A feature extraction and machine learning framework for bearing fault diagnosis

Bodi Cui a,1, Yang Weng b,*, Ning Zhang c,*

Abstract

Wind power generation has been widely adopted due to its renewable nature and decreasing capital cost per kW. However, existing equipment ages rapidly, leading to higher failure rates, greater operation and maintenance costs, and worsening safety conditions, calling for improved condition monitoring and fault diagnosis for wind turbines. Past methods utilize physical models, but they are only successful in laboratory environments. As increasing data are becoming available, there are methods applying machine learning without careful discrimination, leading to low accuracy. To solve this problem, first this paper proposes to conduct unsupervised learning to understand data properties, e.g., structural density. Subsequently, the sensitivity analysis is conducted to extract the significant features and to avoid overfitting. The sensitivity of various features that are characteristics of wind turbine bearings may vary significantly under different working conditions. During such a process, the piece-wise properties are studied to improve supervised learning. By combining the properties of data and regression, a three-stage learning algorithm is proposed to refine and learn the most useful information for turbine bearing fault diagnosis. The proposed framework is validated by using real data from diversified data sets for nonstationary vibration signals of bearings.

© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Wind energy has been widely recognized as one of the most efficient, economical, and environmentally friendly distributed energy resources. For such reasons, installed wind power capacity has increased substantially over the last decade. The worldwide wind power capacity reached 650 GW by the end of 2019. China alone has installed more than 200 GW and will double that number by 2030 [1]. Though wind energy is not currently the largest energy source, it has the largest number of generators in the power system since the capacity of each wind generator is much smaller than that of thermal and hydroelectric generators [2]. With continuous installation, aging wind generators present increasing challenges for sustainable operation and economical maintenance (O&M) of existing installations. Wind generators are mounted on top of high towers with enclosed shells, thus preventing manual monitoring and diagnosis. Therefore, different from solar panels, the higher uncertainty with respect to the rotating mechanism destabilizes revenues, creating a large O&M cost. For example [3], shows a potential 30% reduction in O&M cost in offshore environments if better diagnosis algorithms can be deployed. This potential leads to a strong interest in optimizing the O&M strategies in both industry and academia.

O&M costs are more critical for wind turbines (WTs) than other generators because the corresponding malfunction is much more complex [4]. The reasons are as follows: 1) The turbines continuously operate seven days per week, 24 h per day; 2) There are many rotating machines that cannot possibly operate maintenance-free, even with a perfect design; comparatively, solar panels include no rotating components and can reach a maintenance-free condition if carefully designed; and 3) Wind turbines are directly exposed to the environment, where the wind speed is variable, thus introducing variable mechanical loads, which add greater stress to the mechanical parts. Such a dilemma leads to the fact that some WTs can fail, sometimes frequently.
To reduce the failure rate, it is crucial to apply reliable and cost-effective condition monitoring (CM) techniques so that the best maintenance intervention can be determined in time to allow the machine to return to work as early as possible [5,6]. Additionally, improved CM leads to a better understanding of the design shortcomings, which can be used to improve WT design [7]. For this purpose, CM systems are increasingly installed. Past research shows great potential for economic benefit. Such gains are proven to be substantial and largely dependent on the fault detection rate [8,9]. Therefore, CM systems grew from 25% in 2012 to 36.1% in 2015. The key purpose of CM is to monitor and identify bearing failures on wind turbines. For example, the gearbox and generator are studied due to the high percentage of replacement over the wind farm life and to considerable equipment downtime, leading to substantial production loss. Recent studies show that bearings cause approximately 70% of gearbox downtime and 21%–70% of generator downtime [40].

In addition, corrective or preventive maintenance is inefficient to avoid these failure rates. Corrective maintenance is only performed after the malfunction has occurred, so it would not be able to avoid any failures. Preventive maintenance is time-based way that faces two main challenges: under- and over-maintenance. In other words, the challenges are that the system performance is not effectively monitored, resulting in unexpected failures, or there is an excessive number of maintenance activities, resulting in waste of resources. Therefore, it is crucial to use condition-based maintenance on bearings and to develop methods that can be used to diagnose faults early. In general, bearing diagnosis is a well-recognized field of research; however, this is not the case for machine operating under nonstationary loads, such as wind turbines [11,12]. In the case of varying load/speed, vibrational signals generated by rolling element bearings are affected by operational factors, making the diagnosis difficult [13–15]. These difficulties arise from the variation in vibration-based diagnostic features caused mostly by load/speed variation (operation factors), low energy of sought-after features, and low signal-to-noise levels [16,17]. Analysis of the signal from the main bearing is even more difficult due to the very low rotational speed of the main shaft.

In recent years, machine learning (ML) approaches, such as support vector machine (SVM), artificial neural networks (ANN), and Deep Neural Networks (DNN), have been widely employed to resolve the complex nonlinear and uncertainty problems across various disciplines. Successful applications of ML dealing with wind energy industry can be found in many recent studies [18–20]. These researches can be divided into four general categories: prediction, design optimization, optimal control, and fault detection [21]. The important parameters prediction of wind power generation based on ML, such as wind speed and wind power, has received extensive research [22–24]. Plenty of research has been done on deploying ML to improve the design of wind energy systems, such as wind turbine design and wind farm design [25]. At the same time, ML technologies are showing substantial strategies in the field of control, and are shown in related research on max power tracking control [26], pitch angle control [27], speed control [28] and so on during the process of wind power generation. Research on fault diagnosis of wind energy using ML has also received continuous and widespread attention. Leahy et al. diagnosed and predicted the wind turbine generator faults from SCADA data using support vector machines [29]. High recall and low precision were also found for the diagnostic and prognostic cases. Tang et al. built a classifier to identify gear, bearings, shaft and general transmission failures based on manifold learning and Shannon Wavelet Support Vector Machines [30]. A standard SVM with a radial basis function kernel obtained an accuracy of 76%. Jiang et al. proposed the multiscale convolutional neural networks for fault diagnosis of wind turbine gearbox [31]. Several studies were proposed to detect the wind turbine blades faults (such as cracks, erosion, and pitch angle twists) using ML [32].

For this work, the focus is to study the fault diagnosis of wind turbine bearing in wind energy. Schlechtigen et al. predicted the bearing failure of wind turbine generators by comparative analysis of neural network and regression using the data of bearing temperature [33]. They found that the nonlinear NN approaches outperform the regression models. Bangalore et al. used an artificial neural network approach to predict the early fault in the bearing of a planetary gearbox in wind turbine by bearing temperature measurement [34]. They conducted case studies in which their method predicts a gearbox bearing fault a week before it was indicated by the vibration based condition monitoring system. Bach-Andersen et al. use convolutional neural networks to classify expert-labeled vibration data containing main bearing faults (rotor, planetary stage, and helical stage bearing faults) [35]. They find that the deep CNN approach achieves a better performance on the test data than a logistic classifier or a shallow multilayer perceptron.

In this work, an approach is proposed to diagnose the bearings used in wind turbines based on customized ML framework. For such a design, an observation is firstly done that the wind is characterized by its speed and direction, which are affected by several factors, e.g., geographic location, climate characteristics, height above ground, and surface topography. For example, wind turbines interact with the wind, capturing a portion of its kinetic energy and converting it into usable energy. Therefore, the modeling is customized with these factors for knowledge enhancement. In the process of fault diagnosis, the selection of characteristic parameters significantly influences the diagnostic accuracy. Thus, a systematic method is presented to assess the prediction accuracy of each feature to find the most effective combination for fault diagnosis of wind turbine bearings based on neighborhood component analysis (NCA). Various statistical features are used in this work from notable category differences in time-domain, frequency-domain, and time-frequency. For this purpose, the hybrid-domain feature parameters defined on multiple domains are introduced to describe the state of the rolling bearing. Furthermore, the fault diagnosis of wind turbine bearings is performed based on time-frequency analysis and a supervised learning method.

To demonstrate the performance, the proposed method is validated on different data sets and conditions. The experimental results prove that the proposed method can effectively extract useful features accurately and efficiently for nonstationary vibrational bearing signals. Further, the accuracy of fault diagnosis is significantly higher than that of past methods.

The remaining paper is organized as follows. Section II analyzes the impacts of working conditions on wind turbine bearings. Section III performs the data-driven feature extraction and selection based on neighborhood component analysis. Section IV develops a bearing fault diagnosis method using ML. Section V conducts experiments for numerical validation. Section VI compares the performance between the proposed method and regression-based method. Section VII concludes the paper.

2. Impacts of working conditions on wind turbine bearings

Wind turbines (WTs) are exposed to a variable environment and operate 24 h a day, 7 days a week and experience rapid changes in temperature, air pressure, wind shear, wind speed, and total load. Due to these variations, the bearings of wind turbines undergo continuously changing loads. Therefore, the characteristics of the working conditions of wind turbine bearings are first studied to understand the causes of bearing faults. The observations will subsequently be incorporated into our feature pool.
2.1. Effect of nonstationary mechanical load

Mechanical loads cause fatigue damage to several parts of wind turbine bearings, thereby reducing the useful life of the system. There are two types of mechanical loads: static and dynamic. Static loads result from mean wind speed and determine the absolute mechanical load level. Dynamic loads, induced by turbulence and gusts, are more related to the fatigue damage.

1) Mean wind speed: The mean wind speed determines the basic working load of a wind turbine. This speed is quantified by the hourly mean of wind speeds. It is also used to determine the economic viability of a wind energy project. The probability distribution of the mean wind speed is predicted from measurements collected during several years. The experimentally obtained wind distribution can be approximated by a Weibull distribution given by

$$p(V_m) = \frac{k}{C} \left(\frac{V_m}{C}\right)^{k-1} e^{-\left(\frac{V_m}{C}\right)^k},$$

(1)

where $k$ and $C$ are the shape and scale coefficients, respectively. These coefficients are adjusted to match the wind data at a particular site. The Weibull probability function reveals that large wind speeds rarely occur, whereas moderate winds are more frequent.

2) Turbulence: Turbulence includes all wind speed fluctuations with frequencies above the spectral gap. Therefore, turbulence contains all components in the range from seconds to minutes. In general, turbulence exerts a minor impact on the annual energy capture. However, it has major impacts on aerodynamic loads and power quality.

2.2. Effect of wind farm location

Wind field characteristics exert important impacts on the type of wind turbine bearing failure. Wind farms are primarily divided into offshore wind farms and onshore wind farms from the wind field point of view. Terrain, even in the absence of obstacles, produces friction forces that delay the winds in the lower layers, which produces wind shear. This effect is more important at lower height levels and when the terrain is complex. The wind shear makes the wind turbine generator unbalanced and causes the transmission system to be more easily damaged. Comparatively, the problem is not that obvious for offshore wind farms since the sea level is generally flat and the wind encounters much less resistance. Furthermore, the frequency of the wind change on the sea is lower than that on the land, and therefore, the mechanical loads of offshore wind farms are more stable.

The shear of wind speed is a function of height. Different mathematical models have been proposed to describe wind shear, and one model is the Prandtl logarithmic law \[36\].

$$\frac{V_m(z)}{V_m(z_{ref})} = \frac{\ln(z/z_0)}{\ln(z_{ref}/z_0)},$$

(2)

where $z$ is the height above ground level, $z_{ref}$ is the reference height (usually 10 m) and $z_0$ is the roughness length. Therefore, the discussion in this section is used to build the feature pool for feature selection in the next section.

2.3. Types and impacts of the bearing faults

According to the characteristics of vibration signals during operation, the faults of rolling bearings can be mainly divided into two categories: surface damage faults and wear faults.

1) Surface damage faults: During the operation of the bearing, the rolling ball, the outer race, and the inner race, they are all subjected to the action of periodic pulsating loads, resulting in cyclically changing contact stress. When the number of stress cycles reaches a certain value, fatigue damage occurs on the rolling ball or the working surfaces of the inner and outer race. This phenomenon caused by impact load and alternating stress, also known as bearing fatigue failure, is the main cause of rolling bearing failure. When bearing fatigue failure occurs, significant shock load, vibration, and noise will be generated.

For bearings in wind turbines, this type of fatigue failure is further aggravated due to the non-linear and uncertain non-stationary loads, especially sudden shock loads. When such failures are not detected at an early stage, the surface damage will further increase, resulting in complete bearing invalid and even equipment accidents. Therefore, it is important to diagnose the fatigue failure of rolling bearings.

2) Wear faults: In the normal use of the bearing, the friction between the bearing balls, the outer and the inner race will lead to the occurrence of wear failures. This kind of failure usually takes a long time to occur because it is a gradual process. In general, wear faults will not cause bearing damage immediately, and the degree of harm is far less than surface damage faults.

3. Feature parameter selection for improving the diagnosis accuracy

Feature selection is a critical component in data workflow because high-dimensional data can downgrade the model performance [37]. For example, the training time can increase exponentially with numerous features, which also causes overfitting problems. Feature selection methods help with these problems by reducing the dimensions without much loss of the total information. This approach also helps to make sense of the features and their importance. Feature selection methods aid in creating an accurate predictive model by choosing features that will yield good or better accuracy while requiring fewer data [38].

Feature selection is the process of selecting a subset of relevant features for use in model construction. Given a set of features

$$x = \{x_1, \ldots, x_N\},$$

(3)

the feature selection problem is to find a subset of

$$x' \subseteq x$$

(4)

that maximizes the classification accuracy on an unseen test set, where

$$x' = \{x_i, \ldots, x_{i_0}\}, N \geq M.$$  

(5)

Usually, the $x'$ maximize some score functions. It should be noted that feature selection is different from dimensionality reduction. Both methods seek to reduce the number of attributes in the data set, but a dimensionality reduction method does so by creating new combinations of attributes, whereas feature selection methods include and exclude attributes present in the data without changing them.
3.1. Feature parameters used to describe the status of bearings

Statistical algorithms are used to analyze the time signals of vibration sensors. The employed approach involves calculating statistical values for a trend analysis and learning more about the shape of the vibration signal. It is assumed that vibration signals taken from intact bearings are normally distributed since the sonic pulses will be generated by the rolling elements and surfaces in a random way. If a fault starts to develop on the bearing, sonic pulses with higher energy (spikes) can be found in the vibration signals. The presence of those spikes can be identified with the statistical analysis algorithms as described below. Therefore, these statistical algorithms provide an absolute measure of the condition of the respective bearing; there is no training phase required to learn the baseline values, which correspond to a fault-free condition of the component. The disadvantages of these algorithms include the relatively poor selectivity, which means that a change in the value does not indicate a certain faulty component of a bearing, for example, the inner ring surface.

The various statistical characteristics are used in this study from notable category differences in the time-domain, frequency-domain, and time-frequency features. For example, the hybrid-domain feature parameters defined in multiple domains are introduced to describe the state of the rolling bearing.

1) **Time-domain feature parameters:** The crest factor has a value of approximately 3 for an ideal normally distributed signal. The reason, the maximum value in a normally distributed signal (the so-called white noise) is nearly $3\sigma$, whereas the Root Mean Square (RMS) is $1\sigma$. If a fault is developing in a bearing, the spike amplitudes, and therefore the maximum value of the signal, increase. In an early occurrence of the faulty condition, the RMS value will not be affected since the spike energy is small relative to the overall energy of the signal. Therefore, the RMS value does not change significantly, and the crest factor increases. The fault detection and generation of condition monitoring system (CMS) alarms are performed by trend analysis of the crest factor. Although the calculation of the statistical values of vibrational signals are independent of the measurement time, an appropriate time window should be measured, for example, 1 s. In any case, the measurement time should be a multiple of the grid frequency (20 ms in Europe) because this will help eliminate signal distortions related to grid EMC effects.

2) **Frequency-domain feature parameters:** Energy values of the intrinsic mode function (IMF) components are obtained from vibration acceleration signals of the bearing data. To effectively assess different fault locations and different degrees of performance degradation of a rolling bearing with a unified assessment index, a state assessment method based on the relative compensation distance of multiple feature domains and locally linear embedding is adopted [39]. The time-domain, frequency-domain, and time-frequency features are listed in Table 1, and Fig. 1 is a visualization of four parameters in the table.

<table>
<thead>
<tr>
<th>Feature parameters</th>
<th>Frequency-domain feature parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_{\text{max}} \equiv \max(x_i)$</td>
<td>$F_C = \frac{\sum_{k=1}^{N} f(k)}{\sum_{k=1}^{N} s(k)}$</td>
</tr>
<tr>
<td>$x_{\text{pp}} = \max(x_i) - \min(x_i)$</td>
<td>$p_5 = \frac{\sum_{k=1}^{N} f(k)}{\sum_{k=1}^{N} s(k)}$</td>
</tr>
<tr>
<td>$x_{\text{rms}} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} x_i^2}$</td>
<td>$F_{\text{RMS}} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} s(i)^2}$</td>
</tr>
<tr>
<td>$r = \frac{1}{N} \sum_{i=1}^{N} x_i$</td>
<td>$p_{10} = \frac{p_{6}}{p_{5}}$</td>
</tr>
<tr>
<td>$S_f = \frac{x_{\text{rms}}}{\sqrt{\sum_{i=1}^{N} x_i^2}}$</td>
<td>$p_1 = \frac{\sum_{k=1}^{N} s(k)}{N}$</td>
</tr>
<tr>
<td>$I_f = \frac{x_{\text{max}}}{\sqrt{\sum_{i=1}^{N} x_i^2}}$</td>
<td>$p_2 = \frac{\sum_{k=1}^{N} s(k) - p_1}{N - 1}$</td>
</tr>
<tr>
<td>$C_L = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2$</td>
<td>$p_{11} = \frac{\sum_{k=1}^{N} f(k)}{p_{5} \sqrt{p_{2}}}$</td>
</tr>
<tr>
<td>$G_k = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^4$</td>
<td>$p_{12} = \frac{\sum_{k=1}^{N} f(k) - p_{5}}{p_{2}}$</td>
</tr>
</tbody>
</table>

where $x_i$ is a signal series for $n = 1, 2, \ldots, N$. $N$ is the number of data points.

![Fig. 1. Feature weights for classification.](image)

Consider a multiclass classification problem with a training set containing $n$ observations

$$S = \{ (x_i, y_i), i = 1, 2, \ldots, n \},$$

where $x_i \in \mathbb{R}^p$ are the feature vectors, $y_i \in \{1, 2, \ldots, c\}$ are the class labels, and $c$ is the number of classes. The aim is to learn a classifier $f : \mathbb{R}^p \rightarrow \{1, 2, \ldots, c\}$ that accepts a feature vector and makes a prediction $f(x)$ of the true label $y$.

Consider a randomized classifier that
randomly picks a point, \( \text{Ref}(x) \), from \( S \), as the 'reference point' for \( x \) and
labels \( x \) using the label of the reference point \( \text{Ref}(x) \).

This scheme is similar to that of a 1-nearest neighbor (NN) classifier, where the reference point is chosen to be the nearest
neighbor of the new point \( x \). In NCA, the reference point is chosen randomly, and all points in \( S \) have some probability of being
selected as the reference point. The probability \( P(\text{Ref}(x) = x_j|S) \) that
point \( x_j \) is picked from \( S \) as the reference point for \( x \) is higher if \( x_j \) is
closer to \( x \) as measured by the distance function \( d_w \), where
\[
d_w(x_i, x_j) = \sum_{r=1}^{p} w_r^2 |x_i - x_j|, \tag{7}
\]
and \( w_r \) are the feature weights. Assume that
\[
P(\text{Ref}(x) = x_j|S) = k(d_w(x_i, x_j)) \tag{8}
\]
where \( k \) is some kernel or a similarity function that assumes large
values when \( d_w(x_i, x_j) \) is small. Let
\[
k(z) = \exp \left( -\frac{z}{\sigma} \right). \tag{9}
\]
The reference point for \( x \) is chosen from \( S \), so the sum of
\( P(\text{Ref}(x) = x_j|S) \) for all \( j \) must be equal to 1 [40]. Therefore, it is
possible to write
\[
P(\text{Ref}(x) = x_j|S) = \frac{k(d_w(x_i, x_j))}{\sum_{j=1}^{n} k(d_w(x_i, x_j))}. \tag{10}
\]
Now, consider the leave-one-out application of this randomized
classifier; that is, predicting the label of \( x \) using the data in \( S^{-1} \),
with the training set \( S \) excluding the point \( (x_i, y_i) \). The probability that
point \( x_j \) is picked as the reference point for \( x_i \) is
\[
\pi_j = P(\text{Ref}(x_i) = x_j|S^{-1}) = \frac{k(d_w(x_i, x_j))}{\sum_{j=1}^{n} k(d_w(x_i, x_j))}. \tag{11}
\]
The average leave-one-out probability of a correct classification is
the probability \( \pi \) that the randomized classifier correctly classifies
observation \( i \) using \( S^{-1} \):
\[
\pi_i = \frac{1}{n} \sum_{j=1}^{n} \pi_j P(\text{Ref}(x_i) = x_j|S^{-1}) I(y_i = y_j) \tag{12}
\]
\[
= \frac{1}{n} \sum_{j=1}^{n} \pi_j y_j. \tag{13}
\]
where
\[
y_i = I(y_i = y_j) = \begin{cases} 1 & \text{if } y_i = y_j, \\ 0 & \text{otherwise.} \end{cases} \tag{14}
\]
The average leave-one-out probability of a correct classification using
the randomized classifier can be written as
\[
F(w) = \frac{1}{n} \sum_{i=1}^{n} \pi_i. \tag{15}
\]
The right-hand side of \( F(w) \) depends on the weight vector \( w \). The
goal of neighborhood component analysis is to maximize \( F(w) \) with
respect to \( w \).
\[
F(w) = \frac{1}{n} \sum_{i=1}^{n} \pi_i - \lambda \sum_{r=1}^{p} w_r^2 \tag{16}
\]
where \( \lambda \) is the regularization parameter. The regularization term
drives many of the weights in \( w \) to 0.

After choosing the kernel parameter \( \sigma \) in \( p_{ij} \) as 1, the weight
vector \( w \) is found via the minimization problem for given \( \lambda \),
\[
w = \arg \min_{w} \left\{ \frac{1}{n} \sum_{i=1}^{n} \pi_j y_j \right\} \tag{17}
\]
where
\[
f_i(w) = -F(w) \tag{18}
\]
\[
f_i(w) = -f_i(w) \tag{19}
\]
Note that \( \frac{1}{n} \sum_{i=1}^{n} \pi_j y_j \) is the probability that the randomized
classifier does not change if a constant is added to an objective
function. Therefore, the objective function is rewritten by adding
the constant 1.
\[
w = \arg \min_{w} \left\{ 1 + f_i(w) \right\} \tag{20}
\]
where
\[
l(y_i, y_j) = \begin{cases} 1 & \text{if } y_i = y_j, \\ 0 & \text{otherwise.} \end{cases} \tag{21}
\]
The argument of the minimum is the weight vector that
minimizes the classification error.

4. Bearing fault diagnosis using ML

4.1. ML methods

ML is the process of building an inductive model that learns
from a limited amount of data without specialist intervention.
This learning implies finding an underlying set of structures (or patterns)
are useful to understand relationships in data that might not
be exactly similar to that on which learning occurred. In general, ML
is divided into two categories: supervised learning and unsuper-
vised learning. The supervised learning predicts an output variable
using labeled input data, while unsupervised learning draws in-
fences from data without labeled inputs (those done by clustering
algorithms, recommender systems etc.). For supervised learning, it
can be distinguished between models that predict a numeric var-
iable (regression) or a categorical variable (classifiers). Learning
in models translates into fitting a model's parameters to a specific
data set, iteratively updating them with several passes through the
data until a specific predefined function is minimized.

In the field of wind power, both supervised learning and un-
supervised learning have been successfully applied to fault
detection [33–35,41,42]. Helbing et al. reviewed a selection of studies using ANN and Deep Learning since 2009 [18]. They found that supervised approaches might often outperform unsupervised approaches at fault detection. Furthermore, supervised learning enables an added benefit at fault diagnosis due to the labeled input data. Therefore, supervised ML algorithms including support vector machine, Naive Bayes, k-nearest neighbor, and artificial neural network were adopted to perform the bearing fault diagnosis in the following section.

4.2. Support vector machine (SVM)

SVM is a computational learning method for small sample classification. Algorithmically, SVM builds an optimal separating hyperplane \( f(x) = 0 \) between data sets by solving a constrained quadratic optimization problem based on structural risk minimization (SRM) [43,44].

\[
y(x) = W^T x + b = \sum_{i=1}^{N} W(x_i) + b.
\] (22)

where \( W \) is an \( N \)-dimensional vector and \( b \) is a scalar. The optimal separating hyperplane is the separating hyperplane that creates the maximum distance between the plane and the nearest data. By converting the optimization problem with the Kuhn-Tucker condition into the equivalent Lagrangian dual quadratic optimization problem, the classifier based on the support vector can be obtained.

4.3. Naive Bayes classifier

Naive Bayes is a generative model with high learning and predicting efficiency, which is easy to implement [45]. This method attempts to maximize the posterior

\[
P(Y|X) \sim P(Y)P(X|Y).
\] (23)

Following the conditional independence assumption, it can be expressed as

\[
P(Y|X) = P(X_1X_2 \ldots X_n|Y).
\] (24)

Or, equivalently

\[
P(Y|X) = P(X_1|Y)P(X_2|Y) \cdots P(X_n|Y).
\] (25)

The parameters \( P(Y = y_k) (1 \leq k \leq K) \) can be obtained, where \( K \) is the total number of labels. and \( P(X_i = x_i|Y = y_k) \) \( x_i \) is a value of random variable \( X_i \) via maximum likelihood estimation (MLE) on the data set. When a new sample \( \{x_1, x_2, \ldots, x_n\} \) arrives, its label can be predicted as

\[
\hat{y} = \arg \max_{k \in [1,2,\ldots,K]} P(Y = y_k) \prod_{i=1}^{n} P(X_i = x_i|Y = y_k).
\] (26)

4.4. K-nearest neighbors (KNN)

KNN is an instance-based ML method. For a testing data set, the KNN calculates the distance between the testing sample and all the training samples. Then, the \( K \) nearest neighbors are counted and vote for the label of the testing data [46]. For example, the Euclidean distance is used to calculate sample distance for KNN as follows:

\[
d(x_i, x_j) = \sqrt{\sum_{k=1}^{K}(x_i(k) - x_j(k))^2}.
\] (27)

4.5. Artificial neural networks (ANN)

As the most popular ML methods, Artificial Neural Networks (ANN) simulates the function of biological nervous systems, where a certain number of simple elements called neurons constitute the ANN and operate in parallel. This kind of algorithm performs noticeably well for regression and classification by building the relationship between inputs and outputs throughout adjusting the weights of the connections between its elements and biases.

A two-layer feed-forward network with sigmoid hidden and softmax output neurons is employed to classify the bearing condition in this study. The network is trained with scaled conjugate gradient backpropagation learning algorithm. When a training loop begins, the data are transformed through the network layers and map the input values to the output values, which is called the forward propagation of signal. Then the loss function evaluates the deviations between the target and output values and dynamically tune the weight matrix in each hidden layer to produce better outputs. Then the classification is performed by the output neurons function softmax.

4.6. Fault diagnosis process based on feature extraction and ML

The flow chart of the fault diagnosis process is shown in Fig. 2, where the feature extraction/selection process is integrated in the upper figure and the ML part in the lower figure. Together, an accurate bearing fault diagnosis is achieved.

5. Experiments and discussion

5.1. Experimental data

Numerical validations are extensively conducted with similar results. For example, one of the data sets on a rolling bearing was obtained from the Bearing Data Center of Case Western Reserve University (https://engineering.case.edu/bearingdatacenter) [47]. The bearing type is an SKF6205 deep-groove ball bearing. Acceleration sensors were arranged on the bearing seat at the driving end and fan end of the motor. The vibrational signals were acquired with a 16-channel data logger, and the signal sampling frequency was set as 12 kHz.

This study attempts to use a small number of sample data to obtain the fault diagnosis model. Therefore, the bearing vibration signal is sampled and only 64 sample data under each operating condition for each fault mode are used to train and validate the diagnosis model. Since the bearings of wind turbines are subjected to transiently varying loads and rotational speeds, the fatigue faults of the bearings are diverse. To validate the accuracy and robustness of fault diagnosis, at first the test data are recombined into four sets of data. The first data set corresponds with a relatively uniform fault; that is, the numbers of faults of the inner race, the outer race, and the ball are substantially equal. The inner race, the outer race, and the ball fault separately occupy a majority in the remaining three data sets. Next, these four sets of data for analysis are used to verify the adaptability of the proposed fault diagnosis method.

5.2. Data characteristics of the bearing fault signals

To understand the bearing fault signals, the faults are manually
The normal bearings are first distinguished from faulty bearings by time-domain feature parameters. The peak-peak comparison of normal and faulty bearings is shown in Fig. 3. It can be seen that when the peak-peak value is less than 0.55, the bearing does not malfunction. Therefore, the normal bearings can be distinguished using the peak-peak values.

Next, the remaining faulty bearing signals are classified to determine the type of fault in these bearings. In this step, analysis is performed to distinguish between ball faults and inner/outer race faults. The peak-peak comparison of normal and faulty bearings is shown in Fig. 4. It is determined that when the peak-peak value is greater than 15, a rolling element failure has occurred; when the peak-peak value is greater than 4 and less than 15, a race failure has occurred; and when the peak-peak value is less than 4, the two faults are aliased together and cannot be distinguished. Therefore, the multiple time-domain feature parameters are used to subdivide the data with peak-peak values less than 4. The results show that the distinction is best when using the peak-peak indicator, but there is still a ball fault bearing that cannot be correctly distinguished.

Next, the inner and outer race faults are distinguished. The comparison of the values is shown in Fig. 5. It can be seen that when the peak-peak value is between 2 and 8, the two fault signals are mixed and are difficult to distinguish. When the peak-peak value is found in other ranges, the two fault signals can be directly distinguished. Therefore, the data are subdivided with peak-peak values between 2 and 8. The results show that there are still 5 data points that are mixed and cannot be distinguished.

According to the analysis above, the peak-peak parameters can effectively distinguish between normal bearings and faulty bearings; however, when classifying bearing fault types, time-domain parameters cannot correctly classify all faults, especially when distinguishing between inner and outer race faults. Therefore, the proposed ML methods are used to classify the bearing faults. The validation is presented in the next section.
5.3. Fault classification in its entirety

To match the effects of wind field characteristics on bearing fault types, the bearing fault signals are regrouped. In other words, three sets of bearing fault signals are constructed, and these sets separately contain more ball faults, outer race faults, and inner race faults. This approach allows us to utilize different diagnostic methods for different wind farms, thus improving the relevance and accuracy of diagnosis. The ML algorithms mentioned in Section 4 above were used to train the classifier. Then, the bearing data after the feature extraction and selection process are simultaneously classified into normal bearings, ball fault bearings, inner race fault bearings and outer race fault bearings by the trained classifier.

The classification rate of bearing fault for each method in its entirety was shown in Table 2. The comparison of fault classification accuracy and training time with different methods is shown in Fig. 6 and Fig. 7 respectively. Notice that the classification accuracy obtained by the KNN method is the highest. Again, the KNN method trained with a small number of fault data produces the highest classification accuracy of 89.4%. Although the training time of the KNN method is slightly higher than that of the other three methods, the training of the KNN method is still faster. Furthermore, as shown in Fig. 6, the classification accuracy improves as the training data increases.

Table 2
Classification rate of bearing fault for each method in its entirety.

<table>
<thead>
<tr>
<th>Method</th>
<th>Classification rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>75%</td>
</tr>
<tr>
<td>Naive Bayes classifier</td>
<td>70.83%</td>
</tr>
<tr>
<td>KNN</td>
<td>87.5%</td>
</tr>
<tr>
<td>ANN</td>
<td>83.33%</td>
</tr>
</tbody>
</table>

Fig. 5. Classification of outer and inner race bearing faults with peak-peak values.

Fig. 6. Comparison of fault diagnosis accuracy with different methods.

Fig. 7. Comparison of training time with different methods.

5.4. Fault classification by steps

In this section, a step-by-step method was used to classify the bearing faults with ML. That is, the process was performed with the following steps:

- Step 1: Build a classifier to distinguish the normal bearing from all bearing. At this time, the bearings were divided into two categories, normal bearings and faulty bearings.
- Step 2: Build a classifier to distinguish the ball fault bearing from faulty bearings. At this time, the faulty bearings were divided into two categories, ball fault bearings and race fault bearings including the inner and outer race fault bearings.
- Step 3: Build a classifier to distinguish the inner race fault bearing from race fault bearings. At this point, the bearing fault has been completely classified.

The classification rate of bearing fault for each method by steps was shown in Table 3. Note, it is easy to classify normal bearings and faulty bearings with classification rate 100% for all the proposed ML methods. This means that the proposed method works
very well for fault detection, and can distinguish between normal and faulty bearings at a very high classification rate. The classification rate decreased in the Step 2 stage, but the rate of all methods was still above 80%. Next, the classification performance of the four algorithms further decreased and showed significant differences for classifying the inner race fault and the outer race fault. The KNN method achieved a good performance with classification rate above 90% in the Step 3 stage. The classification rate of Naive Bayes method is relatively poor, below 70%. The results of the average rate show that the KNN method has the best classification performance, which is slightly higher than that of the ANN algorithm.

5.5. Analysis and discussion

By comparing the classification accuracy in Tables 2 and 3, it shows that the classification performance of ANN and KNN methods can be improved by using the step-by-step classification method. At the same time, the KNN method produces the best classification performance in both its entirety and step-by-step classification. Thus the KNN algorithm can provide good results for bearing fault diagnosis.

6. Comparison between the proposed method and regression based method

6.1. Regression-based fault detection method

Regression based fault detection identifies how signals and features are related to a state output in different components. This relationship is captured by fitting regression models when the system is in a healthy state. Therefore, this regression model is also known as the normal behavior model (NBM). When new data is obtained, the NBM prediction value of state variable is compared with the measured values. The difference between the predicted and actual measured values is called residual, and the residual are used for monitoring the condition. Residuals that exceed a given error threshold indicate malfunctioning.

For the fault detection in wind turbine, the residuals indicate how close the wind turbine's behavior resembles the one observed during training of the NBM. It is reasonable to assume that in a normal functioning wind turbine, the residuals are homogeneously randomly distributed. Therefore, trends or excessively large residuals indicate a departure from normal functioning and an alarm is triggered if the residuals exceed the given threshold. A number of studies have established regression models to predict state parameters such as bearing temperature, lubricant pressure, generator temperature and power curves in healthy state, and perform successfully the fault detection in wind turbine [19,33]. It should be noted that the above methods cannot directly classify the bearing fault types when detecting bearing faults.

In this section, an attempt is made to use the regression-based method for fault diagnosis of rolling bearings. The public data are first used to train the regression method for a continuous output or label. Then a threshold is given to discretize the regression result into discrete labels, e.g., fault or no-fault. The comparison between the proposed ML methods and regression-based method is done finally.

6.2. Regression-based modeling of bearing fault

The public dataset obtained from the Bearing Data Center of Case Western Reserve University were used to compare the proposed method with regression-based method [47]. For comparison, the regression model was trained with the same training and testing data used by the proposed ML model. In general, both parametric modeling techniques (linear regression and nonlinear regression) and non-parametric modeling techniques (ANN) are employed to predict the bearing fault. The bearing vibration signal after the feature extraction and selection process is used as the input data of the regression model. The status of bearing is divided into normal with label 0, ball fault with label 1, inner race fault with label 2, and outer race fault with label 3. The label value of the bearing is used as the output data of the regression model. The training of the regression model was performed in MATLAB version 2019a. Subsequently, data samples selected randomly are input into the trained regression model to predict the corresponding bearing state.

6.3. Evaluation criteria

To analyze the difference of each model clearly, the evaluation criteria are introduced for comparison. Mean absolute error (MAE) and root mean square error (RMSE) are used to evaluate the performance. The mathematical expressions are as follows:

\[
\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|
\]

Table 4

<table>
<thead>
<tr>
<th>Method</th>
<th>MAE</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear regression</td>
<td>0.5426</td>
<td>0.6574</td>
</tr>
<tr>
<td>Nonlinear regression</td>
<td>0.5052</td>
<td>0.6258</td>
</tr>
<tr>
<td>ANN</td>
<td>0.5589</td>
<td>0.6658</td>
</tr>
</tbody>
</table>

Fig. 8. Comparison of actual labels with predictive label based on regression model and KNN.
RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^{n} [y_t - \hat{y}_t]^2} \quad (29)

where \( y_t \) is the actual value, \( \hat{y}_t \) is the predicted value, and \( n \) is the number of data.

### 6.4. Comparison and discussion

The trained regression models were evaluated with MAE and RMSE criteria, and the results were listed in Table 4. Notice that the MAE and RMSE values corresponding to the nonlinear regression model are the smallest, that is, the accuracy of the nonlinear regression model is the highest. The labels of actual values and the predictive values are shown in Fig. 8. It can be seen that the actual value of the label and the predictive value generated by the KNN method are both discrete data. The predicted value of the KNN produced three prediction errors, and the others were consistent with the actual value. The label that is predicted based on the regression method is a continuous label. A threshold needs to be given to determine the state of the bearing. The comparison between the actual value and predicted value of bearing label produced with nonlinear regression is shown in Table 5. Currently, how to determine the threshold is the key to fault diagnosis. For example, if the threshold for outer race faults is set to 2.5, five bearings with outer race fault will be correctly identified. At this time, the detection ratio of outer race fault was 71.4%. But unfortunately, the other three bearings with inner race anomalies were mistakenly believed to have outer race fault. Obviously, the design of thresholds heavily depends on the designer’s experience and knowledge. At the same time, the threshold setting will also change for different wind farm environments. The proposed ML method can automatically complete the classification of faults without the need to set thresholds and provides a diagnostic accuracy about 90% that is produced by KNN. Therefore, the proposed method outperforms the regression-based method for bearing fault diagnosis.

### 7. Conclusion

This paper proposes a novel scheme for wind turbine bearing fault diagnosis based on data mining and ML. The proposed method works directly on original data obtained from vibration signals of bearings under different operation conditions. There are two main steps in the proposed method. (1) A systematic method is presented to assess the prediction accuracy of each characteristic to find the combination of the most effective features for fault diagnosis of wind turbine bearings. (2) ML is utilized to perform the fault diagnosis of wind turbine bearings. The experiments demonstrate that this method can effectively extract features from nonstationary signals, obtaining excellent results in the fault diagnosis of wind turbine bearings.

**CRediT authorship contribution statement**

**Bodi Cui**: Conceptualization, Methodology, Writing — original draft, Writing — review & editing. **Yang Weng**: Conceptualization, Methodology, Writing — original draft, Writing — review & editing. **Ning Zhang**: Methodology, Writing — review & editing.

**Declaration of competing interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Acknowledgements**

This work was supported in part by the Department of Energy under grants DE-AR0000858-1631 and DE-EE0009355, the National Science Foundation (NSF) under the grants ECCS-1810537 and ECCS-2048288.

**References**


[14] L. Vamsi, G. Sabareesh, P. Penumakala, Comparison of condition monitoring techniques in assessing fault severity for a wind turbine gearbox under non-