

An integrated generalized discriminant analysis method and chemical reaction support vector machine model (GDA-CRSVM) for bearing fault diagnosis

Nguyen, V.H.^{a,b,c,*}, Cheng, J.S.^{a,b}, Thai, V.T.^{a,b,c}

^aState Key Laboratory of Advanced Design and Manufacturing for Vehicle Body, Hunan University, Changsha, China

^bCollege of Mechanical and Vehicle Engineering, Hunan University, Changsha, China

^cMechanical Engineering Department, Hanoi University of Industry, Hanoi, Vietnam

ABSTRACT

An expert technique in bearing fault diagnosis is proposed for the identification of actual status. A new diagnosis method based on a two-stage hybrid modality in integrating generalized discriminant analysis (GDA) with the chemical reaction support vector machine (CRSVM) classification model, named GDA-CRSVM, is proposed. The GDA reduces high-dimensional data to a more compact data, which serves an optimized CRSVM classification model with input data, in which a support vector machine (SVM) classifier model with the best parameters are selected by the meta-heuristic chemical reaction optimization algorithm (CRO) to build an optimized CRSVM classification model. The implementation of the new proposed method is based on a multi-aspect feature (MAF) set that presents most of the actual aspects of the complex vibration signal. The MAF set is collected from the statistical features in time-domain, frequency-domain, and time-frequency domain features are extracted by local characteristic-scale decomposition (LCD). Experiments have been conducted on two bearing vibration datasets by the expert technique in the bearing fault diagnosis. Results shown that the effectiveness of GDA-CRSVM in terms of classification accuracy and execution time.

© 2017 PEI, University of Maribor. All rights reserved.

ARTICLE INFO

Keywords:

Bearing fault
Expert fault diagnosis technique
Chemical reaction support vector machine (CRSVM)
Multi-aspect feature set
Generalized discriminant analysis (GDA)

*Corresponding author:

hung2009hau@gmail.com
(Nguyen, V.H.)

Article history:

Received 8 June 2017
Revised 23 October 2017
Accepted 26 October 2017

1. Introduction

The bearing is a key component of rotating machinery and is closely allied to system operation. Any failure of bearing may cause unsafe conditions for the operator and inefficient operation, stopping work may also affect associated systems. Hence, advanced fault diagnosis methods in the mechanical maintenance field are a focus of interest to many researchers. These methods can be summarized into several consecutive steps aimed at identifying patterns of fault status. The first step is acquisition vibration data, which may need pre-processing such as denoising or removing artefacts. The second step is feature extraction step to get the most important information. Then, these features are transformed into the pattern diagnosis model to classify patterns. Finally, the pattern diagnosis model determines the pattern type to which the particular fault signal belongs.

One the most important actions for fault diagnosis technique is feature extraction. An effectiveness feature set needs to contain the most salient features, beneficial features of the classification stage. This paper focuses on a multi-aspect feature extraction (MAF), which many actual

aspects of the complex vibration signal. MAF is based on statistical features in time-domain, frequency-domain, which directly represent the outward aspects of signal, and time-frequency domain features, which represent the intrinsic aspects deeply hidden in the vibration signal. Features in the time-frequency domain are especially extracted by local characteristic-scale decomposition (LCD), a method that becomes superior in running time, restraining the end effect and relieving mode mixing [1, 2]. The MAF set can be extracted from the original vibration signal as a high-dimensional feature vector including seven features that represent the aspect in the time-domain, eight features that represent the aspect in frequency-domain, and five features that represent the aspect in the time-frequency domain. The obtained MAF can provide extremely adequate information on various bearing conditions to make an effective diagnosis of performance.

Normally, a feature set of an original vibration signal can also provide more handy information. Increasing the problem of using these features efficaciously in a way that would interfere with the classification stage, such as the computational burden, the processing time is slower, the classification accuracy results are poorer. This paper aims at making classification performance more effective. The high-dimensional feature dataset that was extracted from vibration signals is mapped onto a new feature space to discover the intrinsic structure in these non-linear high-dimensional data and to obtain a more compact feature dataset in a lower dimension. Recently, dimensionality reduction approaches have aroused great interest in the fault diagnosis research field. Principal component analysis (PCA) [3, 4] is one of the most traditionally used, along with multi-dimensional scaling (MDS) [5] and linear discriminate analysis (LDA) [6, 7]. However, while these approaches are remarkably effective on linear data, they may not adequately handle complex non-linear data. This may be cause of low accuracy or misjudgement of a fault diagnosis with non-linear data. To expand the field of non-linear data of LDA, the generalized discriminant analysis (GDA) method was proposed by Baudat and Anouar (2000)[8]. The main idea is to project the input space into an advantageous feature space, where variables are nonlinearly related to the input space. According to the current literature, the GDA method has not been previously applied for fault diagnosis. In the case of medicine [9] and imaging [10], there are previous reference works. An important contribution in this study is the introduction of the GDA method in the discovery of the intrinsic structure of the non-linear high-dimensional feature dataset. Its combination with a classifier model make it effective for classification purposes.

Support vector machine (SVM) based on statistical learning theory is a new machine learning algorithm proposed by Vapnik *et al.* SVM is a powerful supervised machine learning tool and is used in a number of applications such as pattern recognition [11], time-series forecasting [12], robotics [13] and diagnostics [14]. When SVM is used, one should remark that the optimal parameters play a leading role for forming a classification model with high classification efficiency, thus creating something that has aroused the great interest of researchers for the selection of optimal parameters. Recently, several evolutionary based algorithms such as the genetic algorithm (GA) [15], particle swarm optimization (PSO) [16], ant colony optimization (ACO) [17], the simulated annealing algorithm (SA) [18] have been used to optimize the SVM parameters and have also shown promising ability as learning algorithms that can be utilized for diagnosis purposes [19]. However, their performance may vary from one object to another in fault diagnosis and may not be suitable for the different statuses of roller bearings. Besides, the efficiency of these optimization algorithms is characterized by the procedure used for selection the parameters, which requires a deep knowledge of the use of algorithms. The recent chemical reaction optimization (CRO) algorithm, which is a novel computational method, is one optimization of the found meta-heuristics introduced in 2010 [20]. CRO is an evolutionary optimization technique, which is comprehended from the nature of chemical reaction. It performs very well in solving optimization problems in a very short time. In a short period, there have been a few applications of CRO to the recognition field, data mining, classification rule [21, 22], and efficiency has been demonstrated. Indeed, CRO has been applied to solve complex problems successfully, has outperformed many existing evolutionary algorithms in most of the test cases. A guideline can be found in the tutorial introduced in [23] to help readers implement CRO for optimization problems. Motivated by the capability of CRO, an important contribution in this study is the authors' aim to use the CRO algorithm to select the best parameters of the SVM model, which is a fre-

quently used diagnosis technique called CRSVM. Then, the optimized CRSVM model is used for bearing fault diagnosis.

Finally, in this paper, a two-stage hybrid modality for integrating GDA with CRSVM is introduced, called GDA-CRSVM. The proposed GDA-CRSVM method is based on an expert technique that aims to exploit the highest identification accuracy in the fault diagnosis of roller bearings based on an MAF set. GDA is first used to reduce the high-dimensional feature set that acts as the data pre-processing for classifier model. Then, the obtained feature set provides the CRSVM model with input data. For exploration, experiments have been conducted on two bearing vibration datasets with different conditions based on the GDA-CRSVM method. Moreover, the acquired vibration signals have been analysed to extract the MAF set. It is remarkable that the performance of the GDA-CRSVM method is significantly better than that of the other methods and showed the most accurate results for classification purposes along with superior execution time.

The paper is constructed as follows. Section 2 presents materials and methods for bearing fault diagnosis. In Section 3, the GDA-CRSVM method is proposed by a two-stage hybrid modality by integrating the GDA method with the CRSVM classification model for the expert fault diagnosis technique. In Section 4, we present experiments for bearing fault diagnosis, where vibration data is acquired for roller bearings, MAF is used to extract vibration signals, and the actual fault statuses are identified by our proposed method. Section 5 is the conclusion. Acknowledgments and a list of references follow.

2. Materials and methods

2.1 Generalized discriminant analysis (GDA) method

Dimensionality reduction can be done by feature transformation to a low-dimensional data space once features have been extracted from the vibration signals. The purpose of the dimensionality reduction is to select the most superior features of the original feature set, which can provide dominant actuality-related information. Irrelevant or redundant factors must be discarded to improve classification performance, to avoid problems with dimensionality. Therefore, the GDA method is presented and used to select the superior features from the original feature set [8].

The objective of GDA is to find mapping for the input feature set into a lower dimensional space/ new space. The ratio of centre-class partner P_b to within-class partner P_w can be maximized [8]. A set of input patterns S of training features-set can be given as:

$$S = \sum_{c=1}^C S_c X_{ci} \quad c = 1, 2, \dots, C; i = 1, 2, \dots, S_c \quad (1)$$

This is a C-class problem, S_c is the number of samples in class c . The mapping $\psi: R^T \rightarrow T$ is non-linear for training patterns in the new space, thus $X \rightarrow \psi(X_{ci}), c = 1, 2, \dots, C; i = 1, 2, \dots, S_c$ is represented.

The center-class partner P_b to within-class partner P_w of the training feature set can be calculated as below:

$$P_w = \frac{1}{C} \sum_{c=1}^C \frac{1}{S_c} \sum_{i=1}^{S_c} \psi(X_{ci}) \psi^T(X_{ci}) \quad (2)$$

$$P_b = \frac{1}{C} \sum_{c=1}^C (\mu_c - \mu)(\mu_c - \mu)^T \quad (3)$$

We have to calculate the eigenvalues and eigenvectors $v, v \in T \setminus \{0\}$ to satisfy the equation:

$$\lambda P_w v = P_b v \quad (4)$$

$$\lambda = \frac{v'P_b v}{v'P_w v} \tag{5}$$

The eigenvectors v are combinations of $\psi(X_{ci})$ elements and the existing coefficients $f_{ci}, c = 1, 2, \dots, C; i = 1, 2, \dots, S_c$ such that:

$$v = \sum_{c=1}^C \sum_{i=1}^{S_c} f_{ci} \psi(X_{ci}) \tag{6}$$

To simplify, we can write the coefficient vector as below:

$$f = (f_{ci})_{\substack{c=1,2,\dots,C \\ i=1,2,\dots,S_c}} \tag{7}$$

Further, let us consider this vector. We used the kernel technique in the new space. Using the dot product of a sample m from class g and another sample n from class p , the dot product $(k_{mn})_{gp}$ gives the following:

$$(k_{mn})_{gp} = \psi(X_{gm}) \cdot \psi(X_{pn}) \tag{8}$$

First, let K be a $(N \times N)$ matrix defined in terms of the class elements by $(K_{gp})_{\substack{g=1,2,\dots,C \\ p=1,2,\dots,C}}$. In the new space, the K matrix is represented as below:

$$K = (K_{gp})_{\substack{g=1,2,\dots,C \\ p=1,2,\dots,C}} \tag{9}$$

where K_{gp} is a $(S_g \times S_p)$ matrix:

$$(K_{gp}) = (k_{mn})_{\substack{m=1,2,\dots,S_g \\ n=1,2,\dots,S_p}} \tag{10}$$

Then, a $(N \times N)$ matrix A is introduced, A is defined as:

$$A = (A_c)_{c=1,2,\dots,C} \tag{11}$$

where (A_c) is a $(S_c \times S_c)$ matrix with all terms equal to $1 / S_c$.

Finally, from the Eqs. 2, 3, 6 and 4, we found the inner product with vector $\psi(X_{ci})$ on both sides.

$$\lambda K K v = K A K v \tag{12}$$

There, v represents a column vector with values $v_{ci}, c = 1, 2, \dots, C; i = 1, 2, \dots, S_c$.

The solution of Eq. (12) is satisfactory when the eigenvectors of matrix $(K K)^{-1} K A K$ are calculated.

2.2 Optimal classification model

This section emphasizes the superiority of the CRO algorithm, which is then applied to select the best parameters of the SVM. These parameters play a leading role for building an optimal CRSVM classification model. The obtained CRSVM model can be used for fault diagnosis of bearing components, combining the input feature set to become the expert classifier model with high classification accuracy, stability and effectiveness of performance. Fig. 1 depicts the flowchart for using the CRO algorithm to select parameters of the SVM model.

Principle of SVM

Support Vector Machine (SVM) were introduced by Vapnik [24]. The SVM classifier is designed for classification tasks with two-class datasets. The data are separated by a hyperplane in order to maximize distance. The separating hyperplane is defined by the closest points of the training dataset, which are called support vectors. The details of SVM are presented in [14]. The parameter pair (C, γ) in the RBF kernel function (the penalty parameter of C and width parameter of γ) plays an important part in the classification purpose. The parameter pair values cover a broad

range and controls the generalization capability of SVM. The best selection of parameter pair is important and necessary to training the SVM classifier. In this work, values of C , γ are selected using the CRO algorithm for the best performance for accurate bearing fault diagnosis.

The optimized SVM model based on CRO

To fulfil the aim to build an optimal classification model, CRSVM, the CRO algorithm is used to exploit the best parameter pair (C , γ) of the SVM. The obtained classification model is employed to identify the bearing conditions.

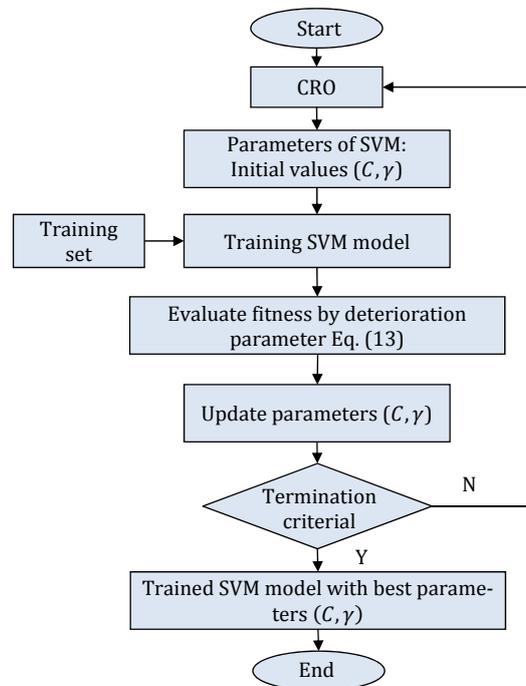


Fig. 1 The architecture of CRSVM classification model

Chemical reaction optimization algorithm

The CRO algorithm is introduced in 2010 [20] in the finding meta-heuristics of computational method which is the efficient optimization technique that enjoys the advantages of previous genetic algorithms and simulated annealing. This algorithm is not only inspired by the elementary chemical reactions, that is different from evolutionary algorithms motivated by biological evolution, but also easily constructed by defining the agents and the energy directional scheme. Consequently, algorithm has been deployed for different problems and has been successfully used to counteract complicated problems, outperforming other prevailing evolutionary algorithms in test conditions.

Furthermore, CRO algorithm carried out parallel of sub-reaction steps in the optimal process which benefits the minimizing for accomplishing time. Algorithm accomplishes local, global search with elementary reactions. In these, four types of elementary reactions are included: (1) on-wall ineffective collision, (2) decomposition, (3) inter-molecular ineffective collision, and (4) synthesis. In fact, each reaction is the interaction (the combination and variation) of molecules at a high energy level to become new products with a low energy level, in a stable status. The details of the CRO algorithm can be seen in [20, 23]. Motivated by the superior capability of CRO, the authors applied select parameters of the SVM model.

The solution to optimize the CRO algorithm involves using the natural chemical reaction of reactants to solve problems. The beginning of the algorithm establishes initial reactants, which play an important role in a solution. Then, the reactants react and produce four types of reactions. The algorithm is stopped when the termination criterion reaches final status, when no

more reactions can take place. In this work, the parameter pair (C, γ) of SVM is set as reactants in four types of reactions. According to this, the CRO algorithm consists of the following steps [21]:

- Step 1: Initialize the parameters.
- Step 2: Set the initial reactants and evaluate enthalpy.
- Step 3: Apply chemical reactions to reactants.
- Step 4: Update and select reactants.
- Step 5: Go to step 3 if termination criterion not satisfied.
- Step 6: Output reactant with best enthalpy.

Classification model CRSVM

The optimization parameter pair (C, γ) of the SVM can be obtained using the CRO algorithm. This CRO algorithm conducts stochastic searches using a population of molecules, each of which represents a possible solution to a problem. A population includes a finite number of molecules, with each molecule defined by an evaluating mechanism to obtain its potential energy.

The principled training phase of the CRSVM model includes seven main steps, which are implemented as follows:

- Step 1: Training and testing datasets are prepared after feature extraction from original vibration signals.
- Step 2: This is initialization step. The initial C, γ parameters are random for SVM. Set the maximum iteration number t_{max} . Set the iterative variable: $t = 0$ and perform the training process for the next steps. The parameters for this optimization algorithm are iteration $t_{max} = 50$, population size $pop = 5$, upper bound $u_p = 2^{12}$ and lower bound $l_p = 2^{-12}$.
- Step 3: Increase the iteration variable by set $t = t + 1$
- Step 4: Deterioration evaluation. The deterioration function is employed to evaluate the quality of every element. Eq. (13) shows the classification accuracy of an SVM classifier:

$$deterioration(\%) = \frac{N_{false}}{N_{\Sigma}} 100 \% \quad (13)$$

where N_{false} is false classified samples, N_{Σ} is total samples in the testing process. The desirable value is small for high classification accuracy.

- Step 5: Stop criteria checking. If the deterioration function satisfies Eq. (13) or iteration is maximal, go to step 7. If not, go to the next step.
- Step 6: Update the new C, γ parameters based on conditions. Go to step 3.
- Step 7: End of the training procedure. Fitting parameters are optimal output values.

The efficient search capability of the chemical reaction algorithm is incorporated with the generalization capability of SVM to bring out synergies of the classification accuracy. The architecture for CRSVM is presented in Fig. 1. Each reactant represents the candidate solution for the model, which includes the parameters C, γ .

3. An expert technique based on the proposed GDA-CRSVM method

In this section, the authors propose a new diagnosis method based on a two-stage hybrid modality for integrating GDA with the CRSVM, called the GDA-CRSVM. This takes special consideration of improved computational time, reduction of calculation memory and enhanced recognition accuracy of fault data. The methodology of dimensionality reduction (GDA) is close, which can obtain the total intrinsic emergent information of the original high-dimensional feature set. Combined, the optimized CRSVM model can obtain effective classification performance.

The process of the proposed method consists of two parts: dimensionality reduction and pattern recognition. First, the authors used the GDA method to reduce the high-dimensional fault feature dataset by taking out the most responsive features to produce a low-dimensional feature set. The obtained feature set increased the overall reliability of the fault diagnosis technique as well as the accuracy of diagnosis of an actual fault condition. Second, the reduced feature is

served as input to the optimized classification mode, which was elaborately optimized by the CRO algorithm based on the SVM, namely CRSVM.

Overall, the expert bearing fault diagnosis technique based on the GDA-CRSVM hybrid method aims to further improve fault diagnosis performance and ensure diagnosis reliability. It is presented in Fig. 2. From this figure, the technique includes four main steps as follows: Step 1 is vibration signal acquisition, Step 2 is MAF extraction, Step 3 is dimensionality reduction, Step 4 is pattern classification. The implementation process is described below:

Step 1: In the first step, the original vibration signals are acquired from acceleration sensors.

Step 2: Feature extraction is an urgent step of the diagnosis process. The extracted feature set represents important information about the actual bearing conditions, which governs the final results of the diagnosis process. This feature set contains the time-domain, frequency-domain features, and time-frequency domain features are extracted by the LCD method, which is used to form the MAF set of the original vibration signal.

Step 3: The GDA method is used to discover the intrinsic structure of the MAF set. The reduced feature set as a low-dimensional feature vector has more effective classification performance, such as reduced calculation memory, computational time, and the best classification accuracy results.

Step 4: The reduced feature set is divided into training set and testing set. Each low-dimensional training sample in its respective class labelled as the training set is used to discover the best parameter pair (C, γ) of the SVM by CRO. The obtained CRSVM classifier model is then used to recognize the samples in the testing set. For reliable diagnosis capability, the diagnosis technique based on this proposed GDA-CRSVM method is applied for bearing fault diagnosis.

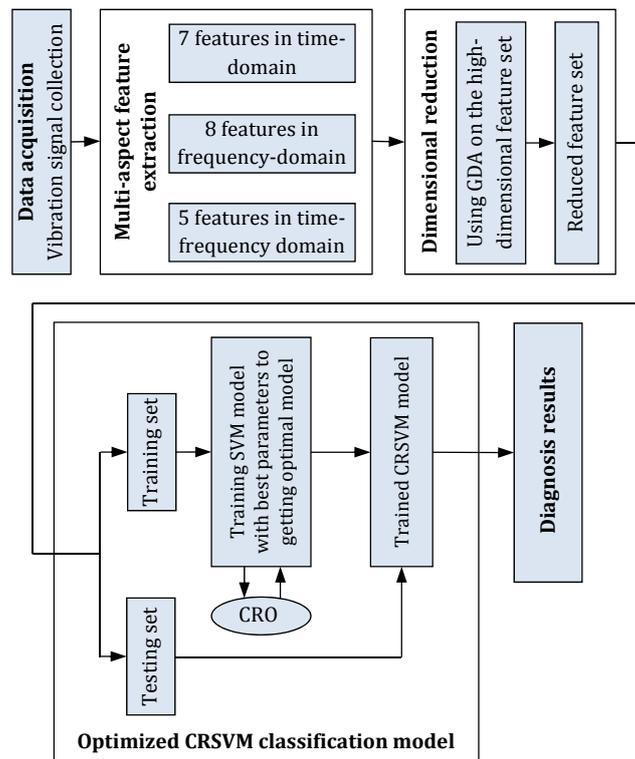


Fig. 2 Struct diagram of expert fault diagnosis technique

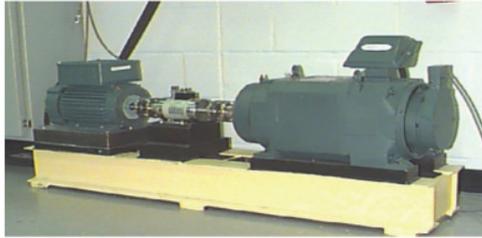


Fig. 3 Schematic of the experimental

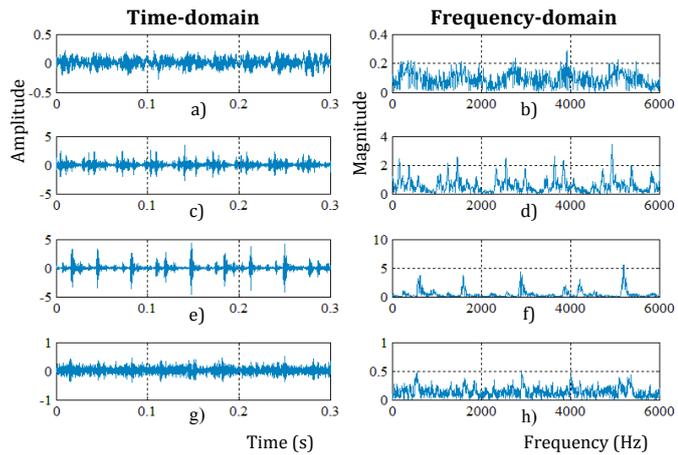


Fig. 4 The bearing conditions in the time domain and frequency domain

- a, b) Normal bearing
- c, d) Inner bearing fault
- e, f) Outer bearing fault
- g, h) Roller element fault

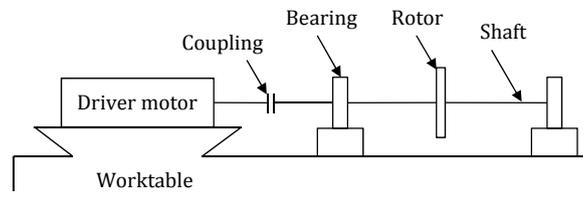


Fig. 5 The schematic drawing of test rig

4. Results and discussion for bearing fault diagnosis performance case study

In this section, two datasets of bearing conditions were collected. They are used for the experiments based on the proposal of proposed expert technique in actual status identification.

4.1 Data acquisition

The first dataset is vibration signals of bearing component fault cross from the Bearing Data Center at Case Western Reserve University (Loparo, 2013). Fig. 3 shows the experimental setup model, which consists of a two-HP reliance electric motor, a torque transducer and a dynamometer. The test bearings were installed on the motor shaft, which was loaded by dynamometer. The accelerometer data at DE were used as original signals for the detection of four bearing conditions: healthy bearing (HB), inner race (IR) fault, outer race (OR) fault, and rolling element (RE) fault. A defect was tested on the IR, OR, RE of test bearing using defect sizes 0.5334 mm (0.021 inches) in diameter with a depth of 0.2794 mm (0.011 inches) generated by electro-discharge machining. Four vibration signal datasets were acquired from the bearings with a sampling frequency of 12 kHz, tested under motor load of two-HP at a speed of 1750 RPM. Fig. 4 presents the vibration signals in time-domain, in frequency-domain from four signal samples of the bearing conditions. In each fault pattern, 25 samples were acquired from vibration signals. Each sample includes 4096 continuous data points in the time-domain. The results obtained groups with 100 vibration signals at various bearing conditions.

To further test the efficacy of the fault detection technique, the second dataset was acquired from test rig, as shown in Fig. 5 [25]. The bearing faults were introduced by laser cutting in the IR or RE with slot width of 0.15 mm and depth of 0.13 mm, respectively. The three experimental conditions tested were healthy bearing (HB), bearing with IR fault and bearing with RE fault. A total of 15 acceleration measuring signals with sampling frequency of 4096 Hz were acquired for each bearing condition.

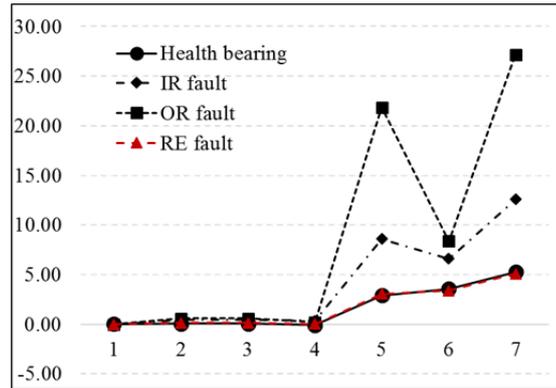


Fig. 6 The time-domain features of a bearing sample in different conditions

4.2 Multi-aspect feature extraction

The MAF set extracted from original vibration signals plays an important role for the achievement of the diagnosis model. The MAF includes time-domain features, frequency-domain features and features in the time-frequency domain, which are considered a high-dimensional feature vector.

- Time-domain features

The signal was analysed to extract seven time-domain statistical features TD_1 to TD_7 . Table 1 shows the seven feature definitions. In this table, the first five dimensions $TD_1 - TD_5$ reflect the vibration amplitude and energy in the time-domain, the last two dimensions TD_6, TD_7 are the crest factor and clearance factor, which represent the time-series distribution of the signals. Fig. 6 describes the time-domain features of the bearing samples in the different conditions.

Table 1 The feature definition equations in time-domain

No.	Time Domain (TD)		Remark
	Feature	Equation	
1	Mean	$TD_1 = \frac{1}{M} \sum_{i=1}^M x_i$	x_i is a vibration signal in time domain $i = 1, 2, \dots, M$, M is the number of data points.
2	Standard deviation	$TD_2 = \sqrt{\frac{1}{M-1} \sum_{i=1}^M (x_i - \bar{x})^2}$	
3	Root Mean Square	$TD_3 = \sqrt{\frac{1}{M} \sum_{i=1}^M x_i^2}$	
4	Skewness	$TD_4 = \frac{1}{TD_2^3(M-1)} \sum_{i=1}^M (x_i - \bar{x})^3$	
5	Kurtosis	$TD_5 = \frac{1}{TD_2^4(M-1)} \sum_{i=1}^M (x_i - \bar{x})^4$	
6	Crest factor	$TD_6 = \frac{\max x_i }{TD_3}$	
7	Clearance Factor	$TD_7 = x_{max} \left[\frac{1}{M} \sum_{i=1}^M \sqrt{ x_i } \right]^{-2}$	

- Frequency-domain features

Some signal information is also described in the frequency-domain and reveals information about the demodulation spectrum, amplitude frequency or distribution, which cannot be found in the time-domain. To extract these features, the Hilbert transform was first used to transform the vibration signals. Eight frequency-domain features FD_1 to FD_8 were then extracted from the

frequency spectrum of vibration signals, as shown in Table 2. The obtained features $FD_1 - FD_4$ describe the convergence or divergence of the spectrum power, FD_5, FD_6 indicate a change in position frequency. The spectrum power energy degree can centralize or de-centralize as described by parameter FD_7 and FD_8 can quantitatively measure disorder in the system. The frequency-domain features of the bearing conditions in an inner race fault, outer race fault, roller element fault and healthy fault are depicted in Fig. 7. The features of the conditional outer race fault and inner race fault are shown especially clearly.

- Time-frequency domain features

LCD is a self-adaptive method used in data decomposition. LCD has been successfully used to analyse non-linear, non-station signals, especially fault signals [2]. Obviously, LCD can extract the deeply hidden features in bearing fault data, as these features are very hard to distinguish only from the time-domain and frequency-domain statistical characteristics. In this study, the authors investigated the energy correlation coefficients between the first several intrinsic scale components (ISC), which were decomposed by the LCD method. These coefficients can reveal the original vibration signal in the time-frequency amplitude and distribution view, which is well and good for accurate diagnosis of a bearing fault. Further, any complex vibration signal $x(t)$ is decomposed into ISC and the residue by LCD, as in the equation below [1]:

$$x(t) = \sum_{i=1}^n ISC_i(t) + r_n(t) \tag{14}$$

where $ISC_i(t)$ is the i^{th} ICS of original signal obtained by LCD, the residue is $r_n(t)$.

Table 2 The feature definition equations in frequency-domain

No.	Frequency domain (FD)		Remark
	Feature	Equation	
1	Mean frequency	$FD_1 = \frac{1}{M} \sum_{i=1}^M p_i$	P_k is the energy probability distribution defined as: $P_k = p_i ^2 / \sum_{i=1}^M p_i ^2$
2	Standard deviation frequency	$FD_2 = \frac{1}{M-1} \sum_{i=1}^M (p_i - FD_1)^2$	
3	Skewness frequency	$FD_3 = \frac{1}{M} \frac{\sum_{i=1}^M (p_i - FD_1)^3}{\sqrt{\left(\frac{1}{M-1} \sum_{i=1}^M (p_i - FD_1)^2\right)^3}}$	where p_i is the spectrum of $x(i)$ vibration signal, $i = 1, 2, \dots, M$, M is the number of spectrum lines.
4	Kurtosis frequency	$FD_4 = \frac{1}{M} \frac{\sum_{i=1}^M (p_i - FD_1)^4}{\left(\frac{1}{M-1} \sum_{i=1}^M (p_i - FD_1)^2\right)^2}$	
5	Frequency centre	$FD_5 = \frac{\sum_{i=1}^M f_i p_i}{\sum_{i=1}^M p_i}$	f_i is the frequency value of the i^{th} spectrum line.
6	Root mean square frequency	$FD_6 = \sqrt{\frac{\sum_{i=1}^M f_i^2 p_i}{\sum_{i=1}^M p_i}}$	
7	Root variance frequency	$FD_7 = \sqrt{\frac{\sum_{i=1}^M (f_i - FD_5)^2 p_i}{M}}$	
8	Shannon Entropy	$FD_8 = - \sum_{i=1}^M p_k \log P_k$	

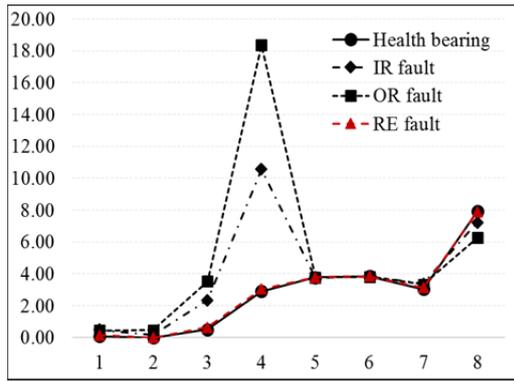


Fig. 7 The frequency-domain features of a bearing sample in different conditions

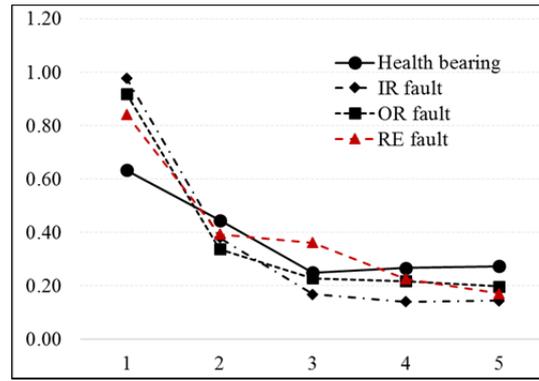


Fig. 8 Time-frequency domain features

The first several ISCs contain almost all valid fault information that characterizes the original signal. There, ISCs have higher energy than the rest. In this work, the first five ISCs were used to calculate the correlation of energy with the original vibration signal. These five energy correlation coefficients formed a feature subset representing the time-frequency domain features of bearing fault status. These following steps were taken:

- Step 1: The original vibration signals were collected from fault samples of the roller bearing.
- Step 2: The LCD method decomposed the vibration signals into ISCs. The first h ISCs were chosen.
- Step 3: The energy $E(ISC_i)$ of the first h ISCs was calculated, as related in Eq. 15 and Eq. 16:

$$E(ISC_i) = \sum_{n=1}^M a_k(t_n) \tag{15}$$

where $n = 1, 2, \dots, M$, M is data length of i^{th} ISC, t_n is the amplitude of point m in the ISC_i component, and $a_k(t)$ is the obtained amplitude of the i^{th} ISC by Hilbert transform:

$$ISC_i(t) = a_i(t)e^{i \int \omega_i(t) dt} \tag{16}$$

Step 4: A feature vector F was constructed with the energy correlation coefficient as element:

$$F = [E'_1, E'_2, \dots, E'_k] \tag{17}$$

where $E'_i, i = 1, 2, \dots, k$ are the energy correlation coefficients.

$$E'_i = \frac{E_i}{\sum_{i=1}^m E_i} \tag{18}$$

Fig. 8 also shows time-frequency domain features of the bearing conditions. The deeply hidden features in the bearing fault signal were extracted by LCD. The feature values show that the energy level of ISCs gradually decreased.

The obtained time-frequency domain features were added into the feature set and thus the complete MAF was formed containing 20 features. The obtained MAF represents a bearing fault condition as a high-dimensional feature vector that serves the GDA-CRSVM method with input data.

4.3 Diagnosis analysis based on GDA-CRSVM

The high-dimensional MAF discovered non-linear characteristics by the GDA method as dimensionality reduction, which can be given as a low-dimensional feature set. The obtained low-dimensional feature set was then randomly divided into a training-testing partition, 70 % : 30 %. Finally, as mentioned in Section 3, the training set was used to train the optimal CRSVM diagnosis model in actual bearing conditions. The obtained optimal diagnosis model was employed to classify the samples in the testing set.

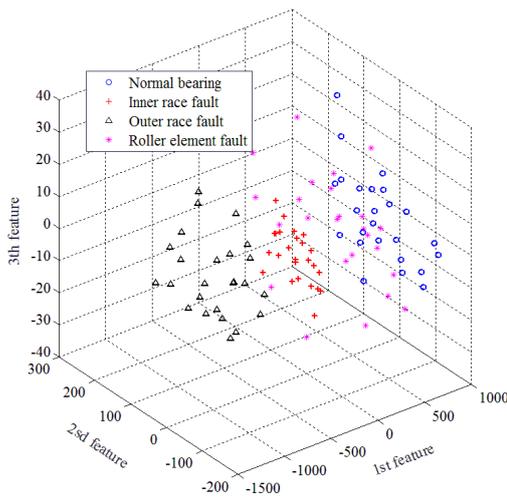


Fig. 9 Scatter plot for the reduced feature set by PCA

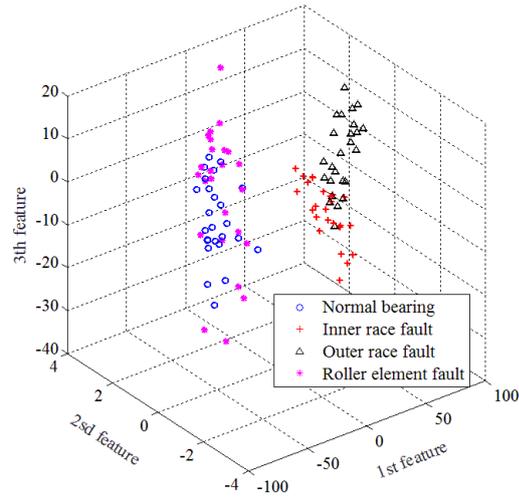


Fig. 10 Scatter plot for the reduced feature set by LDA

Feature-dimensional reduction by GDA method

Practically, the high-dimensional feature dataset involves too much memory for parameters, which results in a complicated and inefficient for classification model. For this goal, the extracted MAF of original vibration signals was reduced to three features using the GDA method mentioned in Section 2. To demonstrate the superiority of the introduced GDA dimensionality reduction method, when GDA is used in the process of the training pattern labelled into C classes, S_c number of samples in each class, C is set to 4, S_c is set to 25.

An experiment was conducted on the feature set, the authors explored this to evaluate the GDA method’s dimensionality reduction performance on the sample feature set. We compared GDA with PCA and LDA as representative methods. The experimental results of the PCA, LDA and GDA methods are shown in Fig. 9 to Fig. 11, respectively, which show that PCA and LDA have dim pattern classification performance, with three classes of overlap. Compared with these, GDA can obtain a clearer separation on the mapping. Therefore, GDA can accurately separate the bearing fault status for the extracted MAF set. In fact, this is because the GDA has greater ability to discover the maximal ratio of centre-class partner to within-class partner in the multi-aspect data by employing class label information. Overall, the best GDA method is used for the high-dimensional MAF dataset to obtain the low-dimensional feature dataset with prominent features as a dimensionality reduction task.

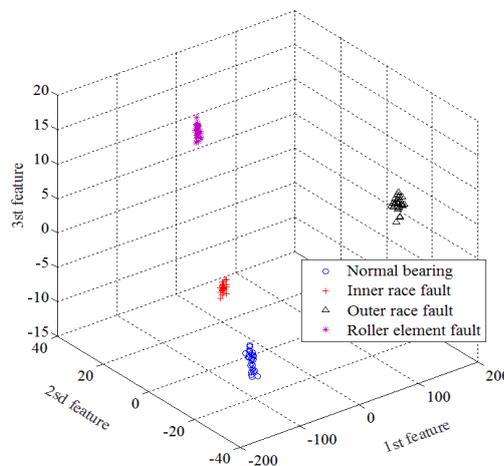


Fig. 11 Scatter plot for the reduced feature set by GDA

CRSVM training

In this study, CRSVM classifier models were designed to identify various bearing conditions. In fact, CRSVM1 was designed to identify normal bearing conditions, with normal bearing condition data assigned to $y = +1$, other data assigned to $y = -1$. CRSVM2 was designed to identify inner race faults of bearings, with inner race fault data assigned to $y = +1$, other data assigned to $y = -1$. CRSVM3 was designed to identify outer race faults, with outer race fault data assigned to $y = +1$, other data assigned to $y = -1$. In the same work, CRSVM4 was designed to identify roller element faults of bearings. To evaluate the performance of the diagnosis technique based on the proposed GDA-CRSVM method, the SVM was adopted to perform bearing condition diagnosis. This is a traditional model, with the parameter set selected to follow experience at $C = 200, d = 1$. These classifiers were trained on both the reduced feature set and the original MAF set to evaluate the recognition accuracy results and time of diagnosis.

Table 3 shows the MAF-GDA-CRSVM diagnosis techniques. They were designed to identify the different bearing conditions of the first dataset based on MAF and the proposed GDA-CRSVM method. In the experimentation in this dataset, these classifiers were trained with the reduced feature set, with 18 samples per class selected randomly as the training set, meaning 72 samples were collected as the training set. They were used to calculate the deterioration function Eq. (13) and construct the optimized classifiers. The seven rest samples per class were used to test the obtained classifier with the best parameters. The archived diagnosis results for bearing conditions are shown in Tables 4 and 5.

To demonstrate the effectiveness of the MAF-GDA-CRSVM diagnosis technique, we designed diagnosis models based on GDA-CRSVM to identify bearing conditions of the second dataset in Table 6. Similar to the process for the first dataset, the bearing fault diagnosis results are listed in Tables 7 and 8.

Table 3 The MAF-GDA-CRSVM diagnosis technique of the first dataset.

Bearing condition	A sample feature vector			Diagnosis technique				
				MAF-GDA-CRSVM1	MAF-GDA-CRSVM2	MAF-GDA-CRSVM3	MAF-GDA-CRSVM4	GDA-CRSVM4
HB	-53.2173	0.2071	0.1407	(+1)	(-1)	(-1)	(-1)	(-1)
IR fault	49.6764	8.0919	-0.2715	(-1)	(+1)	(-1)	(-1)	(-1)
OR fault	35.4594	-8.7378	0.8780	(-1)	(-1)	(+1)	(-1)	(-1)
RE fault	-31.0524	2.6843	-0.4135	(-1)	(-1)	(-1)	(-1)	(+1)

Table 4 The diagnosis accuracy result (%) of first dataset with various feature sets

Bearing condition	Samples		Original feature set		Reduction feature set	
	Training	Test	SVM	CRSVM	GDA-SVM	GDA-CRSVM
HB	72	28	75	99.35	98.70	100
IR fault	72	28	75	97.08	92.85	99.03
OR fault	72	28	75	96.42	85.71	99.67
RE fault	72	28	75	88.63	96.42	98.70

Table 5 The time cost (s) of first dataset with various feature sets

Bearing condition	Original feature set		Reduction feature set	
	SVM	CRSVM	GDA-SVM	GDA-CRSVM
HB	1.3249	1.4095	1.2473	1.3035
IR fault	1.1137	1.4454	1.0277	1.0883
OR fault	0.9838	1.4490	0.9833	1.2997
RE fault	0.8671	1.5266	0.8869	1.5039

According to these results, the MAF set extracted from original vibration signals can cope well with both the SVM and CRSVM diagnosis models for all four bearing conditions. The use of all input features does not ensure an improvement in the classification accuracy results for the various classifiers. In fact, the diagnosis accuracy results in Tables 4, 7 shown that evaluation on the conventional SVM model with the high-dimensional feature set obtained the very poor accuracy, the accuracy is only 75 % in every bearing condition. Thus, this diagnosis technique usually tends to produce a rejection and not reuse the model. It should be emphasized that the low-

dimensional feature set gained from GDA generated the better results in comparison with high-dimensional feature set by the same SVM models, the accuracy got maximum of 98.70 % for the first dataset and getting maximum of 90 % for the second dataset.

Additionally, in this work, we explored the CRSVM model to exploit accuracy results for classification purpose. The CRSVM model has achieved good diagnosis accuracy result in the healthy bearing (HB) and inner race (IR) fault conditions even with using the high-dimensional feature set. Thus, this CRSVM model is more appropriate with diagnosis technique for individual condition of bearing. In particular, Table 4 and 7 also showed that the proposed GDA-CRSVM-based diagnosis technique provides the best results for most bearing conditions. Its meaning that GDA method generated the compact feature set that inputted the optimized CRSVM classification model in integrating to produce effectiveness. Consequently, this diagnosis technique can be more used in the mechanical engineering environment to satisfy with the expected results. Fig. 12, 13 presented the results in comparison the proposed method with the other methods.

Moreover, the execution time of the CRSVM classifier for bearing fault diagnosis is faster than other classifiers for both the original feature set and the reduced feature set, as the results show in Tables 5 and 8. The considerable usefulness of reducing the original input feature space defined by the GDA was combined with the optimal classifier CRSVM to build an expert bearing fault diagnosis technique.

Table 6 The MAF-GDA-CRSVM diagnosis technique of the second dataset

Bearing condition	A sample feature vector			Diagnosis technique		
				MAF-GDA-CRSVM1	MAF-GDA-CRSVM2	MAF-GDA-CRSVM4
HB	-52.6982	0.2977	-5.8266	(+1)	(-1)	(-1)
IR fault	15.0308	-2.0533	-1.4911	(-1)	(+1)	(-1)
RE fault	-25.3478	0.9476	9.9225	(-1)	(-1)	(+1)

Table 7 The diagnosis accuracy result (%) of second dataset with various feature sets

Bearing condition	Samples		Original feature set		Reduction feature set	
	Training	Test	SVM	CRSVM	GDA-SVM	GDA-CRSVM
HB	30	15	75	99.09	90	100
IR fault	30	15	75	95.45	86.81	96.82
RE fault	30	15	75	93.18	87.27	95

Table 8 The time cost (s) of second dataset with various feature sets

Bearing condition	Original feature set		Reduction feature set	
	SVM	CRSVM	GDA-SVM	GDA-CRSVM
HB	1.0190	1.0456	0.8793	1.0240
IR fault	1.0800	1.1376	0.9420	1.1522
RE fault	1.0495	1.1607	0.8901	0.9278

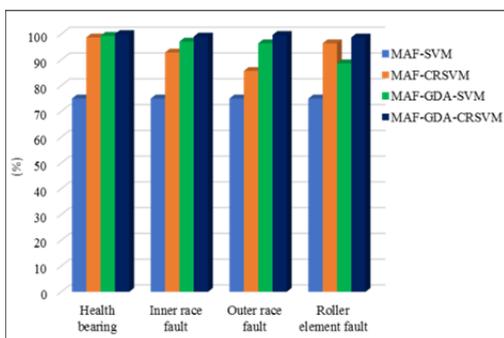


Fig.12 The classification accuracy of different diagnosis techniques for the first dataset

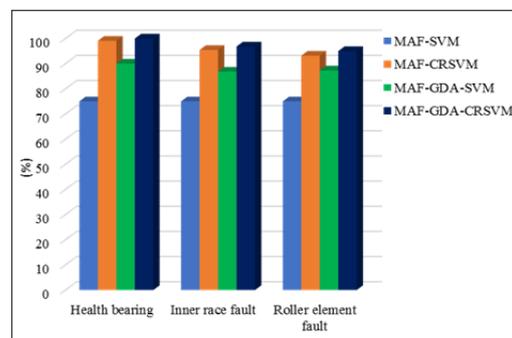


Fig.13 The classification accuracy of different diagnosis techniques for the second dataset

5. Conclusion

In this paper, a GDA-CRSVM-based expert fault diagnosis technique is proposed. The GDA-CRSVM method is a two-stage hybrid method that integrates GDA with CRSVM for an expert diagnosis technique. The original vibration dataset is firstly extracted the high-dimensional feature set by the MAF extraction. This feature set then provides GDA-CRSVM, in which the GDA method exploits to produce a reduced feature set which serves as input to the CRSVM classification model. The most of reduced feature set is used for training the optimized CRSVM classifier and the rest use for evaluation. The experimental results demonstrate the high efficiency of the proposed method and its expertness in bearing fault diagnosis.

In fact, the MAF extraction produces features in the different domain to represent the bearing status which can restrain the effect of proposed method. Furthermore, the proposed method capacity can be restricted due to the operating of GDA method depends on the classes label which reveals the non-objective condition in supervised feature learning. Thus, a feature reduction method is very useful and necessary for un-supervised feature learning in the future. Nevertheless, we have unshaken confidence that the GDA-CRSVM method can help to improve the fault classification performance of any diagnosis technique with different subjects. The practical applications of GDA-CRSVM is enquired and attached the most of features corresponding to the real subject status.

Acknowledgements

This research is supported by the National Key Research and Development Program of China (2016YFF0203400), and the National Natural Science Foundation of China (51575168 and 51375152). The authors would also like to thank the Collaborative Innovation Center of Intelligent New Energy Vehicles and the Hunan Collaborative Innovation Center for Green Car for support.

References

- [1] Zheng, J., Cheng, J., Yang, Y (2013). A rolling bearing fault diagnosis approach based on LCD and fuzzy entropy, *Mechanism and Machine Theory*, Vol. 70, 441-453, doi: [10.1016/j.mechmachtheory.2013.08.014](https://doi.org/10.1016/j.mechmachtheory.2013.08.014).
- [2] Liu, H., Wang, X., Lu, C. (2015). Rolling bearing fault diagnosis based on LCD-TEO and multifractal detrended fluctuation analysis, *Mechanical Systems and Signal Processing*, Vol. 60-61, 273-288, doi: [10.1016/j.ymsp.2015.02.002](https://doi.org/10.1016/j.ymsp.2015.02.002).
- [3] Chen, J., Liao, C.-M. (2002). Dynamic process fault monitoring based on neural network and PCA, *Journal of Process Control*, Vol. 12, No. 2, 277-289, doi: [10.1016/S0959-1524\(01\)00027-0](https://doi.org/10.1016/S0959-1524(01)00027-0).
- [4] Jolliffe, I.T. (2010). *Principal Component Analysis, Second Edition*, Springer, New York, USA.
- [5] Cox, T.F., Cox, M.A.A. (1994). *Multidimensional Scaling, Second Edition*, Chapman & Hall, London, UK.
- [6] Martinez, A.M., Kak, A.C. (2001). PCA versus LDA, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 2, 228-233, doi: [10.1109/34.908974](https://doi.org/10.1109/34.908974).
- [7] Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J. (1997). Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, 711-720, doi: [10.1109/34.598228](https://doi.org/10.1109/34.598228).
- [8] Baudat, G., Anouar, F. (2000). Generalized discriminant analysis using a kernel approach, *Neural Computation*, Vol. 12, No. 10, 2385-2404, doi: [10.1162/089976600300014980](https://doi.org/10.1162/089976600300014980).
- [9] Dogantekin, E., Dogantekin, A., Avci, D. (2011). An expert system based on generalized discriminant analysis and wavelet support vector machine for diagnosis of thyroid diseases, *Expert Systems with Applications*, Vol. 38, No. 1, 146-150, doi: [10.1016/j.eswa.2010.06.029](https://doi.org/10.1016/j.eswa.2010.06.029).
- [10] Li, C.-H., Kuo, B.-C., Lin, L.-H., Wu, W., Lan, D. (2013). Apply an automatic parameter selection method to generalized discriminant analysis with RBF kernel for hyperspectral image classification, In: *2013 International Conference on Machine Learning and Cybernetics*, Tianjin, China, 253-258, doi: [10.1109/ICMLC.2013.6890477](https://doi.org/10.1109/ICMLC.2013.6890477).
- [11] Abbasion, S., Rafsanjani, A., Farshidianfar, A., Irani, N. (2007). Rolling element bearings multi-fault classification based on the wavelet denoising and support vector machine, *Mechanical Systems and Signal Processing*, Vol. 21, No. 7, 2933-2945, doi: [10.1016/j.ymsp.2007.02.003](https://doi.org/10.1016/j.ymsp.2007.02.003).
- [12] Lau, K.W., Wu, Q.H. (2008). Local prediction of non-linear time series using support vector regression, *Pattern Recognition*, Vol. 41, No. 5, 1539-1547, doi: [10.1016/j.patcog.2007.08.013](https://doi.org/10.1016/j.patcog.2007.08.013).

- [13] Pelossof, R., Miller, A., Allen, P., Jebara, T. (2004). An SVM learning approach to robotic grasping, In: *2004 IEEE International Conference on Robotics and Automation, 2004, Proceedings. ICRA '04.*, Vol. 4, 3512-3518, doi: [10.1109/ROBOT.2004.1308797](https://doi.org/10.1109/ROBOT.2004.1308797).
- [14] Gryllias, K.C., Antoniadis, I.A. (2012). A support vector machine approach based on physical model training for rolling element bearing fault detection in industrial environments, *Engineering Applications of Artificial Intelligence*, Vol. 25, No. 2, 326-344, doi: [10.1016/j.engappai.2011.09.010](https://doi.org/10.1016/j.engappai.2011.09.010).
- [15] Samanta, B. (2004). Gear fault detection using artificial neural networks and support vector machines with genetic algorithms, *Mechanical Systems and Signal Processing*, Vol. 18, No. 3, 625-644, doi: [10.1016/S0888-3270\(03\)00020-7](https://doi.org/10.1016/S0888-3270(03)00020-7).
- [16] Dou, D., Zhou, S. (2016). Comparison of four direct classification methods for intelligent fault diagnosis of rotating machinery, *Applied Soft Computing*, Vol. 46, 459-468, doi: [10.1016/j.asoc.2016.05.015](https://doi.org/10.1016/j.asoc.2016.05.015).
- [17] Zhang, X., Chen, W., Wang, B., Chen, X. (2015). Intelligent fault diagnosis of rotating machinery using support vector machine with ant colony algorithm for synchronous feature selection and parameter optimization, *Neurocomputing*, Vol. 167, 260-279, doi: [10.1016/j.neucom.2015.04.069](https://doi.org/10.1016/j.neucom.2015.04.069).
- [18] Gomes, T.A.F., Prudêncio, R.B.C., Soares, C., Rossi, A.L.D., Carvalho, A. (2012). Combining meta-learning and search techniques to select parameters for support vector machines, *Neurocomputing*, Vol. 75, No. 1, 3-13, doi: [10.1016/j.neucom.2011.07.005](https://doi.org/10.1016/j.neucom.2011.07.005).
- [19] Yang, D., Liu, Y., Li, S., Li, X., Ma, L. (2015). Gear fault diagnosis based on support vector machine optimized by artificial bee colony algorithm, *Mechanism and Machine Theory*, Vol. 90, 219-229, doi: [10.1016/j.mechmachtheory.2015.03.013](https://doi.org/10.1016/j.mechmachtheory.2015.03.013).
- [20] Lam, A.Y.S., Li, V.O.K. (2010). Chemical-Reaction-Inspired Metaheuristic for Optimization, *IEEE Transactions on Evolutionary Computation*, Vol. 14, No. 3, 381-399, doi: [10.1109/TEVC.2009.2033580](https://doi.org/10.1109/TEVC.2009.2033580).
- [21] Alatas, B. (2012). A novel chemistry based metaheuristic optimization method for mining of classification rules, *Expert Systems with Applications*, Vol. 39, No. 2, 11080-11088, doi: [10.1016/j.eswa.2012.03.066](https://doi.org/10.1016/j.eswa.2012.03.066).
- [22] Li, J.-Q., Pan, Q.-K. (2013). Chemical-reaction optimization for solving fuzzy job-shop scheduling problem with flexible maintenance activities, *International Journal of Production Economics*, Vol. 145, No. 1, 4-17, doi: [10.1016/j.ijpe.2012.11.005](https://doi.org/10.1016/j.ijpe.2012.11.005).
- [23] Lam, A.Y.S., Li, V.O.K. (2012). Chemical reaction optimization: A tutorial, *Memetic Computing*, Vol. 4, No. 1, 3-17, doi: [10.1007/s12293-012-0075-1](https://doi.org/10.1007/s12293-012-0075-1).
- [24] Vapnik, V.N. (1995). *The Nature of Statistical Learning Theory*, Springer, New York, USA, doi: [10.1007/978-1-4757-2440-0](https://doi.org/10.1007/978-1-4757-2440-0).
- [25] Yu, Y., YuDejie, Junsheng, C. (2006). A roller bearing fault diagnosis method based on EMD energy entropy and ANN, *Journal of Sound and Vibration*, Vol. 294, No. 1-2, 269-277, doi: [10.1016/j.jsv.2005.11.002](https://doi.org/10.1016/j.jsv.2005.11.002).