Wind Turbine Drivetrain Gearbox Fault Diagnosis Using Information Fusion on Vibration and Current Signals

Yayu Peng, *Student Member*, *IEEE*, Wei Qiao, *Fellow*, *IEEE*, Fangzhou Cheng, *Member*, *IEEE*, and Liyan Qu, *Senior Member*, *IEEE*

***Abstract*—To improve the reliability of the conventional vibration-based wind turbine drivetrain gearbox fault diagnosis system, this paper proposes a novel fault diagnosis method by fusing the information from gearbox vibration and generator current signals. First, the fault features contained in the gearbox vibration signals and the generator current signals are analyzed, respectively. Second, a multiclass support vector machine (SVM) model with probabilistic output is proposed to design two classifiers which output the probabilities of different gearbox fault types according to the input fault features extracted from the vibration signals and the current signals separately. Then, a non-trainable combiner and a trainable combiner are designed based on the Dempster–Shafer theory and the softmax regression technique, respectively, to fuse the information from the vibration and current SVM classifiers at decision level. The output of each combiner is the final diagnosis result. The proposed method is validated by experimental results obtained from a test gearbox with different types of faults. The validation results show that the proposed method can increase the fault diagnostic accuracy and is more robust than the conventional fault diagnosis systems that only use one type of signals.**

***Index Terms*—current signal, fault diagnosis, gearbox, information fusion, vibration signal, wind turbine**

I. Introduction

OPERATION and maintenance (O&M) cost constitutes approximately 20-25% of the levelized cost of energy (LCOE) of wind power assets. One effective way to reduce O&M cost to make wind power more competitive in the electricity market is to implement condition-based maintenance [1]. To achieve this, a condition monitoring system (CMS) is needed to online monitor the health conditions of wind turbine subassemblies and perform maintenance timely when a fault that causes an improper operation of the wind turbine is detected. Because gearbox failure causes the longest downtime of wind turbines among various subassembly failures and high maintenance costs, many modern utility-scale wind turbines are equipped with CMSs for drivetrain gearboxes. The majority of these CMSs use vibration signals. A disadvantage of vibration monitoring is the high costs of vibration sensors and associated data acquisition and processing devices [2]. Moreover, false and missing alarms of vibration-based wind turbine CMSs have been reported and some manufacturers have realized the limitation of using vibration signals alone. Furthermore, a problem commonly seen in wind turbine CMSs is sensor failure, which may cause false fault diagnosis results [3]. A feasible way to reduce missing and false alarms of the existing vibration-based CMSs and solve the problem of sensor failure in CMSs is to use additional source(s) of signal(s). For example, General Electric's commercial wind turbine CMS Bently Nevada ADAPT.Wind integrates oil particle sensors to provide a confirmation of suspected damage detected by the ADAPT.Wind vibration monitoring system [4].

To improve the fault diagnosis accuracy and reliability by effectively using the fault-related features extracted from multiple sensors, a variety of methods based on the information fusion technique have been developed. Information fusion is the process of integrating multiple information sources and can be conducted at different levels, such as signal level, feature level, or decision level. In the signal-level fusion, signals from different sensors are combined to create a new signal with a better signal-to-noise ratio than the original signals [5]. For example, [6] proposed a method to enhance the accuracy of sensor fault diagnosis using adaptive extended Kalman filter and signal-level fusion. In [7], a signal-level fusion method was proposed to obtain the health indices for degradation modeling. However, the physical meaning of the result obtained from fusing different types of signals is not clear and signal-level fusion is usually conducted for signals of the same type.

It is more common to conduct feature- and decision-level fusions in fault diagnosis. For example, the work in [8] fused information from different vibration sensors at decision level and achieved higher fault diagnostic accuracy than using the information from individual sensors separately. However, it still suffers from the drawbacks/limitations of vibration monitoring techniques and may fail if the vibration monitoring system fails. In [9], a bearing fault diagnosis system that uses both infrared thermal imaging data and vibration data was proposed. The system fused the features extracted from the two types of data, which were then used by a random forest classifier for fault diagnosis. The system outperformed single type of signal-based systems. However, it is costly to install infrared camera to obtain the thermal imaging data. The work

This work was supported in part by the U.S. National Science Foundation under Grant CMMI-1663562 and CAREER Award ECCS-1554497.

Yayu Peng, Wei Qiao, and Liyan Qu are with the Power and Energy Systems Laboratory, Department of Electrical and Computer Engineering, University of Nebraska-Lincoln, Lincoln, NE 68588-0511 USA (e-mail: yayu.peng@huskers.unl.edu; wqiao3@unl.edu; lqu2@unl.edu).

Fangzhou Cheng is with Amazon Web Services, East Palo Alto, CA 94303 USA (e-mail: chengfz73@gmail.com)

[10] performed fault diagnosis by fusing vibration and acoustic data and achieved higher accuracy than using vibration and acoustic data separately. However, these methods [8]-[10] require installation of additional sensors and data acquisition equipment and, therefore, lead to extra hardware cost and increased wiring complexity of the CMS.

It is reported in [11] that generator current signals contain health condition information of wind turbine drivetrain gearbox and, thus, have been used successfully for fault diagnosis of drivetrain gearbox. Since current signals have been used in generator/power converter control systems, the use of generator current signals for fault diagnosis of drivetrain gearbox does not need to install any additional sensors or data acquisition equipment. Therefore, fusing the information of vibration and current signals is a cost-effective way to improve the accuracy and reliability of the conventional vibration-based CMSs. Information fusion of vibration and current signals has been studied in electric machine fault diagnosis. For example, [12] used a convolutional neural network (CNN) for feature level fusion and [13] used the Dempster-Shafer theory for decision level fusion. However, it is more challenging to perform gearbox fault diagnosis because different gearbox faults commonly have similar features and little work on vibration and current information fusion for gearbox fault diagnosis has been reported. To the best of the authors' knowledge, [14] and [15] are the two most recent works on fusing current and vibration signals for gearbox fault diagnosis. In [14], the vibration and current signals were fused at the feature level by concatenation of the features from the two types of signals, which were used as the input to a deep neural network classifier for fault diagnosis. In [15], vibration, acoustic, current, and angular speed signals were concatenated and deep CNNs were used for fault diagnosis.

Regardless of what information fusion techniques are used, fault diagnosis is a classification problem that is commonly solved by using machine learning techniques. For example, support vector machines (SVMs) were designed in [16] for fault diagnosis of wind turbine drivetrain gearbox using features extracted from stator and rotor current signals of doubly-fed induction generator (DFIG). The work [17] also used an SVM and feature learning on a generator current signal for wind turbine gearbox fault diagnosis. The recent advances in deep learning have achieved remarkable improvements in speech recognition, visual object recognition, and many other domains [18]. The state-of-art deep learning techniques have also been applied in fault diagnosis. For example, deep belief networks and convolutional neural networks have been used for wind turbine gearbox fault diagnosis and achieved satisfactory accuracy [19], [20]. However, a large amount of training data is needed when using deep learning techniques [19]-[22]. Unlike computer vision and natural language processing tasks, the labeled data available for gearbox fault diagnosis is limited [23]. Compared to deep learning-based fault diagnosis methods [19]-[22], SVM is more suitable for the learning problem with a small dataset [24] and, thus, has been adopted for gearbox fault diagnosis using one type of signal(s) only [11], [25].

This paper proposes a wind turbine drivetrain gearbox fault diagnosis method based on the SVM and decision-level information fusion techniques. The proposed method contains a non-trainable combiner and a trainable combiner, which are designed to fuse the diagnosis results from two SVM classifiers that perform fault diagnosis for the gearbox separately using the fault features extracted from gearbox vibration signals and generator current signals, respectively. The features of gearbox faults in vibration and current signals, which are not discussed in recent works on deep learning-based fusion techniques [14], [15], are analyzed in this paper. The proposed method is validated by using the vibration and generator current data collected from the tests performed on a two-stage parallel-shaft gearbox test rig with different types of gear faults in comparison with the state-of-the-art SVM fault diagnosis methods that use vibration or current signal only or fuse the features extracted from the vibration and current signals.

The main contributions of this paper are three-fold. Firstly, this paper discussed why decision-level fusion is preferrable over feature-level fusion for the fault diagnosis using multiple sources of signals. Secondly, this paper developed a non-trainable method and a trainable method for decision-level fusion of vibration and current signals to improve the accuracy and reliability of the conventional vibration-based wind turbine gearbox fault diagnosis methods. Finally, this paper demonstrated the advantage of the non-trainable decision-level fusion method over the trainable decision-level fusion method and feature-level fusion method for the fault diagnosis problems with limited data for which the deep learning-based methods usually do not work.

The remainder of this paper is organized as follows. Section II presents the proposed decision-level vibration and current information fusion-based fault diagnosis method. Section III validates the effectiveness and superiority of the proposed method in comparison with the methods using current or vibration signal alone or fusing the features extracted from both vibration and current signals via experimental test results carried out on a gearbox test rig. Section IV concludes the paper. Section V discusses the recommendations for future research.

II. Proposed Fault Diagnosis Method Based on Vibration and Current Information Fusion

A. Framework of the Proposed Fault Diagnosis Method

In the conventional vibration-based gearbox fault diagnosis system shown in Fig. 1(a), the vibration signals are collected from sensors installed on the case of the gearbox. Fault features are extracted from the vibration signals. An SVM classifier is designed to identify the gearbox fault types according to the input fault features extracted from the vibration signals [25]. One major limitation of the conventional method is that if the vibration data is corrupted due to failure of sensors or data acquisition equipment, fault diagnosis may fail.

The accuracy, reliability, and robustness of the conventional fault diagnosis method could be enhanced by increasing the number of sensors. This, however, will increase hardware cost and wiring complexity of the system. The method proposed in this paper utilizes generator current signals, which are already available in the generator control system and, therefore, does not need installation of any additional hardware. The proposed method, illustrated in Fig. 1(b), consists of four functional modules. The first functional module is feature extraction, which extracts the fault features in the gearbox vibration signals and the generator current signals collected into the proposed

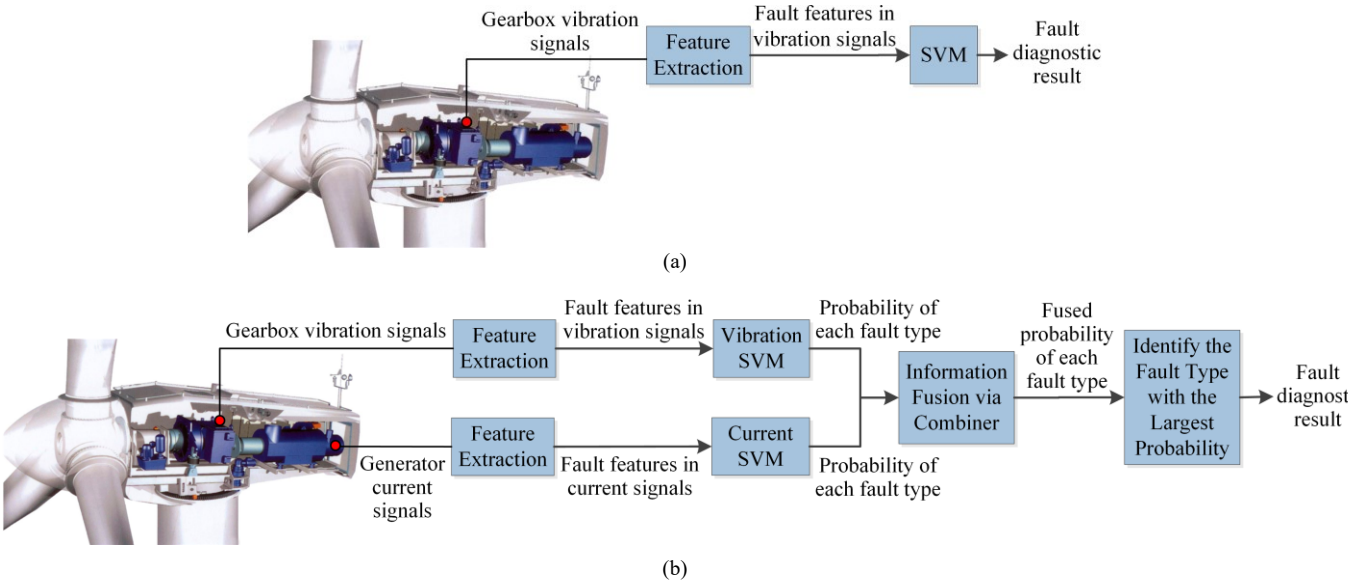


Fig. 1. Block diagram illustration of (a) the existing vibration-based fault diagnosis method and (b) the proposed fault diagnosis method based on information fusion of vibration and current signals for drivetrain gearboxes.

CMS separately. In the second functional module, two SVMs, a current SVM and a vibration SVM, are designed and trained to output the probability of each possible fault type according to the fault features extracted separately from the vibration signals and the current signals. The third functional module is information fusion, which uses a combiner designed to fuse the output information, i.e., the probabilities of possible fault types, from the current SVM and the vibration SVM. The last functional module diagnoses the fault to be the one with the largest probability among the output of the combiner. Compared with the conventional method shown in Fig. 1(a), the proposed method in Fig. 1(b) is expected to achieve higher fault diagnostic accuracy by incorporating the information from generator current signals. Moreover, the proposed method will be effective even when vibration sensor(s) or associated data acquisition equipment fails and, therefore, is expected to improve the reliability and robustness of the CMS.

The goal of this work is not only improving the accuracy but also improving the reliability of the fault diagnosis method so that it can still work properly when some sensor(s) fail or the signal(s) collected from some sensor(s) are corrupted/lost. In these circumstances, the fault diagnosis based on feature-level fusion may not work properly because the incorrect features extracted from the corrupted/lost signals are used for the entire decision, but the fault diagnosis based on decision-level fusion would still work properly because the incorrect features extracted from the corrupted/lost signals are only used in part of the decision by separate SVM classifiers, and the SVM classifier(s) using correct features can still output the correct fault diagnosis result. Therefore, decision-level fusion is adopted in this work.

B. Gearbox Fault Feature Extraction from Vibration and Current Signals

1) Fault Features in Vibration Signals

When a fault appears in a gear, it will alter the stiffness of the teeth or change their geometric parameters and, therefore, lead to changes in the vibration signal. These changes modulate the vibration signal $v(t)$, which can be expressed as [26]

$$v(t) = \sum_m V_m(1 + a_m(t))\cos(2\pi m f_m t + \theta_m + b_m(t)) \quad (1)$$

where m is the meshing harmonic number; f_m is the m^{th} tooth meshing harmonic frequency; V_m and θ_m are the amplitude and initial phase of the m^{th} meshing harmonic, respectively; $a_m(t)$ and $b_m(t)$ are the amplitude and phase modulation functions of the m^{th} meshing harmonic, respectively, expressed as follows.

$$a_m(t) = \sum_n A_{mn}\cos(2\pi n f_i t + \alpha_{mn}) \quad (2)$$

$$b_m(t) = \sum_n B_{mn}\cos(2\pi n f_i t + \beta_{mn}) \quad (3)$$

where f_i ($i = 1, \cdots, I$) is the i^{th} shaft rotational frequency and I is number of shafts of the gearbox; A_{mn} and B_{mn} are the amplitudes and α_{mn} and β_{mn} are the initial phases of the n^{th} harmonic of the amplitude and phase modulation functions of the m^{th} meshing harmonic, respectively.

Gearbox fault features contained in vibration signals can be extracted in time or frequency domain. The commonly used time-domain features include kurtosis and crest factor, which are used in this paper. Kurtosis K is a dimensionless parameter defined as follows.

$$K = \frac{M_4}{\sigma^4} \quad (4)$$

where M_4 is the 4^{th} central moment and σ is the standard deviation of the signal. Kurtosis characterizes the probability distribution of the signal. If the signal follows a normal distribution, its kurtosis is equal to 3 and a fault may lead to an increase in kurtosis [27].

Crest factor C is defined as the ratio of the maximum absolute value to the root mean square value of the signal.

$$C = \frac{|v_{peak}|}{v_{rms}} \quad (5)$$

Crest factor tells how "peaky" the signal is. The higher the crest factor, the peakier the signal.

In addition to kurtosis and crest factor, there are many other time domain features, such as shape factor, clearance factor, etc. [28]. An unhealthy condition of a gearbox usually can be distinguished from healthy condition by using time domain features of vibration signals. However, it is usually difficult to diagnose the fault type of the gearbox using time domain features. To achieve accurate fault diagnosis, the vibration signals need to be analyzed in frequency domain.

The dominant components in the frequency spectrum of a gearbox vibration signal are the gear meshing frequencies and their sidebands caused by modulation by shaft rotating frequencies. Typically, an increase in the number and the amplitudes of sidebands may indicate a fault condition [29]. Therefore, the energies at each gear meshing frequency f_m and its sidebands $f_m \pm f_i$ $(i = 1, \cdots, I)$ in the frequency spectrum of the vibration signal can be used as fault features.

2) Fault Features in Current Signals

Due to the electromechanical coupling between the gearbox and the generator in a wind turbine drivetrain, the generator current signals are modulated by gearbox vibrations. The relationship between gearbox vibration and generator current can be derived from torque and current relationships of electric machines, which has been discussed in detail in [30].

If the gearbox vibrates at a frequency f_i, the generator stator current signal will contain the fundamental component f and its sidebands at the frequencies $f \pm f_i$ whose amplitudes change when fault occurs. Therefore, the energy at $f \pm f_i$ can be selected as fault features in the current signal. Moreover, since the degradation of the gearbox will excite more noise in current signals [31], the noise-to-signal ratio (*NSR*) defined in (6) can also be selected as a fault feature of the current signal.

$$NSR = \frac{P_{noise}}{P_{signal}} \tag{6}$$

where P_{signal} is the energy of the fundamental frequency component of the current signal; $P_{noise} = P_{total} - P_{signal}$; and P_{total} is the total energy of the current signal.

3) Fault Features in Nonstationary Signals

In practice, wind turbines are subject to random fluctuations of speed and load, which make the vibration and current signals used for fault diagnosis nonstationary. This problem can be solved by using an appropriate signal conditioning method reported in the literature [32], [33] or a simple steady-state check which ensures that the signal is collected in steady state. Thus, the proposed information fusion-based fault diagnosis method can be easily extended to nonstationary scenarios.

C. Multiclass SVM Classifier with Probabilistic Output

The SVM is a technique commonly used to solve classification problems with small numbers of data samples. It was originally proposed to solve binary classification problems [34]. The goal of a binary SVM classifier is to find a hyper-plane that separates two classes with the maximal margin. However, gearbox fault type identification is a multiclass classification problem because there are more than two fault types. Suppose that there are total k ($k > 2$) possible gearbox fault types and define $\Theta = \{F_i \mid i = 1, 2, \cdots, k\}$ the set of all possible fault types. Then, gearbox fault diagnosis is a k-class classification problem. To solve such a problem, a multiclass SVM is designed by using $k(k-1)/2$ binary SVM classifiers and the one-against-one method [34]. Each binary SVM classifier is trained to classify two fault types by using two classes of data that characterize the two fault types, respectively. Then, the outputs of the $k(k-1)/2$ binary SVM classifiers are used by the one-against-one method to generate the final fault classification result for the k fault types.

The output of each binary SVM is the class label which represents one of two fault types. A limitation of this approach is that it maps the input fault features to the corresponding fault types deterministically but does not provide the probability of belonging that is needed for the probabilistic information fusion. Here probability is interpreted as a quantification of belief that a particular type of fault will occur, which is known as Bayesian probability [35], instead of the frequency of occurrence of the fault. To enable the probabilistic information fusion of the proposed fault diagnosis method, this paper proposes that each binary SVM outputs the probability of one of the two fault types it classifies using a sigmoid function as follows, instead of the deterministic class label of the fault type.

$$\begin{aligned} \mu_{ij} &= P\left(y = F_i \middle| y = F_i \text{ or } F_j, \boldsymbol{x}\right) \\ &\approx \frac{1}{1 + \exp\left(Gf(\boldsymbol{x}) + H\right)} \end{aligned} \tag{7}$$

where μ_{ij} denotes the probability of the fault type F_i when the SVM classifies the fault type F_i and F_j ($F_i \in \Theta$, $F_j \in \Theta$, and $F_i \neq F_j$); $\boldsymbol{x}$ is the vector of input features to the SVM; $f(\boldsymbol{x}) = \mathbf{w}^T \varphi(\boldsymbol{x}) + b$; $\mathbf{w}$ and b are the parameters of the SVM; and $\varphi(\cdot)$ is the kernel function of the SVM. The parameters G and H of (7) are obtained by minimizing the negative log likelihood of the training data.

Then, the probability of the input features belonging to each of the k fault types, p_i $(i = 1, \cdots, k)$, can be determined by solving the optimization problem (8) based on the pairwise coupling principle [36].

$$\begin{aligned} &\min_{\boldsymbol{p}} \ \frac{1}{2}\sum_{i=1}^{k}\sum_{j:j\neq i}(\mu_{ji}p_i - \mu_{ij}p_j)^2 \\ &s.t. \ \ p_i \geq 0, \forall i \\ &\sum_{i=1}^{k} p_i = 1 \end{aligned} \tag{8}$$

where $\boldsymbol{p} = [p_1, p_2, \cdots, p_k]^{\mathrm{T}}$. The solution of (8) is the probabilistic output of the multiclass SVM.

The proposed multiclass SVM model is applied to design the current SVM and the vibration SVM, shown in Fig. 1(b), of the proposed method.

D. Fusion of the SVM Outputs

The information fusion module of the proposed method fuses the probabilistic outputs of the vibration SVM and the current SVM, which are called base multiclass classifiers, by using a combiner. Depending on whether training is needed, there are two types of combiners: non-trainable and trainable combiners. No training is needed for the non-trainable combiner after the base classifiers are trained individually [37]. This paper designs both types of combiners for the information fusion and provides insights on the combiner selection.

1) Non-trainable Combiner

The non-trainable combiner uses a fixed combination rule, such as majority vote, to fuse the information from different base classifiers and usually assumes that all the base classifiers are equal in determining the final result. The lack of flexibility is the major limitation of the non-trainable combiner. However, it needs less training data than the trainable combiner. This paper uses the Dempster–Shafer theory [38] to design the non-trainable combiner. Let $m_{\text{vibration}}(F_i)$ and $m_{\text{current}}(F_i)$ be the basic probability assignment (BPA) of the fault type F_i ($F_i \in \Theta$) obtained from the vibration SVM and the current SVM, respectively. In other words, $m_{\text{vibration}}(F_i)$ and $m_{\text{current}}(F_i)$ are the probabilistic outputs on the fault type F_i of the vibration SVM and the current SVM, respectively, which are obtained by solving the optimization problem (8). Then, the BPA output on the fault type F_i of the non-trainable combiner, $m(F_i)$, is obtained by fusing the BPA outputs of the vibration and current SVMs using the Dempster's combination rule as follows [39].

$$m(F_i) = m_{\text{vibration}}(F_i) \oplus m_{\text{current}}(F_i) \tag{9}$$

where $\oplus$ denotes orthogonal sum.

After fusing the probabilistic outputs of the vibration and current SVMs, the final BPAs of all possible fault types Θ are obtained. The final diagnosis result is the fault type with the highest BPA, i.e., $\max\big(m(F_i)\big)$, $\forall\, F_i \subset \Theta$. Sometimes the fault diagnosis may fail. To indicate such cases, the probability of false fault diagnosis P_U is defined as follows.

$$P_U = 1 - \max\big(m(F_i)\big), i = 1, 2, \ldots, k \tag{10}$$

If P_U is greater than 0.5, it indicates that the fault diagnosis may fail and the diagnosis result will be set to Not Available (N/A).

The non-trainable combiner needs less training data than the trainable combiner. However, the lack of flexibility is the major limitation of the non-trainable combiner.

2) Trainable Combiner

Instead of using a fixed combination rule, the outputs of the base classifiers can be used as input features of another learning algorithm, which learns an aggregation function of the outputs of the base classifiers based on the training data [37]. Compared with the information fusion using a non-trainable combiner, this method is capable of extracting more information from the training data. This paper designs a simple trainable combiner based on the softmax regression method instead of other complicated trainable combiners used in [10], [14] due to the limited size of the training data.

The input of the proposed softmax regression-based trainable combiner is the combination of the concatenated outputs of the two base classifiers denoted as $\boldsymbol{s}_j = [\boldsymbol{p}_{\text{vibration}}, \boldsymbol{p}_{\text{current}}]_j$, where $\boldsymbol{p}_{\text{vibration}}$ and $\boldsymbol{p}_{\text{current}}$ are the probabilistic outputs of the vibration and current SVMs, respectively; j denotes the j^{th} data sample. Given the input $\boldsymbol{s}_j$, the softmax score for each fault type F_i ($F_i \in \Theta$) can be calculated to be $\boldsymbol{\theta}^{(i)^T}\boldsymbol{s}_j$ $(i = 1, \cdots, k)$, where $\boldsymbol{\theta}^{(i)}$ is the vector of parameters of the softmax regression combiner. Once the softmax score of every fault type is obtained, the probability of $\boldsymbol{s}_j$ belonging to each fault type F_i, $q_i(\boldsymbol{s}_j)$ $(i = 1, \cdots, k)$, can be calculated as follows [40].

$$\boldsymbol{q}(\boldsymbol{s}_j) = \begin{bmatrix} q_1(\boldsymbol{s}_j) \\ q_2(\boldsymbol{s}_j) \\ \ldots \\ q_k(\boldsymbol{s}_j) \end{bmatrix} = \frac{1}{\sum_{i=1}^{k} \exp\,(\boldsymbol{\theta}^{(i)^T}\boldsymbol{s}_j)} \begin{bmatrix} \exp\,(\boldsymbol{\theta}^{(1)^T}\boldsymbol{s}_j) \\ \exp\,(\boldsymbol{\theta}^{(2)^T}\boldsymbol{s}_j) \\ \ldots \\ \exp\,(\boldsymbol{\theta}^{(k)^T}\boldsymbol{s}_j) \end{bmatrix} \tag{11}$$

The parameter matrix of the softmax regression combiner, $\boldsymbol{\theta} = [\boldsymbol{\theta}^{(1)}, \cdots, \boldsymbol{\theta}^{(k)}]$, is determined by training. The goal of training the softmax regression combiner is to adjust the value of $\boldsymbol{\theta}$ to minimize the difference between the output of the combiner and the true probability distribution of different fault types obtained from the training dataset. In specific, the value of $\boldsymbol{\theta}$ is determined by training which minimizes the following cost function.

$$J(\boldsymbol{\theta}) = -\left[\sum_{j=1}^{N}\sum_{i=1}^{k} 1\{y^{(j)} = i\} \log\,(q_i(\boldsymbol{s}_j))\right] \tag{12}$$

where the superscript j denotes the j^{th} training sample; N is the total number of training samples; k is the total number of fault types; and $1\{y^{(j)} = i\}$ is the "indicator function," which is equal to 1 when $y^{(j)} = i$ is true; otherwise, it is equal to 0. Since the minimization problem cannot be solved analytically, the gradient descent algorithm expressed as follows can be used to find the optimal value of $\boldsymbol{\theta}$ that minimizes the cost function $J(\boldsymbol{\theta})$ of (12).

$$\boldsymbol{\theta}_{n+1} = \boldsymbol{\theta}_n - \gamma \nabla J(\boldsymbol{\theta}_n) \tag{13}$$

where $\boldsymbol{\theta}_n$ and $\boldsymbol{\theta}_{n+1}$ are the values of $\boldsymbol{\theta}$ in n^{th} and $(n + 1)^{\text{th}}$ iterations; γ is the learning rate; and $\nabla J(\boldsymbol{\theta}_n)$ is the gradient of $J(\boldsymbol{\theta}_n)$. The training stops when the maximum number of iterations is reached or the gradient is smaller than a predefined threshold.

In contrast to the non-trainable combiner, the parameters of the trainable combiner are learned from a training process through which the trainable combiner can learn the complicated relationship between the final diagnosis result and the diagnosis results from base classifiers. However, the training of a trainable combiner usually requires a large training dataset.

III. Experimental Validation

A. Experiment Setup

The gearboxes used in wind turbine drivetrains usually have multiple (typically three) stages and use parallel-shaft gears and planetary gears for different stages, respectively. For example, the drivetrain gearbox of the 1.6 MW wind turbines studied in [11] has three stages. The low-speed stage uses planetary gears but the intermediate-and high-speed stages uses parallel-shaft gears. However, the three-stage gearbox mixed with parallel-shaft and planetary gears is rarely available at small size for reduced-scale laboratory tests. Therefore, in laboratory tests, the three-stage wind turbine drivetrain gearbox is usually studied by using a one- or two-stage parallel-shaft gearbox and a planetary gearbox separately to validate fault diagnosis algorithms [17], [41]. In this paper, experimental studies were carried out on a two-stage parallel-shaft gearbox with three different fault types to validate the effectiveness of the proposed fault diagnosis method. Fig. 2 shows the experiment setup in which a 5-hp cage induction motor driven by a variable-frequency AC drive was used as a prime mover to drive a DFIG through two identical two-stage helical gearboxes connected back to back. One gearbox was employed to reduce

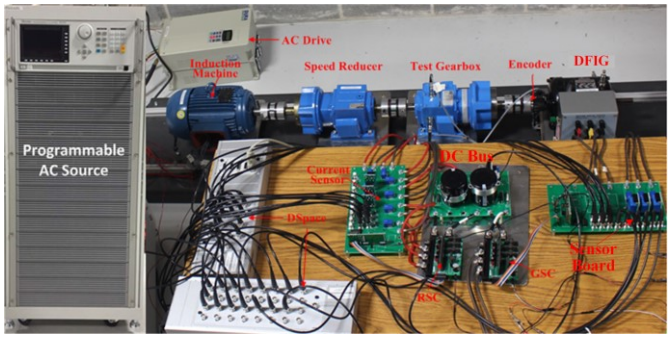


Fig. 2. Experiment setup.

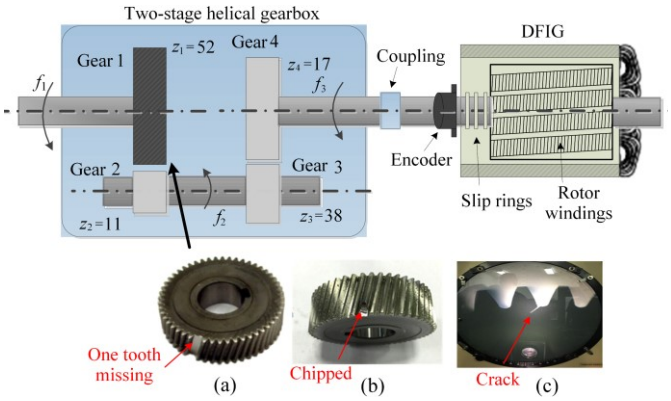


Fig. 3. Configuration of the test gearbox connected to a DFIG with (a) an OTM, (b) a chipped gear, and (c) a crack fault in the test gear.

the induction motor's shaft rotating speed while the other is the test gearbox employed to increase the shaft rotating speed. The stator windings of the DFIG were connected to a programmable three-phase AC source, which was used to emulate a power grid. The rotor of the DFIG was connected to the same power source through power electronics converters. A current signal and a vibration signal were acquired with a sampling frequency of 6 kHz for 100 seconds for each data record from the sensors installed on the DFIG stator and the gearbox case, respectively, by a National Instruments data acquisition system.

The configuration of test gearbox is shown in Fig. 3. It consists of four gears with different tooth numbers Z_1, Z_2, Z_3, and Z_4. The input shaft rotating frequency f_1, pinion shaft rotating frequency f_2, output shaft rotating frequency f_3, the mesh frequency f_{m1} are related to each other as follows.

$$f_1 = \frac{Z_2}{Z_1} f_2 = \frac{Z_2}{Z_1}\frac{Z_4}{Z_3} f_3, f_{m1} = Z_1 \cdot f_1 \quad (14)$$

According to the parameters of the gearbox, the characteristic frequencies are calculated using (14) and given in Table I.

The test gear was mounted on the input shaft of the gearbox. Four different gearbox health conditions were studied: 1) the gearbox is healthy (denoted as F_1) and the test gear has 2) a one-tooth-missing (OTM) fault (denoted as F_2), 3) a chipped gear fault (denoted as F_3), and 4) a crack fault (denoted as F_4), as shown in Fig. 3.

Table II lists the features used for fault diagnosis. For the gearbox vibration signal, the energies at f_3, $2f_3$, $3f_3$, f_{m1}, $f_{m1} \pm f_1$, and $f_{m1} \pm f_2$ of its spectrum as well as the crest factor and kurtosis are used as fault features. For the DFIG stator current signal, the energies at $f \pm f_3$ and $f \pm f_2$ as well as the *NSR* are used as fault features.

TABLE I
CHARACTERISTIC FREQUENCIES OF TEST GEARBOX AND DFIG

Test Gearbox (Hz)					DFIG (Hz)
Shaft			Gear meshing		Current fundamental
Input	Pinion	Output			
f_1	f_2	f_3	f_{m1}	f_{m2}	f
2.26	10.67	23.84	117.32	405.28	60

TABLE II
SUMMARY OF FAULT FEATURES

Vibration Signal	Current Signal
Crest factor, Kurtosis, and energies at f_3, $2f_3$, $3f_3$, f_{m1}, $f_{m1} \pm f_1$, and $f_{m1} \pm f_2$	NSR, and energies at $f \pm f_3$ and $f \pm f_2$

B. Experimental Results

The test gearbox with each of the four different gear health conditions was tested separately. For each health condition, 40 datasets of the gearbox vibration signal and the DFIG current signal were acquired, respectively. Among the 40 datasets 28 were used for training the two multiclass SVMs and the trainable combiner and the remaining 12 were used for testing the proposed method. The inputs of the current SVM and the vibration SVM are the fault features extracted from the gearbox vibration and DFIG stator current signals, respectively.

Fig. 4(a) shows the power spectral density (PSD) of the vibration signal when the gear is healthy. In this case, since the transmission is smooth, only the meshing frequency f_{m1} and output shaft rotating frequency f_3 are dominant. Fig. 4(b) shows the PSD of the vibration signal when the test gear is chipped. There is an increase in the energy at $f_{m1}$and f_3. Moreover, the sideband $f_{m1} + f_1$ and the 2nd and 3rd harmonics of f_3 become more noticeable. Fig. 4(c) shows the result in the case of the cracked gear. Compared to the healthy case, the energy at f_{m1} is higher than and the sideband $f_{m1} + f_1$ as well as the harmonic $2f_3$ are more noticeable. For the OTM case shown in Fig. 4(d), besides the increase of the energy at f_3 and $2f_3$, $f_{m1} - f_2$ and some unknown frequency components are excited, when compared with the healthy case. In summary, gear faults induce additional vibrations or change the amplitudes of existing vibrations. However, the differences between different fault types do not show clear patterns.

Fig. 5(a) shows the PSD of the current signal when the gear is healthy. The $f \pm f_3$ components are noticeable and the amplitude at $f - f_3$ increases in the chipped and cracked gear cases, as shown in Fig. 5(b) and (c), respectively. The $f - f_2$ component is also excited in the cracked gear case. A frequency component denoted as f_u is noticeable in Fig. 5(a)-(d), but is not related to any gearbox fault. Again, the differences between different gear fault types do not have clear patterns. Thus, the fault types cannot be identified directly from the PSD spectra of vibration or current signals.

To address this challenge, the proposed method is applied for diagnosis of the gear faults. The proposed method uses two multiclass SVMs with probabilistic outputs to automatically calculate the likelihood of each fault type based on the features extracted from the gearbox vibration signal and the DFIG stator current signal separately, and then fuses the probabilistic outputs of the two multiclass SVMs to obtain the final fault diagnosis result. The fault diagnostic accuracy of the proposed information fusion methods is compared to that of the vibration

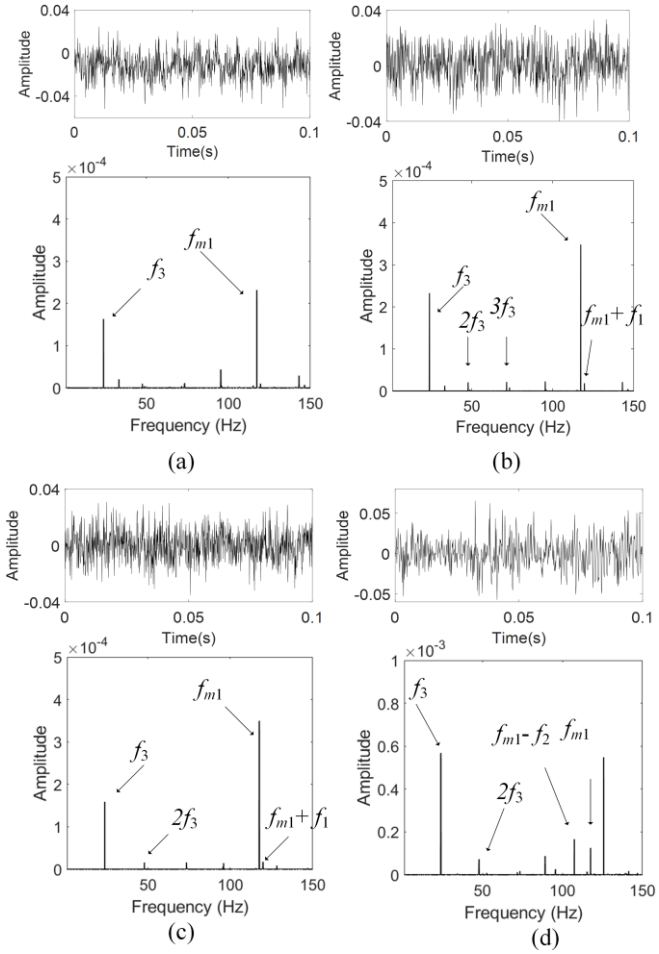


Fig. 4. Vibration signals and their PSD spectra for the test gearbox with (a) healthy gears, (b) a chipped gear, (c) a cracked gear, and (d) a gear with one-tooth missing.

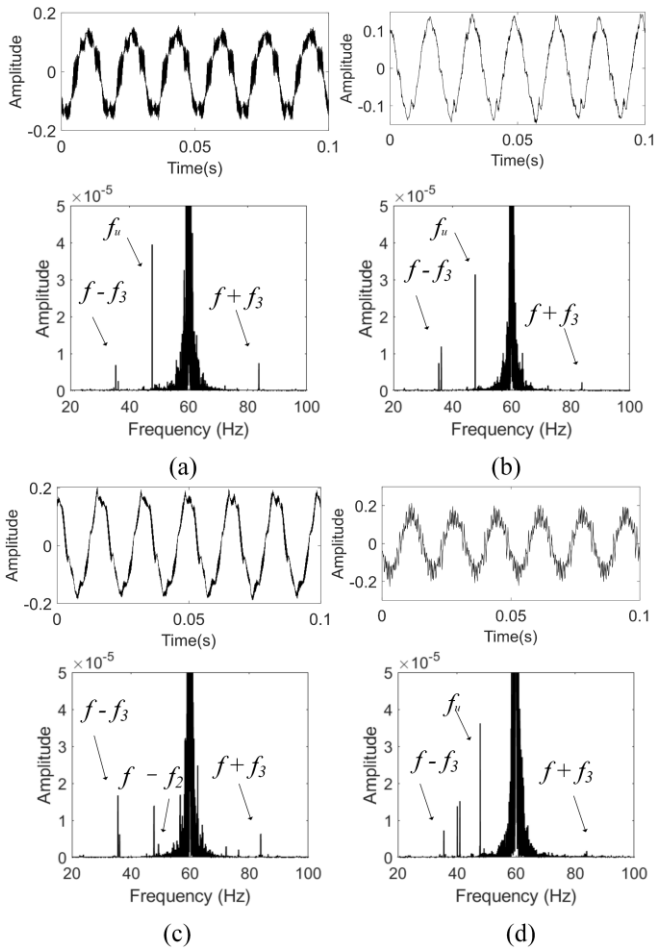


Fig. 5. DFIG current signals and their PSD spectra for the test gearbox with (a) healthy gears, (b) a chipped gear, (c) a cracked gear, and (d) a gear with one-tooth missing.

SVM, the current SVM, and a feature-level fusion method in Table III. In the feature-level fusion method, the features extracted from the vibration and current signals listed in Table II are used directly by a multiclass SVM classifier to output the probability of each fault type. There are 12 datasets used for testing in each health condition. Thus, there are totally 48 testing datasets. The fault diagnostic accuracy of the vibration SVM and the current SVM is 45/48 and 44/48, respectively; and the accuracy is increased to 47/48, 46/48, and 48/48 when using the feature-level fusion and the proposed trainable and non-trainable fusion methods, respectively. These results indicate that both the feature-level fusion and the proposed decision-level fusion-based fault diagnosis methods can improve the diagnostic accuracy and reduce the rate of false fault diagnosis compared to the methods using a single type of signal. The performance of the non-trainable fusion method is even better than that of the trainable fusion method and the feature-level fusion method in this experiment. The failure of the trainable fusion method in 2 of the 48 cases is likely due to insufficient training data for the softmax regression combiner.

According to the confusion matrix of the diagnosis results obtained from the five methods shown in Table IV, it is concluded that F_2, F_3, and F_4 are correctly diagnosed by all of the five methods. However, F_1 is only identified by the current SVM and the vibration SVM from 8 and 9 of the 12 datasets, respectively. Specifically, the current SVM identifies one F_1 dataset to be F_3, two F_1 datasets to be F_4, and one F_1 dataset to be N/A. The vibration SVM identifies one F_1 dataset to be F_3 and two F_1 datasets to be N/A. The feature-level fusion method successfully identified F_1 from 11 out of 12 datasets and, therefore, improved fault diagnosis accuracy. According to the results, there are four scenarios in the results of the decision-level fusion: 1) in 42 out of 48 cases, both the vibration and current signal-based fault diagnosis methods are correct; 2) in 2 out of 48 cases, the vibration signal-based fault diagnosis is wrong but the current signal-based fault diagnosis is correct; 3) in 3 out of 48 cases, the vibration signal-based fault diagnosis is correct but the current signal-based fault diagnosis is wrong; and 4) in 1 out of 48 cases, both the vibration and current signal-based fault diagnosis are wrong. By fusing the diagnosis results from the vibration and current SVMs at the decision level, the proposed method improved the accuracy and reliability of the fault diagnosis.

TABLE III
FAULT DIAGNOSTIC ACCURACY OF INDIVIDUAL MULTICLASS SVMS AND THREE INFORMATION FUSION METHODS.

	Vibration SVM	Current SVM	Feature-level fusion	Non-trainable fusion	Trainable fusion
Accuracy	45/48	44/48	47/48	48/48	46/48

Ten representative cases are studied to illustrate how the proposed method can increase the accuracy and reliability of

TABLE IV
CONFUSION MATRIX OF DIAGNOSIS RESULTS OBTAINED FROM CURRENT AND VIBRATION SVMS, FEATURE-LEVEL FUSION METHOD, AND PROPOSED TRAINABLE AND NON-TRAINABLE FUSION METHODS

Current/Vibration/ Feature-level fusion/ Trainable/Non-trainable		Diagnosed result				
		F_1	F_2	F_3	F_4	N/A
Actual condition	F_1	8/9/11/10/12	0	1/1/0/1/0	2/0/0/0/0	1/2/1/1/0
	F_2	0	12	0	0	0
	F_3	0	0	12	0	0
	F_4	0	0	0	12	0

the fault diagnosis and the results are shown in Table V. The diagnosis result is labelled as the fault type with the highest probability or N/A when the wrong classification probability P_U is larger than 0.5. For example, in Cases 1-4, the actual fault types are F_4, F_2, F_3, and F_1, respectively. They are representative cases of Scenario 1 in which both the vibration and current SVMs correctly diagnosed the fault type. In this scenario, the results from the feature-level fusion and both the non-trainable and trainable fusion methods are also correct but with a higher probability, indicating that it is of a higher confidence to determine the fault type of the gearbox.

Cases 5 and 6 are representative of Scenario 2 in which the vibration SVM failed to diagnose the fault type but the current SVM diagnosed the fault type correctly. In Case 5, the diagnosis result of the vibration SVM is F_3 and in Case 6, the diagnosis result of the vibration SVM is N/A because its P_U is greater than 0.5. However, the final diagnosis results in both cases are corrected by the decision-level fusion while the feature-level fusion only corrects the result in Case 6. This scenario is common when one sensor fails, and the combiner outputs the correct result by taking advantage of the information from the other signal.

Cases 7-9 are representative of Scenario 3 in which the current SVM failed to diagnose the fault type but the vibration SVM diagnosed the fault type correctly. By taking advantage of the information from the vibration signal, both the feature-level fusion and proposed non-trainable fusion methods obtain the correct results in these three cases; and the trainable fusion method obtains correct results in Cases 7 and 8. This demonstrates that the fault diagnosis accuracy and reliability are improved by the proposed method.

Case 10 is representative of Scenario 4 in which both the current and vibration SVMs failed to diagnose the fault type. This is the worst scenario in fault diagnosis. However, by taking advantage of the information from vibration and current signals, the feature-level fusion and non-trainable methods obtain the correct diagnosis result. This case demonstrates the effectiveness of the feature-level fusion and the robustness of the non-trainable method, which can provide the correct fault diagnosis result even when both SVMs output wrong results.

In summary, both the feature-level fusion and proposed decision-level fusion methods can improve the confidence of fault diagnosis when both SVM classifiers output the correct results. Moreover, the proposed non-trainable fusion method can correct wrong fault diagnosis results from one or even both SVM classifiers in all cases 5-10; whereas the trainable combiner can correct wrong fault diagnosis result from one SVM classifier in Cases 5-8 but fails in Cases 9 and 10 and the feature-level fusion fails in Case 5. Therefore, if the size of training data is small, the non-trainable combiner would be the best choice.

IV. CONCLUSION

This paper proposed a method for wind turbine drivetrain gearbox fault diagnosis by fusing the information from gearbox vibration and generator current signals. The features of gearbox faults in vibration and current signals were analyzed. A multiclass current SVM and a multiclass vibration SVM were designed and trained to output the probabilities of different gearbox fault types by using the features extracted from gearbox vibration signals and generator current signals separately. A non-trainable combiner based on the Dempster-Shafer theory and a trainable combiner based on the softmax regression technique were designed to fuse the probabilistic outputs of the two multiclass SVMs to generate a more accurate and more robust fault diagnosis result.

Experiments were conducted on a wind turbine drivetrain test rig to validate the effectiveness of the proposed gearbox fault diagnosis method. The experimental results validated that the proposed method, particularly the proposed decision-level non-trainable fusion method, achieved a more accurate and more reliable fault diagnosis result than the methods using individual type of signal as well as the feature-level information fusion method using both vibration and current signals. The trainable fusion method relies on the training of the combiner and, thus, is applicable when sufficient training data are available. If the size of the training dataset is small, the non-trainable combiner is a better option for information fusion. In contrast, the state-of-the-art deep-learning-based information fusion fault diagnosis methods [10], [14], [15] require much larger datasets for training and, therefore, are not suitable for many real-world applications with limited data of fault scenarios, such as the case studies in this paper, and have higher computational cost than the proposed method.

Since generator current signals are already used in the control systems of wind turbines, the proposed fault diagnosis method can be applied without additional hardware cost for wind turbines with vibration-based CMSs. In contrast, many other works using information fusion for fault diagnosis require installation of additional sensors, such as infrared thermal imaging camera [9] or acoustic sensors [10], leading to extra hardware cost.

V. RECOMMENDATIONS FOR FUTURE WORK

In the future work, a sensor sensitivity analysis can be conducted and only the data from the sensors that are highly sensitive to the faults of interest will be used for information fusion. Moreover, different faults usually have different frequencies of occurrence. The data from maintenance records provide prior probability distribution of different faults and this information can be used as a prior knowledge to increase the accuracy of fault diagnosis. Therefore, appropriate techniques can be developed to utilize the information from maintenance records to further increase the accuracy and reliability of the information fusion method. Finally, the proposed method can also be extended to include other types of signals to further increase the accuracy and robustness of wind turbine fault diagnosis.

TABLE V
DIAGNOSIS RESULTS OBTAINED FROM INDIVIDUAL MULTICLASS SVMS, FEATURE-LEVEL FUSION, AND PROPOSED TRAINABLE AND NON-TRAINABLE FUSION METHODS.

Case and approach		Probabilistic output of SVM and information fusion					Fault type	
		F_1	F_2	F_3	F_4	P_U	Diagnosed result	Actual condition
Case 1	Vibration SVM	0.0309	0.0207	0.0526	**0.8958**	0.1042	F_4	F_4
	Current SVM	0.0680	0.0452	0.0202	**0.8665**	0.1335	F_4	
	Feature-level fusion	0.0344	0.0189	0.0541	**0.8926**	0.1074	F_4	
	Non-trainable fusion	0.0027	0.0012	0.0014	**0.9947**	0.0053	F_4	
	Trainable fusion	6.77e-5	1.05e-4	6.30e-5	**0.9998**	0.0002	F_4	
Case 2	Vibration SVM	0.0173	**0.9654**	0.0131	0.0043	0.0346	F_2	F_2
	Current SVM	0.0263	**0.9509**	0.0179	0.0049	0.0494	F_2	
	Feature-level fusion	0.0325	**0.9549**	0.0092	0.0034	0.0451	F_2	
	Non-trainable fusion	4.94e-4	**0.9992**	2.55e-4	2.26e-5	0.0008	F_2	
	Trainable fusion	3.39e-6	**0.9999**	4.63e-6	5.30e-6	0.0001	F_2	
Case 3	Vibration SVM	0.0552	0.0262	**0.8947**	0.0239	0.1053	F_3	F_3
	Current SVM	0.0584	0.0303	**0.8824**	0.0289	0.1176	F_3	
	Feature-level fusion	0.0808	0.0121	**0.8780**	0.0290	0.1220	F_3	
	Non-trainable fusion	0.0041	9.97e-4	**0.9925**	6.77e-5	0.0075	F_3	
	Trainable fusion	1.16e-4	9.27e-5	**0.9997**	4.78e-5	0.0003	F_3	
Case 4	Vibration SVM	**0.8660**	0.0309	0.1012	0.0019	0.1340	F_1	F_1
	Current SVM	**0.8028**	0.1144	0.0520	0.0308	0.1972	F_1	
	Feature-level fusion	**0.9892**	0.0030	0.0036	0.0042	0.0108	F_1	
	Non-trainable fusion	**0.9874**	0.0050	0.0075	8.33e-5	0.0126	F_1	
	Trainable fusion	**0.9997**	1.47e-4	7.34e-5	5.1e-5	0.0003	F_1	
Case 5	Vibration SVM	0.1513	0.1602	**0.6659**	0.0226	0.3341	F_3	F_1
	Current SVM	**0.9673**	0.0031	0.0248	0.0048	0.0327	F_1	
	Feature-level fusion	0.2390	0.0483	**0.6805**	0.0322	0.3195	F_3	
	Non-trainable fusion	**0.8951**	0.0030	0.1012	0.0006	0.1049	F_1	
	Trainable fusion	**0.9976**	7.82e-4	0.0015	1.68e-4	0.0024	F_1	
Case 6	Vibration SVM	0.3400	0.0892	0.3632	0.2076	**0.6368**	N/A	F_1
	Current SVM	**0.9398**	0.0025	0.0565	0.0012	0.0602	F_1	
	Feature-level fusion	**0.6269**	0.0230	0.3039	0.0463	0.3731	F_1	
	Non-trainable fusion	**0.9384**	6.62e-4	0.0602	7.23e-4	0.0616	F_1	
	Trainable fusion	**0.9989**	3.04e-4	4.26e-4	4.05e-4	0.0011	F_1	
Case 7	Vibration SVM	**0.9850**	0.0018	0.0015	0.0017	0.0150	F_1	F_1
	Current SVM	0.3851	0.0946	0.0370	0.4833	**0.5167**	N/A	
	Feature-level fusion	**0.9696**	0.0071	0.0197	0.0036	0.0304	F_1	
	Non-trainable fusion	**0.9963**	4.45e-4	0.0011	0.0021	0.0037	F_1	
	Trainable fusion	**0.9874**	0.0021	0.0012	0.0093	0.0126	F_1	
Case 8	Vibration SVM	**0.9743**	0.0081	0.0152	0.0024	0.0257	F_1	F_1
	Current SVM	0.1644	0.0412	0.0135	**0.7809**	0.2191	F_4	
	Feature-level fusion	**0.9860**	0.0015	0.0100	0.0025	0.0140	F_1	
	Non-trainable fusion	**0.9850**	0.0020	0.0013	0.0117	0.0150	F_1	
	Trainable fusion	**0.7822**	0.0084	0.0058	0.2036	0.2172	F_1	
Case 9	Vibration SVM	**0.9936**	0.0014	0.0037	0.0013	0.0064	F_1	F_1
	Current SVM	0.0177	0.0637	**0.9171**	0.0015	0.0829	F_3	
	Feature-level fusion	**0.7383**	0.0201	0.1368	0.1047	0.2617	F_1	
	Non-trainable fusion	**0.8360**	0.0043	0.1595	9.61e-5	0.1640	F_1	
	Trainable fusion	0.0472	0.0014	**0.9505**	9.42e-4	0.0495	F_3	
Case 10	Vibration SVM	0.2635	0.2646	0.4602	0.0117	**0.5398**	N/A	F_1
	Current SVM	0.2793	0.0545	0.0015	**0.6646**	0.3354	F_4	
	Feature-level fusion	**0.9346**	0.0103	0.0483	0.0069	0.0654	F_1	
	Non-trainable fusion	**0.7624**	0.1494	0.0073	0.0809	0.2376	F_1	
	Trainable fusion	0.3958	0.1747	0.07	0.3595	**0.6042**	N/A	

REFERENCES

[1] M. Yildirim, N. Z. Gebraeel, and X. A. Sun, "Integrated predictive analytics and optimization for opportunistic maintenance and operations in wind farms," *IEEE Trans. Power Syst*., vol. 32, no. 6, pp. 4319-4328, Nov. 2017.

[2] W. Yang, P. J. Tavner, C. J. Crabtree, and M. Wilkinson, "Cost-effective condition monitoring for wind turbines," *IEEE Trans. Ind. Electron*., vol. 57, no. 1, pp. 263-271, Jan. 2010.

[3] A. G. Kavaz and B. Barutcu, ''Fault detection of wind turbine sensors using artificial neural networks,'' *J. Sensors*, vol. 2018, pp. 1–11, Dec. 2018.

[4] *Bently Nevada ADAPT.Wind condition monitoring solution*, [online]. Available: https://www.bakerhughesds.com/sites/g/files/cozyhq596/files/2020-04/GEA18079F%20Adapt%20Wind%20Brochure_R1.pdf

[5] J. Dong, D. Zhuang, Y. Huang, and J. Fu, "Advances in multi-sensor data fusion: Algorithms and applications," *Sensors*, vol. 9, no. 10, pp. 7771–7784, Sep. 2009.

[6] K. Salahshoor, M. Mosallaei, and M. Bayat, "Centralized and decentralized process and sensor fault monitoring using data fusion based on adaptive extended Kalman filter algorithm," *Measurement*, vol. 41, pp. 1059–1076, Mar. 2008.

[7] C. Song, K. Liu, and X. Zhang, "Integration of data-level fusion model and kernel methods for degradation modeling and prognostic analysis," *IEEE Trans. Rel*., vol. 67, no. 2, pp. 640–650, Jun. 2018.

[8] J. Hang, J. Zhang, and M. Cheng, "Fault diagnosis of wind turbine based on multisensors information fusion technology," *IET Renewable Power Generation*, vol. 8, no. 3, pp. 289-298, Apr. 2014.

[9] O. Janssens, M. Loccufier, and S. V. Hoecke, "Thermal imaging and vibration-based multisensor fault detection for rotating machinery," *IEEE Trans. Ind. Informat*., vol. 15, no. 1, pp. 434-444, Jan. 2019.

[10] M. Ma, C. Sun, and X. Chen, "Deep coupling autoencoder for fault diagnosis with multimodal sensory data," *IEEE Trans. Ind. Informat*., vol. 14, no. 3, pp. 1137-1145, Mar. 2018.

[11] F. Cheng, L. Qu, W. Qiao, C. Wei, and L. Hao, "Fault diagnosis of wind turbine gearboxes based on DFIG stator current envelope analysis," *IEEE Trans. Sustain. Energy*, vol. 10, no. 3, pp. 1044-1053, Jul. 2019.

[12] J. Wang, P. Fu, L. Zhang, R. X. Gao, and R. Zhao, "Multilevel information fusion for induction motor fault diagnosis," *IEEE/ASME Trans. Mechatronics*, vol. 24, no. 5, pp. 2139-2150, Oct. 2019.

[13] L. Hou and N. W. Bergmann, "Novel industrial wireless sensor networks for machine condition monitoring and fault diagnosis," *IEEE Trans. Instrum. Meas*., vol. 61, no. 10, pp. 2787-2798, Oct. 2012.

[14] G. Jiang, J. Zhao, C. Jia, Q. He, P. Xie, and Z. Meng, "Intelligent fault diagnosis of gearbox based on vibration and current signals: a multimodal deep learning approach," in *Proc. Prognostics and System Health Management Conference*, Oct. 2019, pp. 1-6.

[15] L. Jing, T. Wang, and M. Zhao, "An adaptive multi-sensor data fusion method based on deep convolutional neural networks for fault diagnosis of planetary gearbox," *Sensors*, vol. 17, no. 2, pp. 1–15, Feb. 2017.

[16] F. Cheng, Y. Peng, L. Qu, and W. Qiao, "Current-based fault detection and identification for wind turbine drivetrain gearboxes," *IEEE Trans. Ind. Appl.*, vol. 53, no. 2, pp. 878-887, Mar.-Apr. 2017.

[17] Q. He, J. Zhao, G. Jiang, and P. Xie, "An unsupervised multiview sparse filtering approach for current-based wind turbine gearbox fault diagnosis," *IEEE Trans. Instrum. Meas*., vol. 69, no. 8, pp. 5569-5578, Aug. 2020.

[18] L. Yann, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436-444, May 2015.

[19] Y. Qin, X. Wang, and J. Zou, "The optimized deep belief networks with improved logistic sigmoid units and their application in fault diagnosis for planetary gearboxes of wind turbines," *IEEE Trans. Ind. Electron.*, vol. 66, no. 5, pp. 3814-3824, May 2019.

[20] G. Jiang, H. He, J. Yan, and P. Xie, "Multiscale convolutional neural networks for fault diagnosis of wind turbine gearbox," *IEEE Trans. Ind. Electron.*, vol. 66, no. 4, pp. 3196-3207, Apr. 2019.

[21] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. Gao, "Deep learning and its applications to machine health monitoring," *Mechanical Systems and Signal Processing*, vol. 115, pp. 213-237, Jan. 2019.

[22] Z. Pu, C. Li, S. Zhang, and Y. Bai, "Fault diagnosis for wind turbine gearboxes by using deep enhanced fusion network," *IEEE Trans. Instrum. Meas*., vol. 70, pp. 1-11, 2021.

[23] N. Huang, Q. Chen, G. Cai, D. Xu, L. Zhang, and W. Zhao, "Fault diagnosis of bearing in wind turbine gearbox under actual operating conditions driven by limited data with noise labels," *IEEE Trans. Instrum. Meas*., vol. 70, pp. 1-10, 2021.

[24] R. Gholami and N. Fakhari, "Support Vector Machine: Principles, Parameters, and Applications," in *Handbook of Neural Computation*: Elsevier, 2017, pp. 533.

[25] J. Kang, X. Zhang, J. Zhao, H. Teng, and D. Cao, "Gearbox fault diagnosis method based on wavelet packet analysis and support vector machine," in *Proc. IEEE Prognostics and System Health Management Conference*, May 2012, pp. 1-13.

[26] W. Wang and A. K. Wong, "Autoregressive model-based gear fault diagnosis," *J. Vibration Acoustics*, vol. 124, no. 2, pp. 172-179, Apr. 2002.

[27] A. Brandt, *Noise and Vibration Analysis: Signal Analysis and Experimental Procedures*. New York, NY, USA: John Wiley&Sons, 2011.

[28] V. Sharma and A. Parey, "A review of gear fault diagnosis using various condition monitoring indicators," *Procedia Engineering*, vol. 144, pp. 253-263, 2016.

[29] D. Zappala, P. J. Tavner, C. J. Crabree, and S. Sheng, "Side-band algorithm for automatic wind turbine gearbox fault detection and diagnosis," *IET Renewable Power Generation*, vol. 8, no. 4, pp. 380-389, May 2014.

[30] X. Jin, F. Cheng, Y. Peng, W. Qiao, and L. Qu, "Drivetrain gearbox fault diagnosis: vibration- and current-based approaches," *IEEE Ind. Appl. Magzine*, vol. 24, no. 6, pp. 56-66, Nov.-Dec. 2018.

[31] Y. Peng, F. Cheng, W. Qiao, and L. Qu, "Fault prognosis of driventrain gearbox based on a recurrent neural network," in *Proc. IEEE International Conference on Eletro Information Technology*, May 2017.

[32] R. Maheswari and R. Umamaheswari, "Trends in non-stationary signal processing techniques applied to vibration analysis of wind turbine drive train-a contemporary survey," *Mechanical Systems and Signal Processing*, vol. 85, pp. 296-311, Feb. 2017.

[33] J. Wang, F. Cheng, W. Qiao, and L. Qu, "Multiscale filtering reconstruction for wind turbine gearbox fault diagnosis under varying-speed and noisy conditions," *IEEE Trans. Ind. Electron.*, vol. 65, no. 5, pp 4268-4278, May 2018.

[34] C. Hsu, C. Chang, and C. Lin, "A practical guide to support vector classification," Technical Report, Dept. of Computer Science and Information Engineering, National Taiwan Univeristy, 2003.

[35] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006, pp. 21.

[36] T. Wu, C. Lin, and R. C. Weng, "Probability estimates for multi-class classification by pairwise coupling," *The Journal of Machine Learning Research*, vol. 5, pp. 975-1005, Dec. 2004.

[37] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*, Hoboken, NJ, USA: Wiley&Sons, 2004

[38] G. Shafer, *A Mathematical Theory of Evidence*, Princeton, NJ, USA: Princeton University Press, 1976.

[39] X. Fan and M. Zuo, "Fault diagnosis of machines based on D-S evidence theory, Part 1. D-S evidence theory and its improvement," *Pattern Recognition Letters*, vol. 27, no. 5, pp. 366-376, 2006.

[40] A. Ng, J. Ngiam, C. Y. Foo, Y. Mai, and C. Suen, "UFLDL tutorial," [online]. Available: http://deeplearning.stanford.edu/tutorial/supervised/SoftmaxRegression/

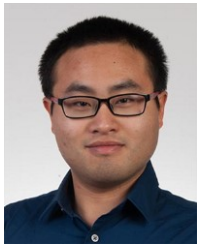
[41] L. Lu, Y. He, Y. Ruan, and W. Yuan, "Wind turbine planetary gearbox condition monitoring method based on wireless sensor and deep learning approach," *IEEE Trans. Instrum. Meas*., vol. 70, pp. 1-16, 2021.

Yayu Peng (S'13) received the B.Eng. degree from Chongqing University, Chongqing, China, in 2013, and the Ph.D. degree from the University of Nebraska-Lincoln, Lincoln, NE, USA, in 2021, both in electrical engineering.

He was a research and development intern at the Global Energy Interconnection Research Institute North America (GEIRINA) in 2018 and the New York Power Authority (NYPA) in 2020. His research interests include condition-based maintenance, intelligent fault diagnosis and prognosis, and the application of machine learning in power and energy systems.

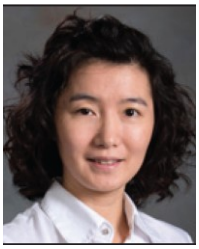
Wei Qiao (S'05–M'08–SM'12–F'20) received the B.Eng. and M.Eng. degrees in electrical engineering from Zhejiang University, Hangzhou, China, in 1997 and 2002, respectively, the M.S. degree in high-performance computation for engineered systems from Singapore-MIT Alliance, Singapore, in 2003, and the Ph.D. degree in electrical engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2008.

Since August 2008, he has been with the University of Nebraska–Lincoln, Lincoln, NE, USA, where he is currently a Professor with the Department of Electrical and Computer Engineering. His research interests include renewable energy systems, smart grids, condition monitoring, power electronics, electric motor drives, energy storage systems, and emerging electrical energy conversion devices. He is the author or coauthor of more than 260 papers in refereed journals and conference proceedings and holds 11 U.S. patents issued. Dr. Qiao was a recipient of the 2010 U.S. National Science Foundation CAREER Award and the recipient of the 2010 IEEE Industry Applications Society Andrew W. Smith Outstanding Young Member Award.



Fangzhou Cheng (S'15–M'18) received the B.Eng. degree in electrical engineering from Zhejiang University, Hangzhou, China, in 2013, and the Ph.D. degree in electrical engineering from the University of Nebraska–Lincoln, Lincoln, NE, USA, in 2017.

From 2018-2019, he was a Data Scientist with the System Sciences Laboratory, Palo Alto Research Center (PARC), CA, USA. He joined Amazon Web Service (AWS), CA, USA, in August 2019 as a Data Scientist. He has been working on different industrial projects in AWS. His research interests include anomaly detection, machine learning, condition-based maintenance, and intelligent fault diagnosis and prognosis for various cyber-physical systems.



Liyan Qu (S'05–M'08–SM'17) received the B.Eng. (with the highest distinction) and M.Eng. degrees in electrical engineering from Zhejiang University, Hangzhou, China, in 1999 and 2002, respectively, and the Ph.D. degree in electrical engineering from the University of Illinois at Urbana–Champaign, Champaign, IL, USA, in 2007.

From 2007 to 2009, she was an Application Engineer with Ansoft Corporation, Irvine, CA, USA. Since January 2010, she has been with the University of Nebraska−Lincoln, Lincoln, NE, USA, where she is currently an Associate Professor with the Department of Electrical and Computer Engineering. Her research interests include energy efficiency, renewable energy, numerical analysis and computer aided design of electric machinery and power electronic devices, dynamics and control of electric machinery, and magnetic devices. Dr. Qu was a recipient of the 2016 U.S. National Science Foundation CAREER Award.