + \|y_i - y_j\|^2)^{-1} \quad (21)$$

where y_i and y_j are the mapping of the high-dimensional data x_i and x_j in the low-dimensional space.

In order to speed up the optimization process and prevent trapping into local minima, a relatively large momentum condition is imposed on the descent process. The current gradient value is summed to the previous gradient value for each iteration and then decays exponentially to determine the coordinates of the low-dimensional data. The momentum formula is as follows:

$$y^{(t)} = y^{(t-1)} + \eta \frac{\delta c}{\delta y} + \alpha(t) (y^{(t-1)} - y^{(t-2)}) \quad (22)$$

where y is the data in the low-dimensional space.

4.2. The Fault Diagnosis Results

In this paper, "one-versus-all" is used to extend LSSVM from two classifications to multiple classifications. That is, each time, one fault is selected as one type, and the rest of the states are selected as another type. In order to produce the posterior probabilities of the four classifications in the vibration feature space, four two-class LSSVM are constructed, and each LSSVM calculates a set of A and B , and then the corresponding posterior probability is calculated according to (5) and (6). In the same way, the probability vectors of the temperature and stator current signal classifiers for the four states can be obtained as the BPA of D-S evidence fusion.

The five-dimensional feature vectors of the vibration signal after t-SNE dimensionality reduction are used as the inputs, and the four working conditions of the transmission system are used as outputs to train the LSSVM, which is optimized by the improved artificial bee colony algorithm. The parameters of the four two-classification LSSVM in the vibration feature space are shown in Table 6. Four samples are selected, such as samples 5, 44, 82, and 130, and the corresponding BPA1 calculated is shown in Table 7.

Table 6. Parameters in four vibration least square support vector machine (LSSVM).

LSSVM	Γ	σ^2	A	B
(normal, the rest)	84.2784	31.1601	−10.8211	−4.2378
(parallel misalignment, the rest)	85.8947	30	−7.4866	−1.9740
(angular misalignment, the rest)	36.0326	39.9616	−6.1128	−1.5637
(integrated misalignment, the rest)	99.4093	96.6015	−6.9786	−2.4798

Table 7. Basic probability assignment 1 (BPA1) of vibration LSSVM.

Sample Number	Normal	Parallel Misalignment	Angular Misalignment	Integrated Misalignment
5	0.8433	0.0314	0.0538	0.0715
44	0.0080	0.9246	0.0295	0.0379
82	0.0025	0.0070	0.8873	0.1032
130	0.0208	0.0147	0.0495	0.9150

The two-dimensional feature vectors of the temperature signal are used as the inputs, and the four operating states of the transmission system are used as the outputs to train the optimized LSSVM. The parameters of the four binary LSSVM in the temperature feature space are shown in Table 8. The BPA2 calculated from the same four samples is shown in Table 9.

Table 8. Parameters in four temperature LSSVM.

LSSVM	Γ	σ^2	A	B
(normal, the rest)	97.4952	89.7017	−3.0178	−0.4262
(parallel misalignment, the rest)	98.3829	30	−2.8893	0.1963
(angular misalignment, the rest)	46.3907	88.9951	−3.9974	−0.3622
(integrated misalignment, the rest)	93.6931	96.3611	−2.4749	0.3805

Table 9. BPA2 of temperature LSSVM.

Sample Number	Normal	Parallel Misalignment	Angular Misalignment	Integrated Misalignment
5	0.7418	0.0387	0.0228	0.1967
44	0.1314	0.7670	0.0243	0.0773
82	0.3644	0.0357	0.5549	0.0450
130	0.4365	0.0394	0.0220	0.5021

The four-dimensional vectors after the dimensionality reduction of the stator current signal are used as inputs, and the four operating states of the transmission system are used as outputs to train the optimized LSSVM. The parameters of the four two-class LSSVM in the stator current feature space are shown in Table 10, and the BPA3 calculated by the same four samples is shown in Table 11.

Table 10. Parameters in four current LSSVM.

LSSVM	Γ	σ^2	A	B
(normal, the rest)	35.2345	53.0811	−4.858	−0.6011
(parallel misalignment, the rest)	87.5286	30	−3.2022	−0.1165
(angular misalignment, the rest)	100	30	−3.2755	−0.2514
(integrated misalignment, the rest)	77.1503	33.2639	−3.2803	−0.2146

Table 11. BPA3 of current LSSVM.

Sample Number	Normal	Parallel Misalignment	Angular Misalignment	Integrated Misalignment
5	0.8635	0.0440	0.0468	0.0457
44	0.0164	0.6193	0.3159	0.0484
82	0.0170	0.1236	0.8093	0.0501
130	0.0127	0.1093	0.0521	0.8259

Then, the probability assignments are calculated after the fusion of the three BPAs. The category with the highest degree of belief is selected as belonging to the class of the fusion model. Table 12 shows the basic and the fusion probability of the three LSSVM outputs for the selected test samples. Table 13 shows the fusion and classification results of the four test samples. Figure 3 shows the test samples' diagnosis results, in which "0" indicates normal operation, "1" indicates parallel misalignment, "2" indicates angular misalignment, and "3" indicates integrated misalignment.

Table 12. Probability assignment of three LSSVMs and fusion.

BPA 1	BPA 2	BPA3	D-S Evidence Fusion
[0.8433, 0.0314, 0.0538, 0.0715]	[0.7418, 0.0387, 0.0228, 0.1967]	[0.8635, 0.0440, 0.0468, 0.0457]	[0.9986, 0.0001, 0.0001, 0.0012]
[0.0080, 0.9246, 0.0295, 0.0379]	[0.1314, 0.7670, 0.0243, 0.0773]	[0.0164, 0.6193, 0.3159, 0.0484]	[0.0001, 0.9991, 0.0005, 0.0003]
[0.0025, 0.0070, 0.8873, 0.1032]	[0.3644, 0.0357, 0.5549, 0.0450]	[0.0170, 0.1236, 0.8093, 0.0501]	[0.0001, 0.0001, 0.9993, 0.0005]
[0.0208, 0.0147, 0.0495, 0.9150]	[0.4365, 0.0394, 0.0220, 0.5021]	[0.0127, 0.1093, 0.0521, 0.8259]	[0.0003, 0.0002, 0.0001, 0.9994]

Table 13. Fusion and classification results of four test samples.

D-S Evidence Fusion	Category	Is the Classification Correct?
[0.9986, 0.0001, 0.0001, 0.0012]	Normal	Yes
[0.0001, 0.9991, 0.0005, 0.0003]	Parallel misalignment	Yes
[0.0001, 0.0001, 0.9993, 0.0005]	Angular misalignment	Yes
[0.0003, 0.0002, 0.0001, 0.9994]	Integrated misalignment	Yes

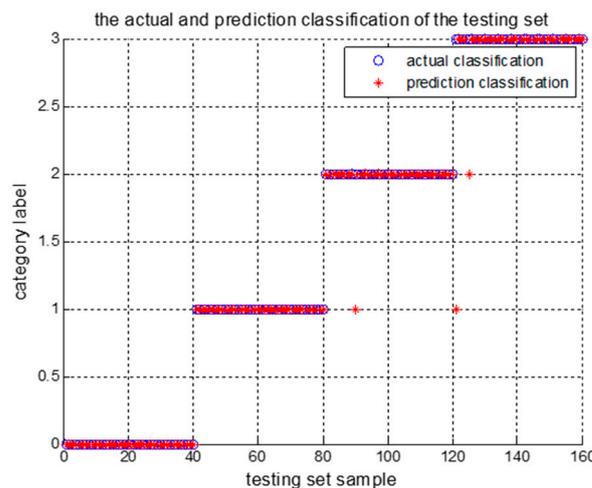


Figure 3. The diagnosis results of the testing set.

In order to better evaluate the performance of the fault diagnosis method, three indexes are adopted: the training set classification accuracy, the testing set classification accuracy, and the fault false alarm rate. The fault false alarm rate means that the fault does not actually occur, but the fault detection alarm is given by the detection system. The false alarm rate equals the number of false alarm samples divided by the total number

of actual fault-free samples. Table 14 compares the results of the sample sets diagnosed by the indexes of a single signal (vibration, temperature, or current signal) with the D–S evidence fusion.

Table 14. Comparison of diagnostic results.

Signal Selection	Training Set Classification Accuracy	Testing Set Classification Accuracy	False Alarm Rate
Vibration signal	100% (240/240)	85.625% (137/160)	5% (2/40)
Temperature signal	90.8333% (218/240)	81.25% (130/160)	35% (14/40)
Current signal	99.5833% (239/240)	84.375% (135/160)	10% (4/40)
D–S evidence fusion	100% (240/240)	98.125% (157/160)	0% (0/40)

From Table 14, it can be seen that the accuracy of D–S fusion is higher than that of any single signal, and the failure false alarm rate is equal to zero, lower than others, which proves the advantage of information fusion in the diagnosis of wind turbine misalignment fault.

5. Experimental Verification of Platform

In this paper, the 1.5 kW misalignment experimental platform is used for experimental verification. The platform is shown in Figure 4a. It includes a generator, coupling, gearbox, driving motor, and so on. The speed of the driving motor is changed by a planetary gear reducer with a transmission ratio of 1:50 to simulate the wind blowing blade speed, then it is accelerated by a planetary gear with a transmission ratio of 40:1 and a spur gear with a transmission ratio of 1.5:1 to drive the generator. The generator can be adjusted by the support to create parallel or angular misalignment.

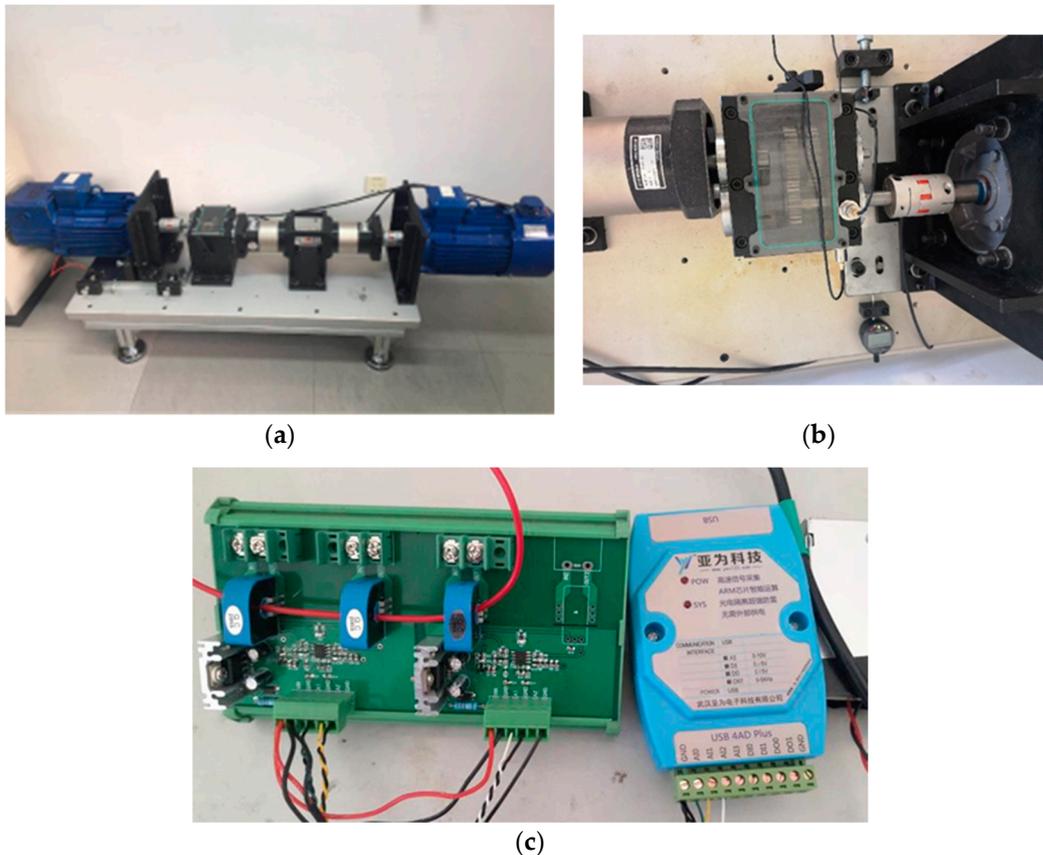


Figure 4. Experiment equipment. (a) The platform of wind turbine; (b) layout of vibration sensor; (c) current signal acquisition card USB 4AD Plus.

The vibration signal of the gearbox is obtained using the DFT5100 dynamic data collector from the acceleration sensor (ICP type) on the experimental platform (Figure 4b). The current signal is transmitted to the USB signal acquisition and recording platform through the signal acquisition card USB 4AD Plus (Figure 4c). In this paper, the rotation speed of the motor is set to 600 rpm; the sampling time is 10 s; and the sampling frequency of vibration and current is 1 kHz and 2 kHz, respectively. In the experiments, the temperature signal is easily affected by the operation time of the unit and the ambient temperature, and it cannot reflect the actual operating temperature of the wind turbine. Therefore, when fusing different signals by D–S evidence theory, we set the temperature signal to 0, regardless of its influence. Four groups for each working condition, with a total of 16 groups, are sampled on the platform. Some characteristic indexes of vibration and current signal are shown in Tables 15 and 16. The actual classification and diagnosis results of fusion signals and individual signals are shown in Figure 5. Table 17 is the calculation of two examples.

Table 15. Part of the characteristic index of the vibration signal.

Fault Type	Root Mean Square Value	Center of Gravity Frequency	IMF1 Energy Entropy	IMF2 Energy Entropy
Normal	0.0286	−118.3859	0.3671	0.0991
	0.0270	−184.0340	0.3461	0.1265
	0.0288	−308.9050	0.3678	0.1624
	0.0626	−993.2476	0.3524	0.3675
Parallel misalignment	0.0248	−272.4819	0.3678	0.1229
	0.0258	−196.5053	0.3677	0.1201
	0.0253	−286.5526	0.3668	0.1769
	0.0607	−1082.4788	0.3678	0.3455
Angular misalignment	0.0266	−166.9377	0.3488	0.1005
	0.0296	−145.9158	0.3658	0.1083
	0.0280	−232.8465	0.3620	0.1347
	0.0569	−1052.415	0.3583	0.3677
Integrated misalignment	0.0284	−261.2838	0.3544	0.1615
	0.0342	−334.0774	0.3621	0.2021
	0.0311	−388.9565	0.3675	0.2138
	0.0670	−1138.6520	0.3603	0.3670

Table 16. Partial characteristic index of current signal.

Fault Type	Root Mean Square Value	Center of Gravity Frequency	Energy Entropy1	Sample Entropy1
Normal	2.4944	−0.0571	0.0001	0.6859
	2.4952	−0.0741	0.0003	0.8285
	3.5641	−0.1023	0.0008	0.9624
Parallel misalignment	2.4948	−0.1238	0.0001	0.6099
	2.5293	−0.2608	0.0004	0.7046
	2.7607	−0.4788	0.0015	0.9455
Angular misalignment	2.4990	−0.1062	0.0002	1.0642
	2.6908	−0.2794	0.0008	1.1659
	3.0569	−0.4415	0.0016	1.3677
Integrated misalignment	2.5051	−0.0524	0.0003	1.2666
	2.8986	−0.3861	0.0009	1.6857
	3.2670	−0.6520	0.0023	1.7670

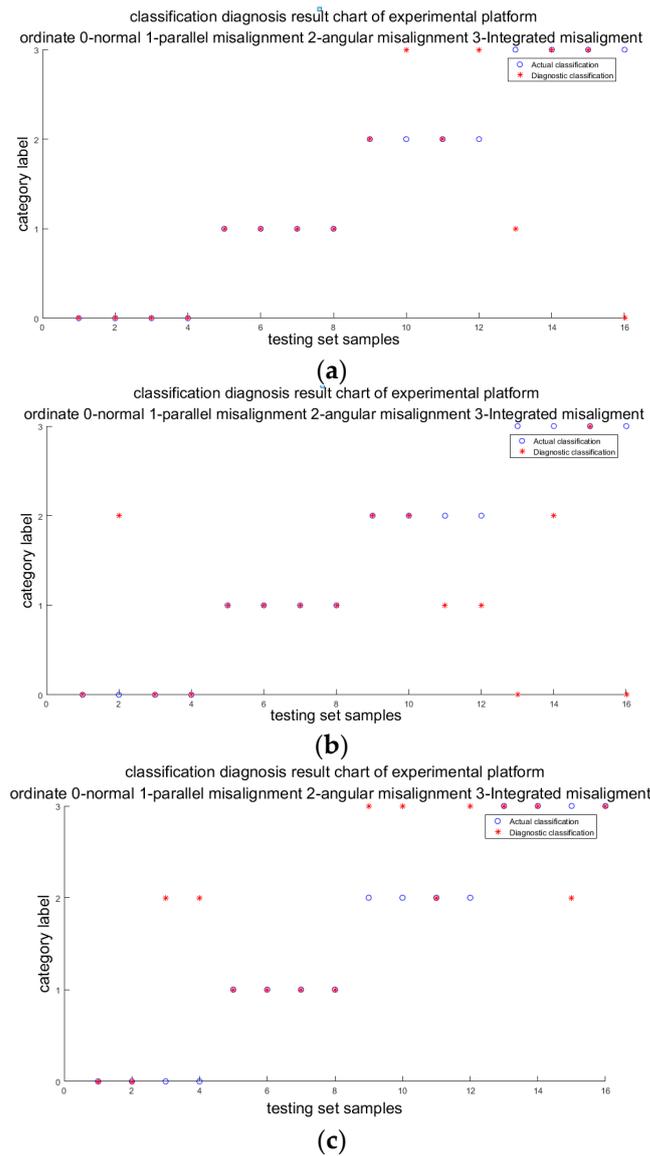


Figure 5. Diagnostic result. (a) Vibration signal + current signal; (b) vibration signal; (c) current signal.

Table 17. Basic probability assignment of two kinds signals and probability after fusion.

Probability Value	Normal	Parallel Misalignment	Angular Misalignment	Integrated Misalignment
Vibration signal	0.5161	0.1663	0.1510	0.1666
Current signal	0.1770	0.1927	0.4194	0.2109
Fusion signal	0.8209	0.0737	0.0181	0.0873
	0.7705	0.0566	0.1063	0.0666
	0.9348	0.0281	0.0050	0.0321
	0.6623	0.0530	0.2164	0.0683

It can be seen from Figure 5 that the classification accuracy of the testing set is 75%, while that of the single vibration signal is 62.5% and that of the single current signal is 62.5%, which indicates that the accuracy of the diagnosis is improved by using the D–S decision fusion method with multi-source signals as the diagnosis information. In addition, the reason the classification accuracy of the experimental results is much lower than that of the simulation results is that there is no temperature signal in the D–S evidence theory

fusion. It can be seen from Table 17 that the first sample is correctly identified using either the single signal or fusion signal, while the second sample is mistakenly diagnosed as angle misalignment using only the vibration signal, but is correctly identified by D–S fusion.

6. Conclusions

This paper proposes an integrated fault diagnosis method for wind turbine transmission system misalignment based on information decision fusion. The method uses multiple sources of signal including vibration signal, temperature signal, and stator current signal as the original source, and extracts different features from their time domain, frequency domain, and time–frequency domain. t-SNE is used to eliminate the correlation of characteristic values of the vibration signal and the stator current signal. Three posterior probability least squares support vector machines optimized using improved artificial bee colony algorithm are constructed respectively. The output probabilities of least squares support vector machines are used as the basic probability distribution of evidence fusion, and the fault diagnosis is completed by D–S synthesis and decision rules. Finally, the simulation experiments and platform verification show that the D–S evidence fusion model has higher diagnostic accuracy than the non-fusion model for the wind turbine misalignment fault.

Author Contributions: Conceptualization, Y.X. and Y.W.; software, Y.W.; validation, M.L.; writing—original draft preparation, Y.W. and Y.X.; writing—review and editing, L.Z. and J.X.; project administration, Y.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (51577008).

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

References

1. Junior, P.R.; Fischetti, E.; Araújo, V.G.; Peruchi, R.S.; Aquila, G.; Rocha, L.C.S.; Lacerda, L.S. Wind power economic feasibility under uncertainty and the application of ANN in sensitivity analysis. *Energies* **2019**, *12*, 2281. [CrossRef]
2. Liu, Z.; Zhang, L.; Carrasco, J. Vibration analysis for large-scale wind turbine blade bearing fault detection with an empirical wavelet thresholding method. *Renew. Energy* **2020**, *146*, 99–110. [CrossRef]
3. Hu, A.; Yan, X.; Xiang, L. A new wind turbine fault diagnosis method based on ensemble intrinsic time-scale decomposition and WPT-fractal dimension. *Renew. Energy* **2015**, *83*, 767–778. [CrossRef]
4. Zheng, H. Case analysis of wind turbine vibration monitoring. *Wind Energy* **2014**, *2017*, 88–92. (In Chinese) [CrossRef]
5. Liao, M.; Liang, Y.; Wang, S.; Wang, Y. Analysis of misalignment of wind turbines. *Mech. Sci. Technol. Aerosp. Eng.* **2011**, *2011*, 173–180. (In Chinese with English abstract) [CrossRef]
6. Liu, Z.; Zhang, L. A review of failure modes, condition monitoring and fault diagnosis methods for large-scale wind turbine bearings. *Measurement* **2020**, *149*, 107002. [CrossRef]
7. Simani, S.; Fantuzzi, C. Dynamic system identification and model-based fault diagnosis of an industrial gas turbine prototype. *Mechatronics* **2006**, *16*, 341–363. [CrossRef]
8. Tang, H.; Lam, K.-M.; Shum, K.-M.; Li, Y. Wake effect of a horizontal axis wind turbine on the performance of a downstream turbine. *Energies* **2019**, *12*, 2395. [CrossRef]
9. Zhao, Q.; Han, T.; Jiang, D.; Yin, K. Application of variational mode decomposition to feature isolation and diagnosis in a wind turbine. *J. Vib. Eng. Technol.* **2019**, *7*, 639–646. [CrossRef]
10. Abdalla, G.M.; Tian, X.; Zhen, D.; Gu, F.; Ball, A.; Chen, Z. Misalignment diagnosis of a planetary gearbox based on vibration analysis. In Proceedings of the 21st International Congress on Sound and Vibration, Beijing, China, 13–17 July 2014; Volume 4, pp. 2775–2783. Available online: <https://www.researchgate.net/publication/286770020> (accessed on 20 January 2021).
11. Huang, Q.; Jiang, D.X.; Hong, L.Y. Application of Hilbert-Huang transform method on fault diagnosis for wind turbine rotor. *Key Eng. Mater.* **2009**, *413*, 159–166. [CrossRef]
12. An, M.-S.; Kang, D.-S. Application of modified empirical mode decomposition method to fault diagnosis of offshore wind turbines. *Int. J. Multimed. Ubiquitous Eng.* **2016**, *11*, 67–80. [CrossRef]
13. Villa, L.F.; Reñones, A.; Perán, J.R.; De Miguel, L.J. Statistical fault diagnosis based on vibration analysis for gear test-bench under non-stationary conditions of speed and load. *Mech. Syst. Signal Process.* **2012**, *29*, 436–446. [CrossRef]
14. He, G.; Ding, K.; Li, L.; Deng, R. Vibration test and analysis of transmission chain of wind turbine based on double-elastic support. *J. South China Univ. Technol. (Nat. Sci. Ed.)* **2014**, *2014*, 90–97. (In Chinese with English abstract) [CrossRef]
15. Bossé, E.; Rogova, G.L. *Information Quality in Information Fusion and Decision Making*; Information Fusion and Data Science; Springer: Berlin/Heidelberg, Germany, 2019; pp. 1–49.

16. Mönks, U. *Information Fusion under Consideration of Conflicting Input Signals*; Technologien für die intelligente Automation; Springer: Berlin/Heidelberg, Germany, 2017; pp. 11–35.
17. Wald, L. Some terms of reference in data fusion. *IEEE Trans. Geosci. Remote Sens.* **1999**, *37*, 1190–1193. [[CrossRef](#)]
18. Zeng, Q.; Liu, X. Research on Bayesian classification algorithm. *Biotechnol. World* **2015**, *35*, 253–255. (In Chinese)
19. Dubois, D.; Liu, W.; Ma, J.; Prade, H. The basic principles of uncertain information fusion. An organised review of merging rules in different representation frameworks. *Inf. Fusion* **2016**, *32*, 12–39. [[CrossRef](#)]
20. Yao, P. *Research on Fuzzy Multiple Attribute Decision Making Method for Process Industry Fault Diagnosis*; Qilu University of Technology: Jinan, China, 2019. (In Chinese with English abstract)
21. Li, W.-W.; Wu, C. A multicriteria interval-valued intuitionistic fuzzy set topsis decision-making approach based on the improved score function. *J. Intell. Syst.* **2016**, *25*, 239–250. [[CrossRef](#)]
22. Su, C.-H.; Chen, T.-L.; Cheng, C.-H.; Chen, Y.-C. Forecasting the stock market with linguistic rules generated from the minimize entropy principle and the cumulative probability distribution approaches. *Entropy* **2010**, *12*, 2397–2417. [[CrossRef](#)]
23. Wang, M.; Zhang, Z.; Pei, P. Application of information fusion technology in fault diagnosis of large generators. *Comput. Simul.* **2012**, *29*, 349–352. (In Chinese with English abstract) [[CrossRef](#)]
24. Li, Y.; Xu, Y.; Chen, G.; Miao, R.; Yu, J. Improvement and application of D-S evidence theory in multi-sensor fault diagnosis. *J. Southeast Univ. (Nat. Sci. Ed.)* **2011**, *41*, 102–106. (In Chinese with English abstract) [[CrossRef](#)]
25. Jiang, W.; Wu, S. Multi-data fusion fault diagnosis method based on SVM and evidence theory. *Chin. J. Sci. Instrum.* **2010**, *31*, 1738–1743. (In Chinese with English abstract) [[CrossRef](#)]
26. Hu, J.; Yu, Z.; Zhai, X.; Peng, J.; Ren, L. Research on fault diagnosis and fusion diagnosis of aero-engine rotor based on improved D-S evidence theory. *Acta Aeronaut. Astronaut. Sin.* **2014**, *35*, 436–443. (In Chinese with English abstract)
27. Tian, Y.; Liu, S.; Jing, Y.; Yang, Y. Fault prediction of wind turbines based on D-S evidence fusion. *Comput. Mod.* **2017**, *2017*, 57–61. (In Chinese with English abstract) [[CrossRef](#)]
28. Kang, J.; Gu, Y.; Li, Y. Multi-sensor information fusion algorithm based on DS evidence theory. *J. Chin. Inert. Technol.* **2012**, *20*, 670–673. (In Chinese with English abstract) [[CrossRef](#)]
29. Zhou, H.; Li, S. The combination of support vector machine and evidence theory in information fusion. *J. Transduct. Technol.* **2008**, *21*, 1566–1570. (In Chinese with English abstract) [[CrossRef](#)]
30. Boudraa, A.O.; Bentabet, L.; Salzenstein, F.; Guillon, L. Dempster-Shafer’s basic probability assignment based on fuzzy membership functions. *ELCVIA* **2004**, *4*, 1–10. Available online: <https://www.researchgate.net/publication/39087167> (accessed on 20 February 2021). [[CrossRef](#)]
31. Haenni, R. Shedding new light on Zadeh’s criticism of Dempster’s rule of combination. In Proceedings of the 2005 7th International Conference on Information Fusion, Philadelphia, PA, USA, 25–28 July 2005; Volume 2, p. 33.
32. Li, W.; Guo, K. Combination rules of D-S evidence theory and conflict problem. *Syst. Eng. Theory Pract.* **2010**, *30*, 1422–1432. (In Chinese with English abstract)
33. Zhang, X.; Xiao, X.; Xu, G. Weighted posterior probability output for support vector machines. *J. Tsinghua Univ. (Sci. Tech.)* **2007**, *47*, 1689–1691. (In Chinese with English abstract) [[CrossRef](#)]
34. Wang, J.P.; De Lin, S.; Bao, Z.F. Neural network and D-S evidence theory based condition monitoring and fault diagnosis of drilling. *Appl. Mech. Mater.* **2012**, *249*, 481–486. [[CrossRef](#)]
35. Lin, H.-T.; Lin, C.-J.; Weng, R.C. A note on Platt’s probabilistic outputs for support vector machines. *Mach. Learn.* **2007**, *68*, 267–276. [[CrossRef](#)]
36. Zhao, F. Detection method of LSSVM network intrusion based on hybrid kernel function. *Mod. Electron. Technol.* **2015**, *38*, 97–99. (In Chinese with English abstract)
37. Yuan, X.; Wang, P.; Yuan, Y.; Huang, Y.; Zhang, X. A new quantum inspired chaotic artificial bee colony algorithm for optimal power flow problem. *Energy Convers. Manag.* **2015**, *100*, 1–9. [[CrossRef](#)]
38. Yi, Y.; He, R. A novel artificial bee colony algorithm. In Proceedings of the 2014 Sixth International Conference on Intelligent Human-Machine Systems and Cybernetics, Hangzhou, China, 26–27 August 2014; pp. 271–274.
39. Chen, W.; Xiao, Y. An improved ABC algorithm and its application in bearing fault diagnosis with EEMD. *Algorithms* **2019**, *12*, 72. [[CrossRef](#)]
40. Xiao, Y.; Kang, N.; Hong, Y.; Zhang, G. Misalignment fault diagnosis of DFWT based on IEMD energy entropy and PSO-SVM. *Entropy* **2017**, *19*, 6. [[CrossRef](#)]
41. Xiao, Y.; Hong, Y.; Chen, X.; Chen, W. The application of dual-tree complex wavelet transform (DTCWT) energy entropy in misalignment fault diagnosis of doubly-fed wind turbine (DFWT). *Entropy* **2017**, *19*, 587. [[CrossRef](#)]
42. Zhang, G. Thermal Characteristics Analysis of High Speed Transmission System of Wind Turbines. Master’s Thesis, Beijing Jiaotong University, Beijing, China, 2017. (In Chinese with English abstract)
43. Xiao, Y.; Wang, Y.; Ding, Z. The Application of heterogeneous information fusion in misalignment fault diagnosis of wind turbines. *Energies* **2018**, *11*, 1655. [[CrossRef](#)]
44. Shao, R.; Hu, W.; Wang, Y.; Qi, X. The fault feature extraction and classification of gear using principal component analysis and kernel principal component analysis based on the wavelet packet transform. *Measurement* **2014**, *54*, 118–132. [[CrossRef](#)]
45. Zhang, Y.; Li, Y. Fisher information metric based on stochastic neighbor embedding. *J. Beijing Univ. Technol.* **2016**, *42*, 863–869. (In Chinese with English abstract)