TRANSACTIONS OF THE INSTITUTE OF MEASUREMENT & CONTROL

Article

Transactions of the Institute of Measurement and Control
2018, Vol. 40(8) 2681–2693

DOI: 10.1177/0142331217708242
journals.sagepub.com/home/tim
SAGE

Comparison of random forest, artificial neural networks and support vector machine for intelligent diagnosis of rotating machinery

Te Han[1], Dongxiang Jiang[1], Qi Zhao[2], Lei Wang[2] and Kai Yin[2]

Abstract

Nowadays, the data-driven diagnosis method, exploiting pattern recognition method to diagnose the fault patterns automatically, achieves much success for rotating machinery. Some popular classification algorithms such as artificial neural networks and support vector machine have been extensively studied and tested with many application cases, while the random forest, one of the present state-of-the-art classifiers based on ensemble learning strategy, is relatively unknown in this field. In this paper, the behavior of random forest for the intelligent diagnosis of rotating machinery is investigated with various features on two datasets. A framework for the comparison of different methods, that is, random forest, extreme learning machine, probabilistic neural network and support vector machine, is presented to find the most efficient one. Random forest has been proven to outperform the comparative classifiers in terms of recognition accuracy, stability and robustness to features, especially with a small training set. Additionally, compared with traditional methods, random forest is not easily influenced by environmental noise. Furthermore, the user-friendly parameters in random forest offer great convenience for practical engineering. These results suggest that random forest is a promising pattern recognition method for the intelligent diagnosis of rotating machinery.

Keywords

Intelligent fault diagnosis, rotating machinery, random forest, artificial neural networks, support vector machine

Introduction

As a kind of industrial infrastructure, rotating machinery has become one of the key equipments in a lot of industry sectors such as power system and aerospace engineering. With the extreme application conditions, the critical components in rotating machinery including bearing, gearbox, rotor, and so forth, are easily subject to faults, which may cause the machine breakdown and economic loss. It is very necessary to develop the condition monitoring and fault diagnosis (CMFD) technologies for rotating machinery. Vibration analysis has been accepted as a major diagnostic tool because vibration signals can be obtained handily and contain abundant information of machine conditions (Yunusa-Kaltungo et al., 2015).

Once we acquire the vibration signal, there are mainly two categories of diagnostic approaches: signal processing-based approaches and pattern recognition-based approaches (Guo et al., 2016). In the first class, fault pattern can be identified by detecting the feature of vibration waveforms or the fault characteristic frequency using advanced signal processing methods, such as wavelet transform (WT), empirical mode decomposition (EMD) and spectral kurtosis (SK) (Han et al., 2016; Sun et al., 2017; Tang et al., 2016; Yaqub and Loparo, 2016), while this procedure requires the operator to grasp a good deal of expertise, which may bring difficulties for online diagnosis. Consequently, as the second ones, the data-driven diagnostic techniques, which can realize the automated and intelligent diagnosis with pattern recognition methods, receive wide attention and develop rapidly in recent years (Bogoevska et al., 2017; Cheng et al., 2016; Han et al., 2017; Liu et al., 2014). This strategy includes three important parts: data acquisition, feature extraction and pattern recognition. Adopting the extracted features from the vibration signals of different failure modes, the classification model can be trained, so as to make decision intelligently when similar patterns come afterwards. Usually, the capability of classification model has a significant influence on diagnostic results, indicating the need to attempt appropriate pattern recognition algorithm for rotating machinery (Han and Jiang, 2016). The most used classifiers in this field are artificial neural networks (ANNs) including back propagation neural network (BPNN), radical basis function (RBF), learning vector quantization

[1]State Key Lab of Control and Simulation of Power System and Generation Equipment, Department of Thermal Engineering, Tsinghua University, China
[2]AECC Commercial Aircraft Engine Co., Ltd, Shanghai, China

Corresponding author:
Te Han, State Key Lab of Control and Simulation of Power System and Generation Equipment, Department of Thermal Engineering, Tsinghua University, Beijing 100084, China.
Email:hant15@mails.tsinghua.edu.cn

(LVQ) (Jiang and Liu, 2011), wavelet neural network (WNN) (Lei et al., 2011), extreme learning machine (ELM) (Tian et al., 2015), probabilistic neural network (PNN) (Zhou and Cheng, 2016; Dou and Zhou, 2016) and so forth, and support vector machine (SVM) (Li and Zhang, 2011).

ANN, which essentially comes from the abstraction of human brain neural network, is a hot spot in the field of artificial intelligence since 1980s. To date, there are nearly 40 kinds of neural network models in which many feedforward neural network models have been extensively studied and documented for fault detection and diagnosis in mechanical system. The problems of slow training speed, local minima and sensitive learning rate in many traditional models such as BPNN, have become the main bottlenecks, constraining the development of this technology. As two kinds of modified models from the large family of feedforward neural networks, ELM and PNN can successfully avoid these defects with the excellent ability of self-organizing and self-learning. Nevertheless, the signal pre-processing is usually needed for most of ANNs owing to the sensibility to feature magnitudes. Additionally, the performance of ANNs may be undesirable when dealing with high-dimensional features without dimension reduction or prior feature selection. Moreover, a good deal of typical training samples are necessary to ensure the ergodicity and improve robustness against overfitting (Xu and Chen, 2013).

SVM, developed based on statistical learning theory, is a generic and effective pattern recognition method for rotating machinery fault diagnosis. To be more accurate, SVM approximately realizes the goal of structural risk minimization for binary classification problems. Some research has shown that nonlinear SVM not only can cope with high-dimensional features, but has a stronger generalization capability than ANNs when solving small sample learning problems (Yang et al., 2007). Unfortunately, the recognition accuracy will be severely degraded if two model parameters, namely the penalty factor and kernel function parameter, are not properly selected. Although the cross-validation (CV) method and some heuristic methods such as genetic algorithm (GA) and particle swarm optimization (PSO) can be utilized to guide the selection of parameters, the whole procedure of parameters tuning would become rather complex and time-consuming.

Faced with the aforementioned problems, there is, in particular, a need to employ more efficient pattern recognition methods in the field of fault diagnosis for rotating machinery, especially those are robust to the features, model parameters selection and even the number of training samples. Random forest is such a powerful pattern recognition method that may meet these requirements based on existing studies. However, it is relatively unfamiliar in this field. Up to now, to the best of our knowledge, there is no comprehensive comparison of performance between random forest and other common models using the fault datasets of rotating machinery. Random forest is developed from a novel learning strategy called "ensemble learning", which combine many base classifiers and synthetically consider their results (Breiman, 2001). Before random forest, the bagging algorithm with bootstrap sampling proposed by Breiman (1996) and the random decision forests (RDF) using the idea of random subspace presented by Ho (1998) are also based on this learning strategy. With the successful applications in the 1990s, ensemble learning has been proven to remarkably improve the accuracy and generalization capability of the whole system (Hansen et al., 1992; Schwenk and Bengio, 2000). In 2001, Breiman (2001) creatively put forward the random forest merging the bagging and the random split selection of features. In recent years, random forest has drawn wide attention in the fields of E-tongue (Liu et al., 2013), acoustic emission (Morizet et al., 2016), mass spectrometry (Coomans et al., 2006), digital soil mapping (Yang et al., 2016), eye state estimation (Dong et al., 2016), remote sensing image (Luo et al., 2016), et al.

In this paper, we will introduce the novel random forest algorithm to rotating machinery fault diagnosis and focus on its performance, comparing with currently more popular methods, that is, ELM, PNN and SVM. We use three classes of features from time-domain, frequency-domain and multiple scale components, respectively, and two validation datasets that are the bearing dataset from Case Western Reserve University bearing data center and a gearbox dataset from our test rig. The remaining part is organized as follows. In section 2, the principles of random forest, ELM, PNN and SVM are introduced briefly. The main diagnosis procedure is given in section 3. In section 4, the two datasets are described and the experimental verification is conducted, including the comparison and discussion with different methods in the aspects of accuracy, standard deviation, sensitivity to noise, parameters tunning and time consumption. A comparative discussion between this work and some published literatures are also given. Finally, the conclusions can be drawn in section 5.

Background knowledge

Random forest

Random forest is a typical ensemble learning method that operates by integrating multiple weak decision tree classifiers and reaches a final decision by the majority of votes. In practice, the performance of a single classifier always differs greatly depending on the different types of applications or the dimension of feature space. Therefore, researchers put forward the ensemble learning methods to improve the recognition capability of individual classifier. Dietterich (2000) demonstrated the superiority of ensemble learning in the aspects of statistics, local search and function expression by theoretical analysis. In earlier studies, how to keep the diversity of base learners plays a critical role in a successful ensemble learning method. Three novel algorithms were effectively developed, including Breiman's "bagging" (Breiman, 1996), Schapire's "boosting" and Ho's "random subspace" (Ho, 1998).

In fact, random forest merges the ideas of bagging and random subspace, resulting in two valuable randomness (see Figure 1). In bagging algorithm, multiple training subsets can be drawn from the training set with bootstrap method (randomly sampling with replacement). Each bootstrap retains the same size of original training set while some molecules are repeated and some others are left out. Actually, two-thirds of training samples are applied to grow each tree. The other lost one-third of training samples called Out-of-Bag (OOB) data can be used for performance estimation,

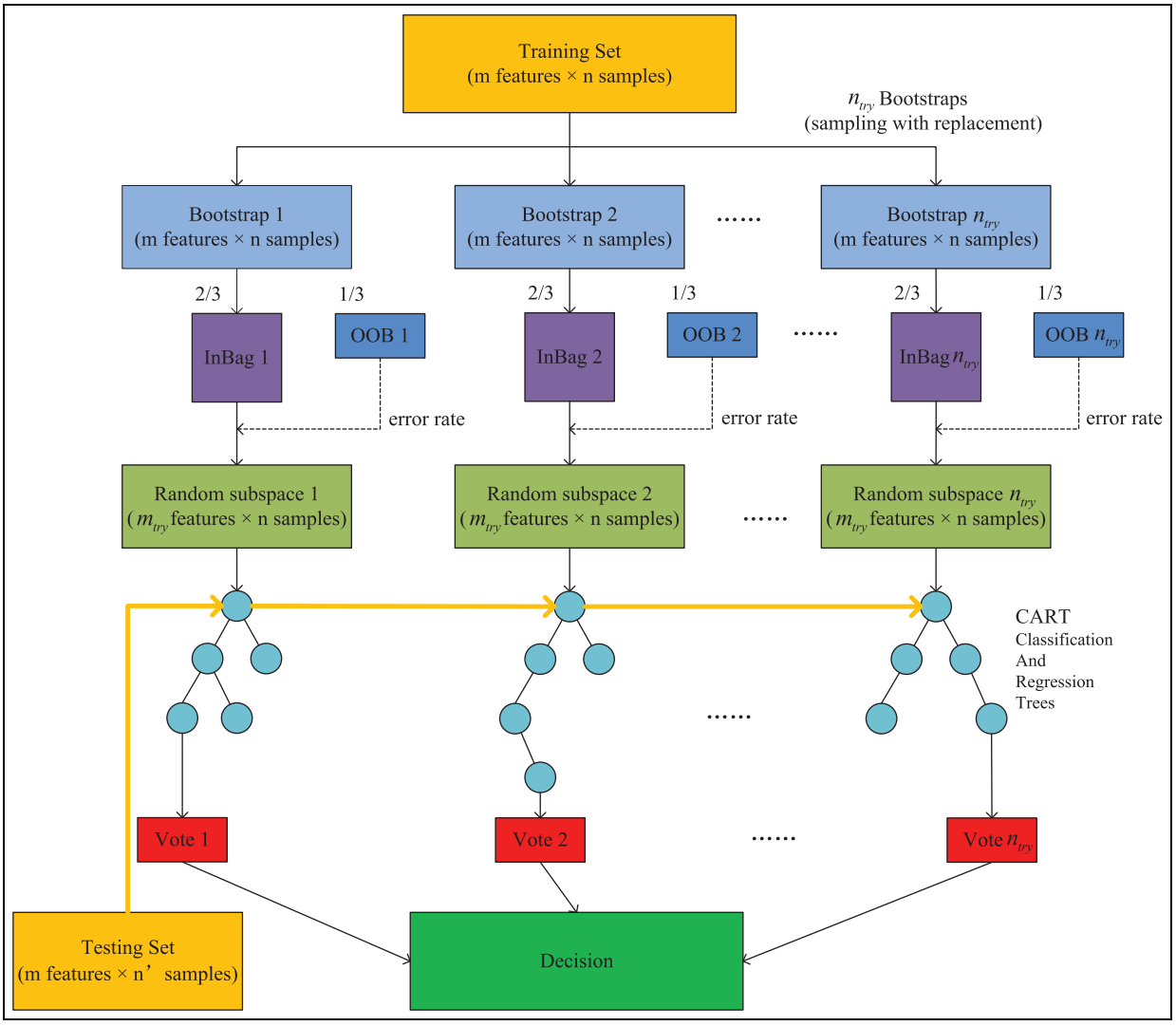

Figure 1. Workflow of random forest algorithm.

suggesting that there is no requirement of additional cross-validation for random forest. With a forest consisting of n_{tree} trees, corresponding n_{tree} bootstraps can be generated, indicating the randomness of training samples for each tree. Essentially, the bagging algorithm can prominently reduce the variance of the base classifier and improve the generalization error. The second randomness is introduced by Ho's "random subspace", which means only a randomly selected subset (m_{try} descriptors) is considered for each tree instead of all descriptors in the input feature vector. On the one hand, this idea further promotes the diversity of base learners. On the other hand, it makes the tree less greedy and increases the possibility that some weak features can get into the tree. The effect of all descriptors in the input feature vector will be magnified and these weak features may become beneficial in combination with other features. All the n_{tree} trees in the forest are grown without pruning and the growing algorithm is CART (classification and regression trees). A comprehensive prediction can be evaluated by the ensemble n_{tree} votes for a new testing sample (Svetnik et al., 2003).

Based on the existing application of random forest, some attracted advantages can be presented such as the robustness against overfitting, only two user-friendly parameters: n_{tree} and m_{try}, insensitive to prior feature selection, good ability to handle with badly unbalanced dataset, and so forth, showing its great potentials for intelligent fault diagnosis of rotating machinery.

ELM

As a new single-hidden layer feedforward artificial neural network (SLFN), ELM can randomly generate the input weights and hidden bias with no adjustment in the process of training, and calculate output weights according to Moore–Penrose generalized inverse. Provide a training set $\{(\boldsymbol{x}_i, \boldsymbol{t}_i) | \boldsymbol{x}_i \in R^n, \boldsymbol{t}_i \in R^m, i = 1, 2, \ldots, Q\}$ where x_i is the *ith* training

sample and m denotes the number of total classes. The output of ELM can be expressed as below

$$\begin{aligned} \boldsymbol{T} &= \left[\boldsymbol{t}_1, \boldsymbol{t}_2, \ldots, \boldsymbol{t}_Q\right]_{m\times Q},\ \boldsymbol{t}_j = \begin{bmatrix} t_{1j} \\ t_{2j} \\ \vdots \\ t_{mj} \end{bmatrix} \\ &= \begin{bmatrix} \sum_{i=1}^{l} \beta_{i1} g\left(\boldsymbol{w}_i \boldsymbol{x}_j + b_i\right) \\ \sum_{i=1}^{l} \beta_{i2} g\left(\boldsymbol{w}_i \boldsymbol{x}_j + b_i\right) \\ \vdots \\ \sum_{i=1}^{l} \beta_{im} g\left(\boldsymbol{w}_i \boldsymbol{x}_j + b_i\right) \end{bmatrix}_{m\times 1} \quad (j = 1, 2, \ldots, Q) \end{aligned} \tag{1}$$

where l is the number of hidden neurons, $g(x)$ is the activation function, $\boldsymbol{w}_i = [w_{i1}, w_{i2}, \ldots, w_{in}]$ is the input weights connecting all the input nodes and the *ith* hidden node, b_i is the bias of the *ith* hidden node and β_{im} is the output weight between *ith* hidden neuron and *mth* output neuron. Equation (1) can be rewritten as the following form

$$\begin{aligned} &\boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{T}', \\ &\boldsymbol{H}\left(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_l, b_1, \ldots, b_l, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_Q\right) \\ &= \begin{bmatrix} g(\boldsymbol{w}_1 \cdot \boldsymbol{x}_1 + b_1) & \ldots & g(\boldsymbol{w}_l \cdot \boldsymbol{x}_1 + b_l) \\ \vdots & \ddots & \vdots \\ g\left(\boldsymbol{w}_1 \cdot \boldsymbol{x}_Q + b_1\right) & \ldots & g\left(\boldsymbol{w}_l \cdot \boldsymbol{x}_Q + b_l\right) \end{bmatrix}_{Q\times l} \end{aligned} \tag{2}$$

where $\boldsymbol{H}$ is the output matrix of hidden layer of the network. For the fixed $\boldsymbol{H}$, the training process is equal to find an optimal $\boldsymbol{\beta}$ that can be decided by the least-squares solution of the following equation

$$\min_{\beta} \|\boldsymbol{H}\boldsymbol{\beta} - \boldsymbol{T}'\| \tag{3}$$

Once the single parameter: the number of hidden neurons is set, the unique optimal solution can be obtained.

PNN

PNN is developed from the RBF network and the theoretical foundation is the Bayes decision theory. The fault diagnosis based on PNN is one kind of widely accepted decision method in probability statistics. It can be described as: supposing two assigned fault classes (θ_A and θ_B), for a testing sample $\boldsymbol{X} = (x_1, x_2, \ldots, x_n)$,

$$\begin{aligned} &\text{if } h_A l_A f_A(X) > h_B l_B f_B(X), \text{ then } \boldsymbol{X} \in \theta_A \\ &\text{if } h_A l_A f_A(X) < h_B l_B f_B(X), \text{ then } \boldsymbol{X} \in \theta_A \end{aligned}$$

where h_A, h_B are the priori probability of class θ_A, θ_B, l_A is the cost factor that the fault sample belonging to θ_A is misclassified into θ_B, l_B is the cost factor that the fault sample belonging to θ_B is misclassified into θ_A, f_A and f_B are the PDFs of classes θ_A and θ_B, respectively. The key PDFs can be estimated by Parzen method

$$f_A(\boldsymbol{X}) = \frac{1}{(2\pi)^{P/2}\delta^P} \frac{1}{m} \sum_{i=1}^{m} exp\left\{ -\frac{(\boldsymbol{X} - \boldsymbol{X}_{Ai})^T (\boldsymbol{X} - \boldsymbol{X}_{Ai})}{2\delta^2} \right\} \tag{4}$$

where m is the total number of training samples in class A, $\boldsymbol{X}$ is the testing sample, $\boldsymbol{X}_{Ai}$ is the *ith* training sample of class A, δ is the smoothing parameter and p the dimensionality of the space.

As a feedforward neural network, PNN consists of four layers, which are input layer, pattern layer, summation layer and output layer. The input layer transmits input vectors to the nodes of pattern layer. The pattern layer calculates the distance between the input vector and the patterns of training set by a nonlinear operator. Then, the summation layer simply sums the output from pattern layer and estimates the probability densities using all multivariate Gaussian distributions. Finally, the output layer classifies the vector into a pattern depending on the maximum of these probabilities.

SVM

The core idea of SVM is to create a classification hyperplane as decision surface, maximizing the margin of separation between the two classes. Given a training sample set $S = \{x_i, y_i\}_{i=1}^{l}$, each sample x_i is an n-dimensional input feature vector and has a binary class label $y_i \in \{-1, +1\}$. For linearly separable samples, the optimal hyperplane can be found by solving the following optimization problem

$$\begin{aligned} &\min \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{l} \xi_i \ s.t.\ y_i(\omega \cdot x_i + b) \\ &\geq 1 - \xi_i,\ \ \xi_i \geq 0,\ \ i = \{1, 2, \ldots, l\}, \end{aligned} \tag{5}$$

where ω is the normal vector of the hyperplane, b is the bias, ξ_i are the slack variables and C is the penalty factor. When the training samples are linearly inseparable in the feature space, a non-linear function $\phi(x)$ can be implemented to map them into a high-dimensional feature space. The kernel function returns a dot product of the mapped feature space, that is $K\left(x_i.x_j\right) = \left(\phi^T(x_i) \cdot \phi\left(x_j\right)\right)$. The universally used radial basis function (RBF) kernel function in the field of fault diagnosis can be expressed as follows

$$K(x, x_i) = exp\left(-\frac{\|x - x_i\|^2}{2g^2} \right), \tag{6}$$

where g is the kernel function parameter. After introducing the Lagrange multipliers $\alpha_i \geq 0$, the problem of equation (8) can be transformed as

$$\begin{aligned} &\min \frac{1}{2}\sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j K\left(x_i.x_j\right) \\ &- \sum_{i=1}^{l} \alpha_i \ s.t.\ 0 \leq \alpha_i \leq C,\ \sum_{i=1}^{l} \alpha_i y_i = 0. \end{aligned} \tag{7}$$

Then, SVM can predict the label of samples via the optimal classification hyperplane and the decision function can be given by

Table 1. The statistical characteristic parameters in time-domain.

Number	Feature expression	Number	Feature expression	Number	Feature expression
$tp1$	$\frac{\sum_{n=1}^{N} x(n)}{N}$	$tp7$	$max(x(n))$	$tp13$	$\frac{tp_7}{tp_4}$
$tp2$	$\sqrt{\frac{1}{N}\sum_{n=1}^{N} x_n^2}$	$tp8$	$min(x(n))$	$tp14$	$\frac{tp_7}{tp_3}$
$tp3$	$\left(\frac{1}{N}\sum_{n=1}^{N}\sqrt{\lvert x_n\rvert}\right)^2$	$tp9$	$tp_7 - tp_8$	$tp15$	$\frac{tp_5}{(tp_2)^3}$
$tp4$	$\frac{\sum_{n=1}^{N}\lvert x(n)\rvert}{N}$	$tp10$	$\frac{\sum_{n=1}^{N}(x(n)-tp_1)^2}{N-1}$	$tp16$	$\frac{tp_6}{(tp_2)^4}$
$tp5$	$\frac{\sum_{n=1}^{N}(x(n)-tp_1)^3}{N}$	$tp11$	$\frac{tp_2}{tp_4}$		
$tp6$	$\frac{\sum_{n=1}^{N}(x(n)-tp_1)^4}{N}$	$tp12$	$\frac{tp_7}{tp_2}$		

Where $x(n)$ is a sampling series of raw signal with N points.

Table 2. The statistical characteristic parameters in frequency-domain.

Number	Feature expression	Number	Feature expression	Number	Feature expression
$fp1$	$\frac{\sum_{k=1}^{K} y(k)}{K}$	$fp6$	$\frac{\sum_{k=1}^{K} f_k y(k)}{\sum_{k=1}^{K} y(k)}$	$fp11$	$\frac{\sum_{k=1}^{K}(f_k-fp_5)^3 y(k)}{K(fp_6)^3}$
$fp2$	$\frac{\sum_{k=1}^{K}(y(k)-fp_1)^2}{K-1}$	$fp7$	$\sqrt{\frac{\sum_{k=1}^{K} f_k^2 y(k)}{\sum_{k=1}^{K} y(k)}}$	$fp12$	$\frac{\sum_{k=1}^{K}(f_k-fp_5)^4 y(k)}{K(fp_6)^4}$
$fp3$	$\frac{\sum_{k=1}^{K}(y(k)-fp_1)^3}{K\left(\sqrt{fp_2}\right)^3}$	$fp8$	$\sqrt{\frac{\sum_{k=1}^{K} f_k^4 y(k)}{\sum_{k=1}^{K} f_k^2 y(k)}}$	$fp13$	$\frac{\sum_{k=1}^{K}(f_k-fp_5)^{1/2} y(k)}{K\sqrt{fp_6}}$
$fp4$	$\frac{\sum_{k=1}^{K}(y(k)-fp_1)^4}{K(fp_2)^2}$	$fp9$	$\frac{\sum_{k=1}^{K} f_k^2 y(k)}{\sqrt{\sum_{k=1}^{K} y(k)\sum_{k=1}^{K} f_k^4 y(k)}}$		
$fp5$	$\sqrt{\frac{\sum_{k=1}^{K}(f_k-fp_5)^2 y(k)}{K}}$	$fp10$	$\frac{fp_6}{fp_5}$		

Where $y(k)$ is the frequency spectrum of the discrete signal and k means the sequence number of spectrum line; f_k represents the frequency value of the kth. spectrum line.

$$f(x) = sign\left[\sum_{i=1}^{l} \alpha_i y_i K(x_i x) + b\right]. \quad (8)$$

Since the algorithm above is designed for binary classification, when dealing with k classes, we can convert the problem into $k(k-1)/2$ binary classifications, indicating one SVM model is designed between any two classes. The label of a sample can be identified according to the most votes from all binary classifiers. The popular libSVM toolbox (Chang and Lin, 2011) exactly utilizes this approach. It should be noted that the multi-class SVM is fundamentally different from the ensemble learning of random forest. All the individual learners aim to the same problem in ensemble learning methods while the strategy of multi-class SVM is to divide and conquer the entire problem.

Data-driven fault diagnostic techniques for rotating machinery

As the data-driven diagnostic technique, the first step is data acquisition. In this experimental analysis, a number of vibration signal samples can be monitored using acceleration sensors when the machine works. Different types of artificial faults are introduced to rotating machinery test rig, and thus the samples under different machine conditions can be acquired. In real industrial case, the history database collected by condition monitoring systems contains abundant signal types after the long-time operation. Hence, the signal samples can be manually selected. With the obtained fault datasets, the samples of each machine condition are divided into two classes: training sample and testing sample.

For feature extraction, different features may lead to different diagnostic results. To have a comprehensive comparison, three classes of popular features are extracted for further model training: time-domain statistical features (TDF), frequency-domain statistical features (FDF) and multiple scale features (MCF). (1) When a fault occurs in a mechanical system, the time-domain waveform of vibration signal may change both in amplitude and distribution. Sixteen statistical characteristic parameters from time-domain in Table 1 are taken. (2) Similar to time-domain, the frequency spectrum of fault signal may also alter since the impact of faults will arouse the resonance of the system and some high frequency components may appear. Thirteen statistical characteristic parameters from frequency-domain are displayed in Table 2. (3) With the development of non-stationary analysis methods

Table 3. Description of two datasets.

Experimental validation	Object	Health status	Load speed (rpm)	Training samples	Testing samples	Label	Sample length
Dataset A	Rolling bearing	Health	1797, 1772, 1750, 1730	5, 10, 20, 50	50	1	1024
		Inner race fault	1797, 1772, 1750, 1730	5, 10, 20, 50	50	2	
		Outer race fault	1797, 1772, 1750, 1730	5, 10, 20, 50	50	3	
		Ball fault	1797, 1772, 1750, 1730	5, 10, 20, 50	50	4	
Dataset B	Gearbox	Health	1500, 1200, 900	5, 10, 20	40	1	2048
		Tooth broken	1500, 1200, 900	5, 10, 20	40	2	
		Tooth surface spalling	1500, 1200, 900	5, 10, 20	40	3	
		Gear root crack	1500, 1200, 900	5, 10, 20	40	4	

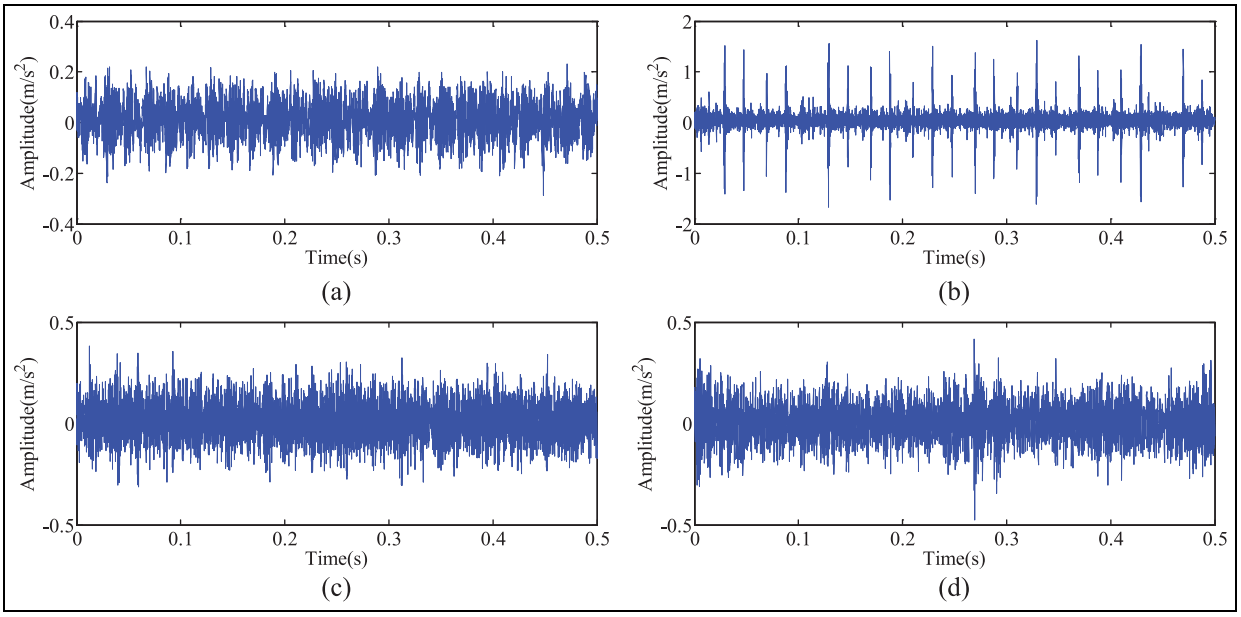

Figure 2. Vibration waveform of bearing with four states: (a) health; (b) inner race fault; (c) outer race fault; (d) ball fault.

for vibration signal of rotating machinery, variational mode decomposition (VMD) (Dragomiretskiy and Zosso, 2014) is utilized to decompose the vibration signal into several stable components, which represent different frequency bands and reflect the multi-scale information from raw signal. Then, an AR model can be established for each component. All the parameters of AR model are served as the feature vectors, which can denote the condition of the mechanical system effectively.

The next stage is to realize intelligent diagnosis via pattern recognition techniques. The classification model can be trained with the extracted feature vectors from training samples. In this step, the parameters tunning is crucial for some pattern recognition methods such as ANNs and SVM. Then, the testing set is applied to examine the diagnostic performance, so as to ensure the learned classifier can make diagnosis precisely when similar faults occur in the rotating machinery afterwards. In this paper, three indexes are used for quantitative assessment, namely mean accuracy, standard deviation and time consumption based on many random tests. Additionally, the complexity level of parameters tunning in different classifiers is taken into consideration.

Experimental verification and discussion

Experimental data description

In this paper, the comparative study was conducted on two datasets: the bearing dataset from Case Western Reserve University bearing data center and a gearbox dataset from our test rig. Dataset A consists of four classes, which are health, inner race fault, outer race fault and ball fault, considering four load speeds; that is, 1797rpm, 1772rpm, 1750rpm and 1730rpm. Single point faults are introduced to the test bearings using electro-discharge machining with fault diameters of 0.014 inches (1 inches = 25.4 mm). The vibration waveforms of the bearing with 1797rpm load speed are presented in Figure 2. In each load condition, the vibration signal of every bearing status is truncated to 100 samples with the length of 1024 points. The training set and testing set can be partitioned at random. To compare the performances of classifiers with different size of training set, in each bearing status, we employed the different number of samples (5, 10, 20 and 50 respectively) to train the models and the other 50 samples are utilized for verification.

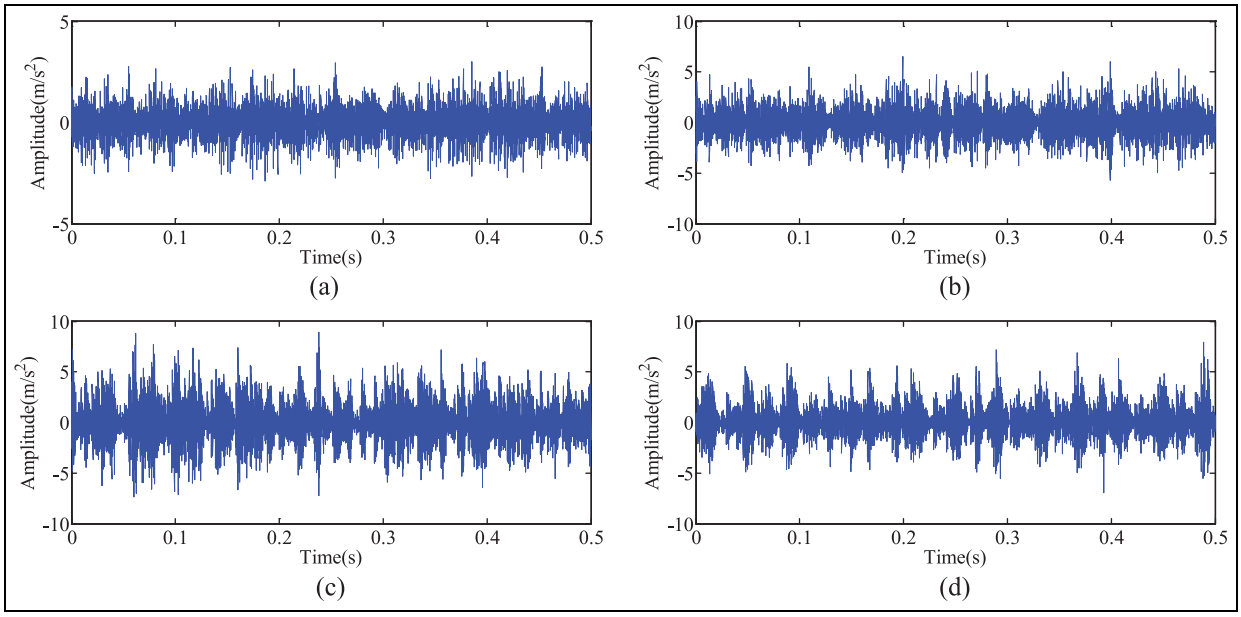

Figure 3. Vibration waveform of gearbox with four states: (a) health; (b) tooth broken; (c) tooth surface spalling; (d) gear root crack.

Similarly, the gearbox dataset B is composed of four classes, which are health, tooth broken, tooth surface spalling, and gear root crack, respectively, under three operating conditions of high speed shaft, that is, 1500rpm, 1200rpm and 900rpm. Tooth broken fault is simulated by cutting off a tooth in gear with the depth of 1.5mm. For tooth surface spalling, we process a deep groove with the width of 1.5mm and the depth of 1mm on the surface of a tooth. And the failure mode of gear root crack is made by manufacturing a crack on the dedendum with the width of 0.1mm and the depth of half the dedendum width. Figure 3 displays the waveforms of the gearbox with 1500rpm load speed. A total of 60 samples of 2048 dimensions are generated in which different numbers of training samples (5, 10 and 20, respectively) and 40 testing samples are picked out randomly to form the training set and testing set in one class. For both the two datasets, each test is repeated 50 times to get a mean accuracy and standard deviation. The final results can be acquired by the average of overall load conditions.

Performance comparison with different features in two datasets

In this part, we will compare the average diagnosis rate and standard deviation by random forest, ELM, PNN and SVM using different features in the two datasets. Firstly, the selection criteria of pivotal parameters in each classifiers are introduced. (1) Random forest: the default vue of n_{try}, representing the number of independent trees in the forest, is 500. The number of random feature subset, m_{try}, out of m total features are set to the recommended value, that is $m_{try} = \lfloor\sqrt{m}\rfloor$ (Svetnik et al., 2003). (2) ELM: the optimal number of hidden neurons in the hidden layer is determined through the 5 fold CV on training samples in each test, ranging from 1 to 200. (3) PNN: similarly, the spread value is determined by the 5 fold CV. (4) SVM: the most popular 'RBF' kernel function is adopted. The selection of another two crucial structural parameters in SVM, penalty factor c and kernel function parameter g, are decided by GA in the sense of 5 fold CV. According to literature (Luo et al., 2015), we set the maximum generation to 50 and the number of populations to 20 in GA. Detailed parameters discussion will be interpreted in the subsequent part.

Tables 4 and 5 give the diagnosis performance in two datasets respectively. In dataset A, it is clear that random forest achieves the higher diagnosis accuracies and the lower standard deviations than that of the other three classifiers in all cases. When we apply plenty of samples to train classifiers (n = 50), all the four models can reach high recognition rates with FDF and MSF, whereas the diagnosis rate of PNN drops to 89.82% using the TDF, which means PNN cannot handle with TDF very well in these tests. Similarly, random forest also performs much better than ELM and SVM using TDF. Besides, it is worth mentioning that a quite satisfactory result can be obtained by random forest only with a small training set (n = 5, 10, for example), which proves its strong generalization ability and excellent self-learning capacity. For ELM and SVM, when training sample size is reduced to 5, the accuracies distinctly decrease to 77.00% and 86.41% respectively using TDF, 84.66% and 84.14% respectively using FDF. Simultaneously, the higher standard deviations reflect the unstable behaviour of ELM and SVM. In essence, most ANNs and SVM need a lot of typically labelled samples to train the model structure, implying that it may be nontrivial to popularize the algorithms in industrial tasks due to the lack of fault samples. A more intuitional change of diagnostic accuracies with different features and training sample sizes in the four classifiers is displayed in Figure 4. Similar results can be observed from dataset B. Based on the comparison above,

Table 4. Comparison of diagnosis results with different features in dataset A.

Feature	No. of training samples	Random forest		ELM		PNN		SVM	
		Acc(%)	Std(%)	Acc(%)	Std(%)	Acc(%)	Std(%)	Acc(%)	Std(%)
TDF	50	98.04	1.09	95.89	1.22	89.82	1.96	93.90	1.68
	20	96.18	1.81	93.44	2.16	85.51	3.32	92.25	2.27
	10	94.31	2.00	89.01	4.41	80.93	4.15	90.42	3.76
	5	92.69	2.41	77.00	8.71	77.12	5.38	86.41	6.38
FDF	50	99.76	0.30	99.43	1.09	99.10	0.73	98.33	1.09
	20	99.59	0.56	98.40	2.13	98.53	1.17	95.10	2.65
	10	99.08	1.21	96.06	3.80	97.23	2.17	89.60	4.18
	5	97.77	1.86	84.66	7.34	94.24	4.64	84.14	4.88
MSF	50	99.85	0.28	99.29	1.33	99.84	0.30	99.57	0.82
	20	99.61	0.53	98.45	2.76	99.48	0.76	99.06	1.38
	10	99.24	0.95	96.95	3.61	98.81	1.30	97.92	2.43
	5	98.64	1.29	95.46	5.13	97.93	1.67	96.04	4.04

Table 5. Comparison of diagnosis results with different features in dataset B.

Feature	No. of training samples	Random forest		ELM		PNN		SVM	
		Acc(%)	Std(%)	Acc(%)	Std(%)	Acc(%)	Std(%)	Acc(%)	Std(%)
TDF	20	97.64	1.29	94.13	2.55	84.20	3.03	96.48	1.64
	10	96.31	2.13	87.46	4.33	76.90	4.38	93.75	3.16
	5	94.72	2.89	76.66	7.34	67.55	6.26	88.45	6.38
FDF	20	99.57	0.46	98.99	1.40	98.51	1.45	98.14	1.88
	10	99.34	0.59	96.63	3.19	96.63	2.39	95.96	2.55
	5	98.53	1.23	93.46	5.38	93.24	4.74	91.80	3.82
MSF	20	97.50	1.16	97.19	2.08	95.28	2.13	97.04	1.77
	10	95.98	2.27	92.08	6.13	93.51	2.77	93.81	3.25
	5	92.47	4.61	87.00	7.77	89.38	5.22	86.27	9.89

the following points can be drawn. (1) Random forest is capable of handling with three types of features without prior selection, showing it is a powerful and efficient pattern recognition approach in the field of fault diagnosis. (2) Random forest always retains a superior diagnostic performance regardless of the number of training samples and significantly outperforms the two kinds of ANNs and SVM in the case of small training set.

Performance comparison under noise environment

In real industrial tasks, the raw vibration signal always contains much noise, which has great effects on the diagnostic results. However, it is difficult to acquire all the labelled training sets under different noisy environments. To conform this condition, the additive white Gaussian noise is artificially added to testing samples with different signal to noise ratios (SNRs). The noise tests are carried out to investigate the robustness of these classification models with different SNRs. This experiment is conducted in dataset A using FDF as input vectors, and sufficient training samples ($n = 50$) are adopted. The diagnosis results for the analysed classifiers are shown in Table 6. It is clear to see that random forest achieves statistically significant win over the comparative methods when handling the test samples with different levels of noise. As shown in Figure 5, for ELM, PNN and SVM, an obvious downward trend subsequently appears when the SNR is lower than 22 dB, while random forest still performs significantly well in a wider range of SNR owing to the strong anti-noise ability of bagging algorithm used in random forest (Lu et al., 2017).

Parameters discussion in different classifiers

For an excellent classification algorithm, we should demand not only the accuracy, but also the robustness of its parameters. If the technique can only perform well with the parameters in a narrow area, the procedures of parameter tunning are usually tough and costly, increasing the difficulties for practical engineering project. Consequently, it is necessary to discuss the diagnostic accuracy with respect to the crucial parameters in the analysed classifiers. The experiment is conducted in the dataset A with 1797 load speed using FDF as input vectors. Figure 6(a) shows the diagnostic accuracy by random forest versus the number of trees and the number of random feature subset. Figure 6(b) is the zoomed change

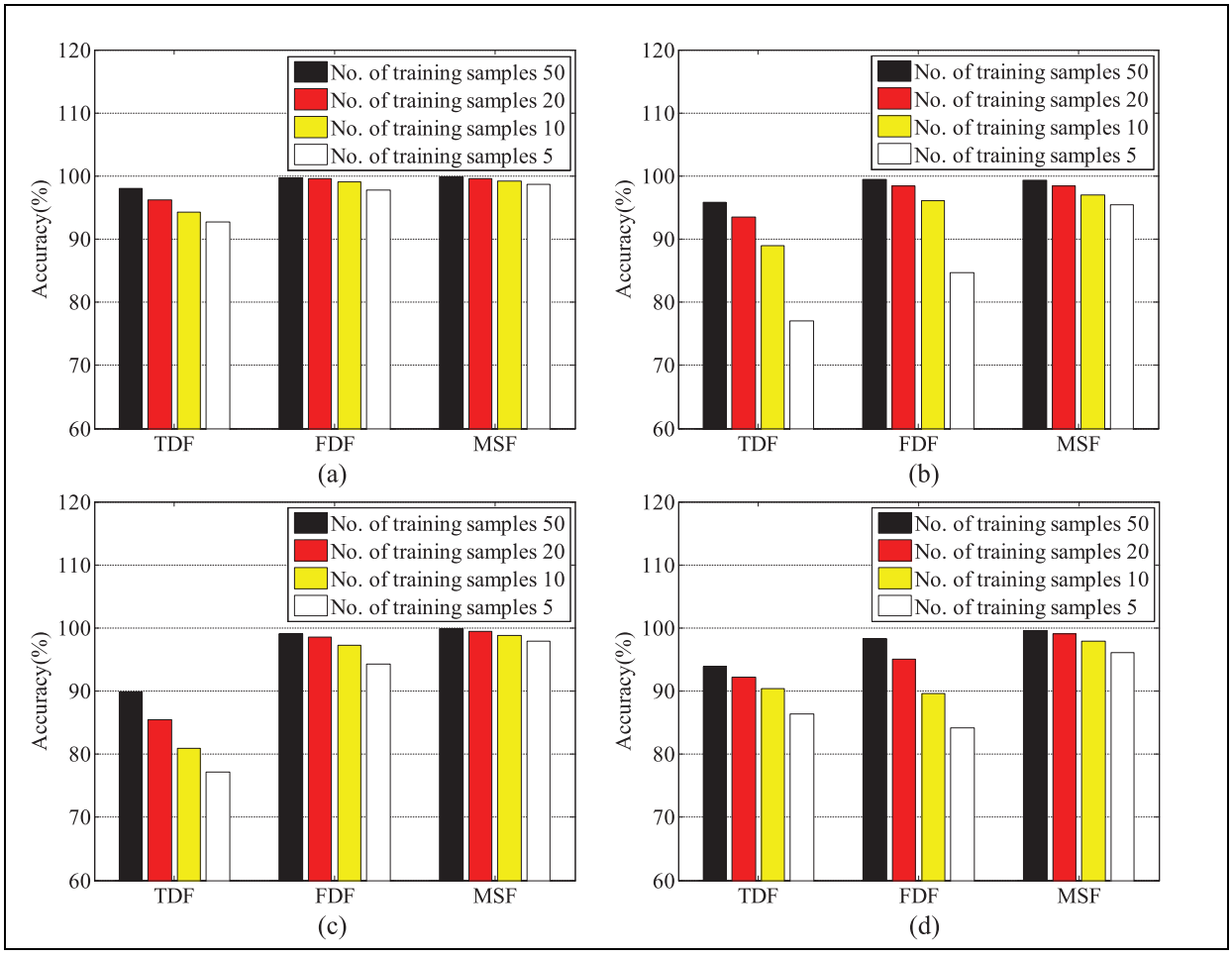

Figure 4. Comparisons of accuracy using different classifiers in dataset A: (a) random forest; (b) ELM; (c) PNN; (d) SVM.

Table 6. Comparison of diagnosis results under different noise environment.

Diagnosis rates (%)										
Classifier	SNR									
	12 dB	13 dB	14 dB	15 dB	16 dB	18 dB	20 dB	22 dB	24 dB	26 dB
Random forest	74.78	85.82	92.95	95.94	97.74	98.60	99.00	99.26	99.46	99.53
ELM	65.53	69.53	73.01	76.36	79.72	85.48	91.40	96.81	98.83	99.37
PNN	63.25	71.64	79.16	84.84	88.77	93.43	96.51	97.81	98.28	98.87
SVM	66.87	73.32	78.85	83.59	87.66	92.28	95.68	97.24	98.11	98.67

curves. It is clear to find, when the number of trees is over 5 with each m_{try}, the accuracy fluctuates slightly over 98% and all the standard deviations are lower than 0.3%, validating the algorithm is substantially robust to the number of trees. Some research has shown that the random forest can achieve a good recognition performance for most applications when the number of trees is below 50 (Dong et al., 2016), while more trees will guarantee the error convergence, that is the law of large numbers. In this work, the default 500 trees are utilized for comparisons. Furthermore, the diagnosis rate also has a good robustness for m_{try}. The recommended value $m_{try} = 3$ reaches a relatively high accuracy.

On the other hand, the accuracy curves of other 3 classifiers are described in Figure 7. From Figure 7(a), one can apparently observe the accuracy changes dramatically with the number of hidden neurons for ELM. As the number of hidden neurons grows, it does not mean a better performance will be obtained. In addition, the training set also has a

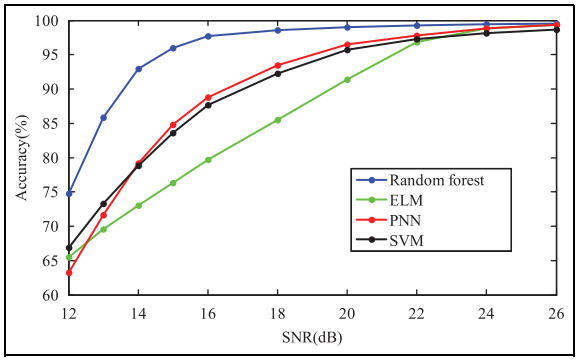

Figure 5. Diagnostic results of employed classifiers with different SNRs.

significant influence on the tendency of accuracy, meaning it may be hard to choose the optimal parameter directly. Likewise, from Figure 7(b), we can find a good recognition rate heavily relies on the proper spread value for PNN. Generally, a larger value can generate smooth PDFs while results in increased computational cost. In this work, we applied the 5 fold CV to guide the selection of the number of hidden neurons and the spread value. For SVM, penalty factor c and kernel function parameter g play a key role in achieving a significant higher accuracy. Figure 7(c) shows the contour of accuracy with varying parameter pairs. The three-dimensional curve is illustrated in Figure 7(d). The choice of the two parameters has received considerable attention in recent years and the heuristic optimization algorithms has been approved as the most widely used means of parameters selection, however, undoubtedly leading to computational burden.

Time consumption analysis

Here, the same dataset and input feature vectors in section 4.3 are adopted. The tests were performed on Windows 10 Edu (64 bits). ELM and PNN were run by use of MATLAB version 8.2. SVM and GA optimization procedure was conducted by using libSVM version 3.14 (Chang and Lin, 2011). Random forest was run with the help of open source toolbox of Abhishek Jaiantlal (University of Colorado Boulder, USA) on MATLAB platform. The hardware environment mainly includes a quad-core i7-4790 3.6Ghz CPU and DDR3-1600 16G memory. The average processing time of every model is presented in Table 7. In the first stage, compared with other three models, random forest possesses the marked advantages due to the robust parameters with no need of tunning. Although random forest fails to run faster than ELM and SVM in the predicting step, most of the tests can be finished within less than 5ms. In addition, it is easy to develop parallel computing for random forest on the basis of the characteristics of this algorithm.

Comparison with existing studies

To further verify the potential values of this scheme, a comparative analysis between this work and some published literatures adopting the same bearing data from Case Western Reserve University Bearing Data Center was made (see Table 8). Yang et al. (2007) employed 11 TDF in tandem with three fractal dimensions (FD) as input feature vectors and identified the bearing faults using a SVM with 10 fold cross validation. Li and Zhang (2011) presented a feature extraction method of supervised locally linear embedding projection (SLLEP) and a minimum-distance classifier in their diagnostic scheme. Dou and Zhou (2016) applied PNN to diagnose the feature vector which is composed of 6 TDF and 5 FDF. Luo et al. (2015) exploited chemical reaction optimization (CRO) algorithm to determine the structural parameters of SVM along with a feature extraction method via local characteristic-scale decomposition (LCD) and singular value decomposition (SVD). In our work, the novel random forest is utilized for fault diagnosis without complicated procedure of signal preprocessing, prior feature selection or parameters tunning. Random forest can still achieve a high diagnosis accuracy in cases where the training set is small, which demonstrates its powerful capacity for pattern recognition problems.

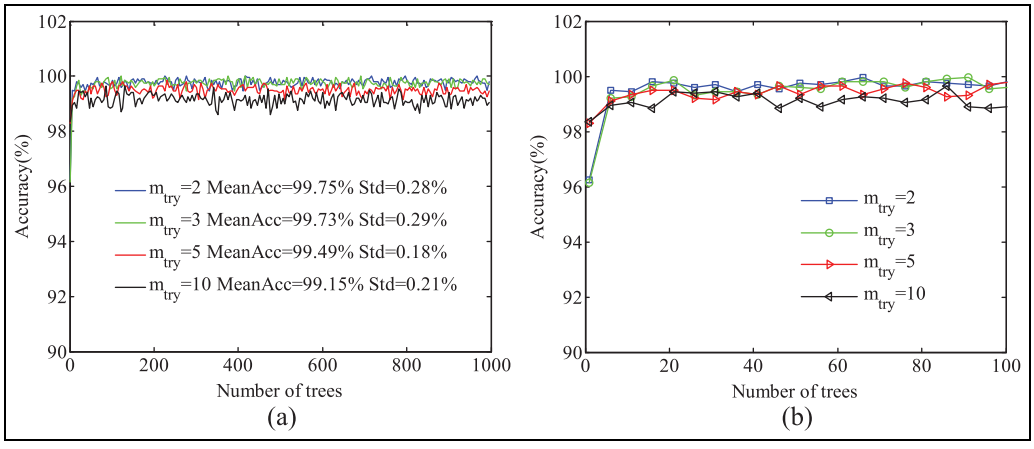

Figure 6. Diagnostic accuracy versus n_{try} and m_{try} in random forest.

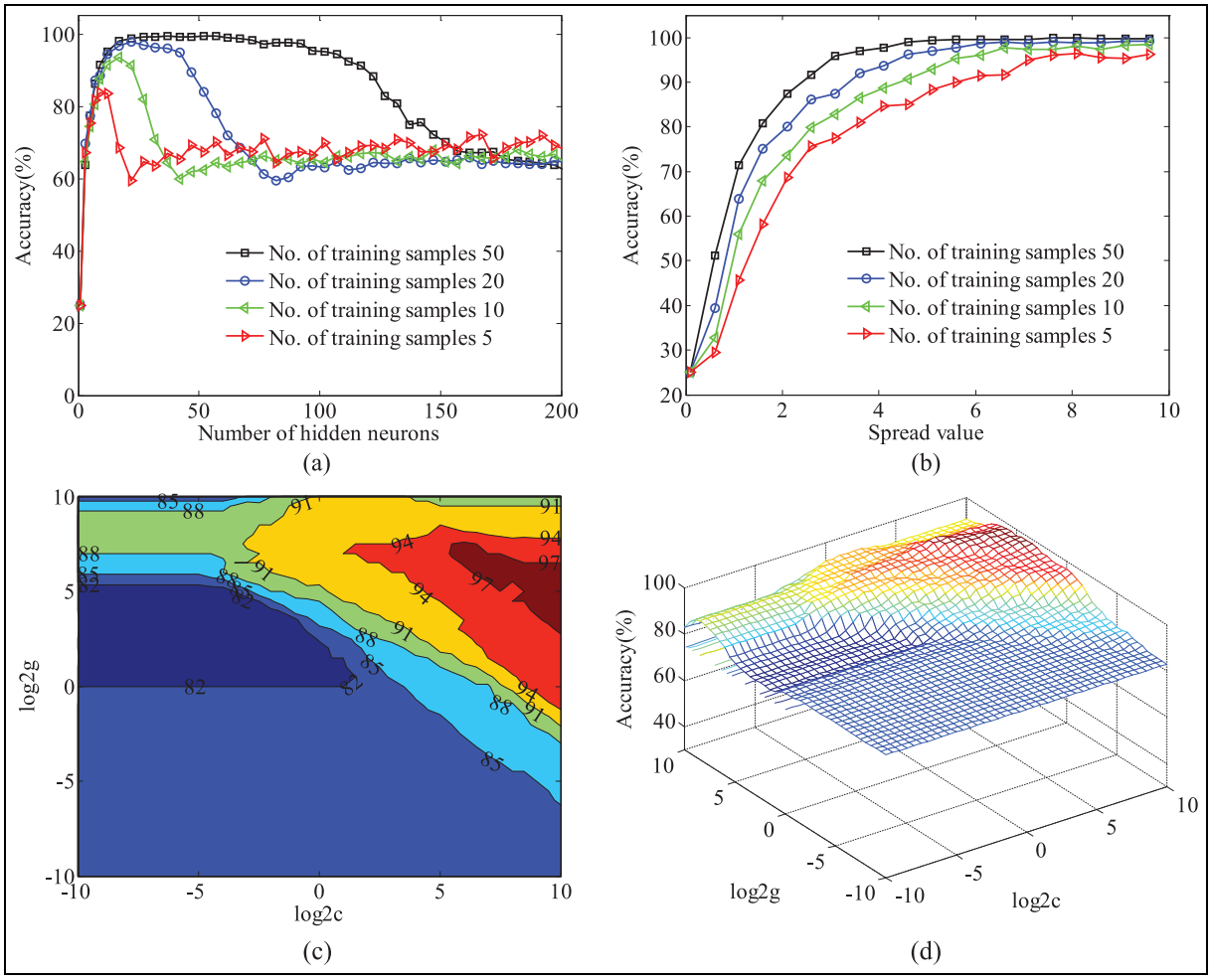

Figure 7. Diagnostic accuracy versus parameters in each classifier: (a) ELM; (b) PNN; (c) two-dimensional contour of SVM; (d) three-dimensional curve of SVM.

Table 7. Time consumption in different classifiers (Para.-Parameter tunning).

No. of training samples	Random forest		ELM		PNN		SVM	
	Model training	Prediction	Para. and model training	Prediction	Para. and model training	Prediction	Para. and model training	Prediction
50	62.64 ms	5.08 ms	3.12 s	0.60 ms	3.01 s	8.44 ms	2.85 s	0.50 ms
20	29.30 ms	4.74 ms	1.33 s	0.64 ms	2.94 s	7.70 ms	1.44 s	0.36 ms
10	18.46 ms	4.64 ms	0.79 s	0.52 ms	2.89 s	7.94 ms	0.94 s	0.18 ms
5	12.92 ms	4.76 ms	0.54 s	0.60 ms	2.53 s	5.46 ms	0.65 s	0.22 ms

Table 8. Comparisons between this work and some published literatures.

References	Feature	Classifier	No. of training samples	No. of testing samples	No. of classes	Diagnosis accuracy
Yang et al., 2007	TDF-FD	SVM	472	472	4	95.253%
Li and Zhang, 2011	SLLEP	Minimum-distance *classifier*	300	300	3	98.33%
Luo et al., 2015	LCD-SVD	CRO-SVM	240	80	4	100%
Dou and Zhou, 2016	TDF-FDF	PNN	240	80	4	94.38%
Present work	TDF	Random forest	200	200	4	98.04%
	FDF	Random forest	40	200	4	99.08%
	MSF	Random forest	40	200	4	99.24%

Conclusions

Currently, ANNs and SVM have received increasingly attention while random forest is relatively unfamiliar in this field. It is very necessary to attempt more efficient pattern recognition methods to facilitate the intelligent diagnosis of mechanical failures. In this work, we explored the performances of random forest, two advanced ANNs (ELM, PNN) and SVM with different features using two datasets from rotating machinery. The comparative analysis demonstrates the excellent property of random forest in terms of classification accuracy, stability and robustness to features. In particular, in a number of diagnostic experiments, we find that random forest has significant superiority when the training samples are limited. Additionally, compared with traditional methods, random forest performs significantly well in a wider range of SNR. Moreover, its only two parameters usually keep a weak sensitivity, which largely solves the drawback of parameter tuning in other pattern recognition algorithms and reduces the time consumption. The overall results indicate that random forest is satisfactorily able to deal with industrial tasks, leading to a promising application prospect.

Declaration of conflicting interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This research is supported by the National Natural Science Foundation of China (No. 11572167) and the project of State Key Lab of Power Systems (No. SKLD16Z12).

References

Bogoevska S, Spiridonakos M, Chatzi E, et al. (2017) A data-driven diagnostic framework for wind turbine structures: A holistic approach. *Sensors* 17(4): 1–28.

Breiman L (1996) Bagging predictors. *Machine Learning* 24(2): 123–140.

Breiman L (2001) Random forests. *Machine Learning* 45(1): 5–32.

Chang C-C and Lin C-J (2011) LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* 2(3).

Cheng G, Chen X-h, Shan X-l, et al. (2016) A new method of gear fault diagnosis in strong noise based on multi-sensor information fusion. *Journal of Vibration and Control* 22(6): 1504–1515.

Coomans D, Donald D, Hancock T, et al. (2006) Bagged super wavelets reduction for boosted prostate cancer classification of seldi-tof mass spectral serum profiles. *Chemometrics and Intelligent Laboratory Systems* 82(1–2): 2–7.

Dietterich TG (2000) Ensemble methods in machine learning. *Lecture Notes in Computer Science* 1875: 1–15.

Dong Y, Zhang Y, Yue J, et al. (2016) Comparison of random forest, random ferns and support vector machine for eye state classification. *Multimedia Tools and Applications* 75(19): 11763–11783.

Dou D and Zhou S. (2016) Comparison of four direct classification methods for intelligent fault diagnosis of rotating machinery. *Applied Soft Computing* 46: 459–468.

Dragomiretskiy K and Zosso D (2014) Variational mode decomposition. *IEEE Transactions on Signal Processing* 62(3): 531–544.

Freund Y and Schapire RE (1997) A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences* 55(1): 119–139.

Guo L, Gao H, Huang H, et al. (2016) Multifeatures fusion and nonlinear dimension reduction for intelligent bearing condition monitoring. *Shock and Vibration* (2016): 1–10.

Han T and Jiang D (2016) Rolling bearing fault diagnostic method based on VMD-AR model and random forest classifier. *Shock and Vibration* (2016): 1–11.

Han T, Jiang D and Wang N (2016) The fault feature extraction of rolling bearing based on EMD and difference spectrum of singular value. *Shock and Vibration*: 1–14.

Han T, Jiang D, Zhang X, et al. (2017) Intelligent diagnosis method for rotating machinery using dictionary learning and singular value decomposition. *Sensors* 17(4): 1–18.

Hansen LK, Liisberg C and Salamon P (1992) Ensemble methods for handwritten digit recognition. In: *Proceedings of the IEEE workshop on neural networks for signal processing II*, Copenhagen, Denmark, pp.333–342.

Ho TK (1998) The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20(8): 832–844.

Jiang D and Liu C (2011) Machine condition classification using deterioration feature extraction and anomaly determination. *IEEE Transactions on Reliability* 60(1): 41–48.

Lei Y, He Z and Zi Y (2011) EEMD method and WNN for fault diagnosis of locomotive roller bearings. *Expert Systems with Applications* 38(6): 7334–7341.

Li B and Zhang Y (2011) Supervised locally linear embedding projection (SLLEP) for machinery fault diagnosis. *Mechanical Systems and Signal Processing* 25(8): 3125–3134.

Liu C, Jiang DX and Yang WG (2014) Global geometric similarity scheme for feature selection in fault diagnosis. *Expert Systems with Applications* 41(8): 3585–3595.

Liu M, Wang M, Wang J, et al. (2013) Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: Application to the recognition of orange beverage and Chinese vinegar. *Sensors and Actuators B-Chemical* 177: 970–980.

Lu C, Wang Z-Y, Qin W-L, et al. (2017) Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification. *Signal Processing* 130: 377–388.

Luo S, Cheng J and Ao H (2015) Application of LCD-SVD Technique and CRO-SVM Method to Fault Diagnosis for Roller Bearing. *Shock and Vibration* (2015): 1–8.

Luo Y-M, Huang D-T, Liu P-Z, et al. (2016) A novel random forests and its application to the classification of mangroves remote sensing image. *Multimedia Tools and Applications* 75(16): 9707–9722.

Morizet N, Godin N, Tang J, et al. (2016) Classification of acoustic emission signals using wavelets and Random Forests: Application to localized corrosion. *Mechanical Systems and Signal Processing* 70–71: 1026–1037.

Schwenk H and Bengio Y. (2000) Boosting neural networks. *Neural Computation* 12(8): 1869–1887.

Sun P, Liao Y and Lin J (2017) The shock pulse index and its application in the fault diagnosis of rolling element bearings. *Sensors* 17(3): 1–23.

Svetnik V, Liaw A, Tong C, et al. (2003) Random forest: A classification and regression tool for compound classification and QSAR modeling. *Journal of Chemical Information and Computer Sciences* 43(6): 1947–1958.

Tang G, Luo G, Zhang W, et al. (2016) Underdetermined blind source separation with variational mode decomposition for compound roller bearing fault signals. *Sensors* 16(6): 1–17.

Tian Y, Ma J, Lu C, et al. (2015) Rolling bearing fault diagnosis under variable conditions using LMD-SVD and extreme learning machine. *Mechanism and Machine Theory* 90: 175–186.

Xu H and Chen G (2013) An intelligent fault identification method of rolling bearings based on LSSVM optimized by improved PSO. *Mechanical Systems and Signal Processing* 35(1–2): 167–175.

Yang J, Zhang Y and Zhu Y (2007) Intelligent fault diagnosis of rolling element bearing based on SVMs and fractal dimension. *Mechanical Systems and Signal Processing* 21(5): 2012–2024.

Yang R-M, Zhang G-L, Liu F, et al. (2016) Comparison of boosted regression tree and random forest models for mapping topsoil organic carbon concentration in an alpine ecosystem. *Ecological Indicators* 60: 870–878.

Yaqub MF and Loparo KA (2016) An automated approach for bearing damage detection. *Journal of Vibration and Control* 22(14): 3253–3266.

Yunusa-Kaltungo A, Sinha JK and Nembhard AD (2015) A novel fault diagnosis technique for enhancing maintenance and reliability of rotating machines. *Structural Health Monitoring* 14(6): 604–621.

Zhou B and Cheng Y (2016) Fault diagnosis for rolling bearing under variable conditions based on image recognition. *Shock and Vibration* (2016): 1–14.