



Feature Selection Scheme Based on Pareto Method for Gearbox Fault Diagnosis

Ridha Ziani^(✉), Hafida Mahgoun,
Semcheddine Fedala, and Ahmed Felkaoui

Laboratory of Applied Precision Mechanics,
Institute of Optics and Precision Mechanics, Ferhat Abbes University,
Setif 1, 19000 Setif, Algeria
ziani_lmpa@yahoo.fr

Abstract. Fault diagnosis based on pattern recognition approach has three main steps viz. feature extraction, sensitive features selection, and classification. The vibration signals acquired from the system under study are processed for feature extraction using different signal processing methods. Followed by feature selection process, classification is performed. The challenge is to find good features that discriminate the different fault conditions of the system, and increase the classification accuracy. This paper proposes the use of Pareto method for optimal feature subset selection from the pool of features. To demonstrate the efficiency and effectiveness of the proposed fault diagnosis scheme, numerical analyses have been performed using the Westland data set. The Westland data set consists of vibration data collected from a US Navy CH-46E helicopter gearbox in healthy and faulty conditions. First, features are extracted from vibration signals in time, spectral, and time-scale domain, then ranked according to three different criterions namely: Fisher score, correlation, and Signal to Noise Ratio (SNR). Afterward, data formed by only the selected features is used as input for the classification problem. The classification task is achieved using Support Vector Machines (SVM) method. The proposed fault diagnosis scheme has shown promising results. Using only the feature subset selected by Pareto method with Fisher criterion, SVMs achieved 100% correct classification.

Keywords: Signal processing · Support Vector Machine · Feature selection
Vibration · Fault diagnosis

1 Introduction

The gears are one of the major components of rotating machines, and proper maintenance of gear system is very essential to ensure reliability, safety, and performance of machines. The most of the developed methods for fault diagnosis of these systems are based on pattern recognition approach (Rafiee et al. 2007, 2010; Gryllias and Antoniadis 2012; Zhang et al. 2013; Ziani et al. 2017). The advantage of this approach is that it doesn't require large priori knowledge of the process under study. In this case, the diagnosis is assimilated to a classification problem (healthy or faulty condition).

The specialty of condition features is to provide accurate information regarding the condition of various components at different levels of damage (initial, heavy, or growing).

Vibration analysis is considered as a the most suitable tool for rotating machines faults diagnosis, thus it has attracted greater attention towards the researchers to acquire, analyze and quantify this parameter for improving the diagnosis precision. A multitude of methods have been developed. The yield of these techniques is, to distinguish changes in the signal brought on because of damaged or faulty components. These techniques are generally based on signal processing in different domains: time, spectral, time-frequency, and time-scale.

In time domain, the analysis is generally based on statistical features which provide an overall picture of some aspect of the time-series under investigation. Examples of these features include arithmetic mean, root mean square (RMS), variance (or standard deviation), skewness, kurtosis, peak-to-peak, crest factor (Ziani et al. 2017). Time Synchronous Averaging (TSA) (Abdul Rahman et al. 2011) is a pre-processing technique which was widely used for signal denoising before performing the feature extraction procedure.

In frequency domain, the most popular technique is Fast Fourier Transform (FFT) which provides a representation of the frequency content of a given signal. Various techniques resulted from FFT such as Power Spectral Density (PSD), cepstrum analysis, and envelope analysis. Many authors used amplitudes, entropy, and significant energy, calculated around fault characteristic frequency, to form the feature vector.

Time frequency distributions represent a good way to analyze the non stationary mechanical signals in which the spectral content changes with time. Short Time Fourier Transform (STFT), and Wigner–Ville distribution (Baydar and Ball 2001) are the well known time frequency distributions employed to overcome this problem, and widely used to processing the vibration signals of systems operating in non stationary modes.

The non-stationary signals can be considered as a superposition of components with respect to a set of basis functions which are each more or less localized in time. These basis functions can then be used to represent different frequency content simply by scaling them with respect to time. Signal decomposition using such called functions results in the so-called time-scale representations—and this leads directly to the wavelet transform (Worden et al. 2011).

Another group of features which have grown in popularity in recent years are those based on the Empirical Mode Decomposition (EMD) and Hilbert–Huang transform (HHT). These nonlinear analysis methods were employed to deal with the non-stationary vibrations to extract the original fault feature vector (Mahgoun et al. 2016).

A review on the application of the above signal processing methods and others for gear fault diagnosis can be found in Goyal et al. (2016).

In fault diagnosis methods based on pattern recognition approach, irrelevant features spoil the performance of the classifier and reduce the recognition accuracy (Kudo and Sklansky 2000). Hence it is necessary to reduce the dimension of the data by finding a small set of important features which can give good classification performance.

Dimensionality reduction is one of the most popular techniques to remove irrelevant and redundant features. Dimensionality reduction techniques may be divided in two main categories, called feature extraction (FE) and feature selection (FS) (Kotsiantis 2011). Feature extraction approaches map the original feature space to a new feature space with lower dimensions by combining the original feature space. This transformation may be a linear or nonlinear combination of the original features. These methods include Principle Component Analysis (PCA), Linear Discriminant Analysis (LDA) and Canonical Correlation Analysis (CCA). Bartkowiak and Zimroz (2014) cited other transformation methods used for reducing the dimensionality of the data, such as: Independent Component Analysis (ICA), Isomap, local linear embedding, kernel PCA, and curvilinear component analysis. On the other hand the term feature selection refers to algorithms that output a subset of the input feature set.

Both Feature extraction and feature selection are capable of improving learning performance, lowering computational complexity, building better generalizable models, and decreasing required storage (Tang et al. 2014). While feature selection selects a subset of features from the original feature set without any transformation, and maintains the physical meanings of the original features, it is better to select and process original data than create new features because by projections the physical meaning of the original variables may be lost (Bartkowiak and Zimroz 2014).

For the classification problem, algorithms used to select features are divided into three categories: filter, wrapper, and embedded methods (Tang et al. 2014). Filter methods rank features or feature subsets independently of the classifier, while wrapper methods use the predictive accuracy of a classifier to assess feature subsets, thus, these methods are usually computationally heavy and they are conditioned to the type of classifier used. Another type of feature subset selection is identified as embedded methods. In this case, the feature selection process is done inside the induction algorithm itself, i.e. attempting to jointly or simultaneously train both a classifier and a feature subset. They often optimize an objective function that jointly rewards the accuracy of classification and penalizes the use of more features (Kotsiantis 2011).

The goal of this study is to present a feature selection scheme based on Pareto method combined with three different criterions namely: Fisher score, Correlation criterion, and Signal to Noise Ratio (SNR). This approach was tested using vibration data acquired from a helicopter gearbox. In this study, Support Vector Machines (SVM) was used to achieve the classification task. This method has a good generalization capability even in the small-sample cases of classification and has been successfully applied in fault detection and diagnosis in Gryllias and Antoniadis (2012), Ziani et al. (2017), Konar and Chattopadhyay (2011).

The rest of this paper is organized as follow: In the second section we present the basic principle of SVM. Vibration data and feature extraction procedure are given the third section. In the fourth section we present the proposed feature selection method. Results are presented and discussed in the fifth section. Finally, the sixth section is dedicated to the conclusion.

2 Support Vector Machines (SVMs)

SVMs is a relatively a new computational learning method proposed by Vapnik (1998). The essential idea of SVMs is to place a linear boundary between two classes of data, and adjust it in such a way that the margin is maximized, namely, the distance between the boundary and the nearest data point in each class is maximal. The nearest data points are known as Support Vectors (SVs) (Konar and Chattopadhyay 2011). Once the support vectors are selected, all the necessary information to define the classifier is provided.

If the training data are not linearly separable in the input space, it is possible to create a hyper plane that allows linear separation in the higher dimension. This is achieved through the use of a transformation that converts the data from an N -dimensional input space to Q -dimensional feature space. A kernel can be used to perform this transformation. Among the kernel functions in common use are linear functions, polynomials functions, Radial Basis Functions (RBF), and sigmoid functions. A deeper mathematical treatise of SVMs can be found in the book of Vapnik (1998) and the tutorials on SVMs (Burges 1998; Scholkopf 1998).

SVMs is essentially a two-class classification technique, which has to be modified to handle the multiclass tasks in real applications e.g. rotating machinery which usually suffer from more than two faults. Two of the common methods to enable this adaptation include the One-against-all (OAA) and One-against-one (OAO) strategies (Yang et al. 2005).

In the One-against-all strategy, each class is trained against the remaining $N - 1$ classes that have been collected together. The “winner-takes-all” rule is used for the final decision, where the winning class is the one corresponding to the SVM with the highest output (discriminant function value). For one classification, N two-class SVMs are needed.

The One-against-one strategy needs to train $N(N - 1)/2$ two-class SVMs, where each one is trained using the data collected from two classes. When testing, for each class, score will be computed by a score function. Then, the unlabeled sample x will be associated with the class with the largest score.

3 Vibration Data and Feature Extraction

3.1 The CH46 Gearbox

Vibration data used in this paper is acquired from the Westland CH46 Helicopter gearbox (Cameron 1993). The gearbox is relatively complex, driving both the main shaft and many auxiliary devices. This vibration data have been widely used to validate the effectiveness of new algorithms for gear fault diagnosis (Williams and Zalubas 2000; Loughlin and Cakrak 2000; Chang et al. 2009; Nandi et al. 2013).

Figure 1 shows the simplified main section of the CH46 helicopter gearbox including the input, quill and output shafts, the spur pinion/collector gear pair and the spiral bevel pinion/gear pair. In this study we interest only to fault of gear 5, (spiral bevel pinion tooth spalling). For this element, the vibration data is composed of twenty

four (24) signals: nine (9) signals acquired in normal condition (Fig. 2a), six (6) with defect Level 1 condition (Fig. 2b), and nine (9) with defect level 2 (Fig. 2c).

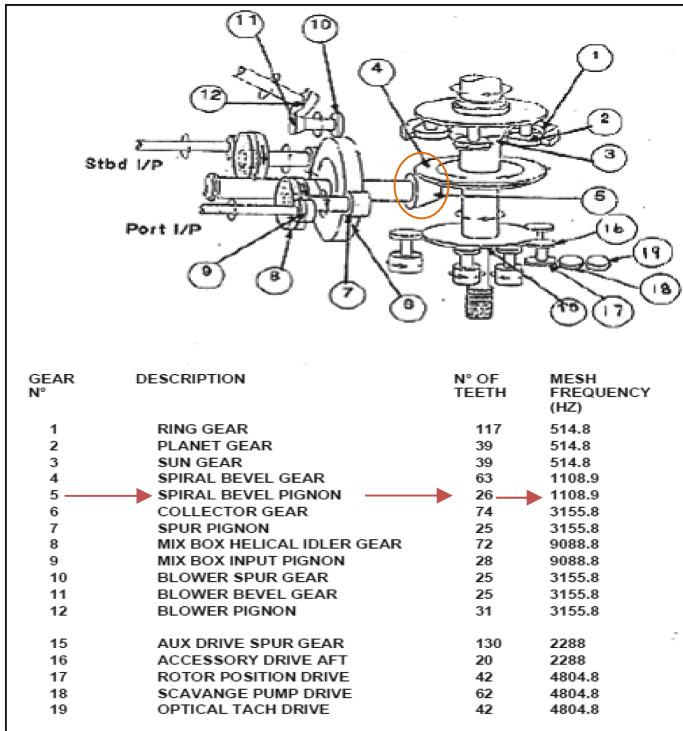


Fig. 1. Simplified main section of the CH46 helicopter gearbox

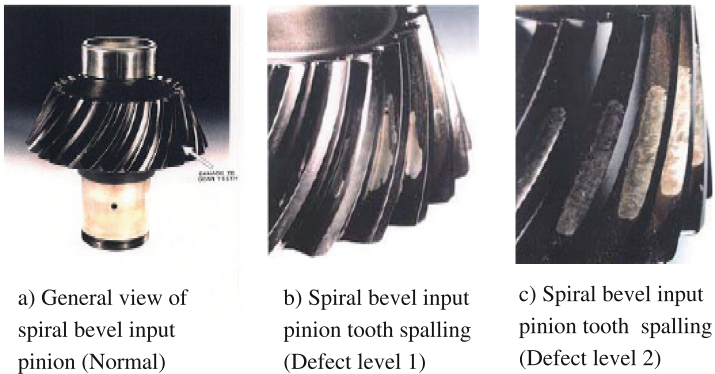


Fig. 2. The Bevel input pinion used in the test

Signals are composed of 412464 samples acquired with a sampling frequency of 103116 Hz. The following parameters are given:

- Number of teeth of spiral bevel pinion/gear: $n1 = 26$; $n2 = 63$;
- Rotating frequency $fr1 = 42.65$ Hz; $fr2 = 17.60$ Hz;
- Meshing frequency: $fm1 = 1108.9$ Hz, $fm2 = 3155$ Hz.

3.2 Features Extraction

In order to obtain sufficient samples for training and testing SVMs, each signal was divided into ten (10) samples of 41246 points. Afterwards, different feature subsets were extracted from each sample using different signal processing methods. These features were extracted in time domain, frequency domain, and time frequency domain.

Statistical Features. In time domain (Fig. 3), signals are processed to extract the nine following statistical features: mean, Root Mean Square (RMS), skewness, kurtosis, Peak factor, Peak to Peak value, Clearance factor, Shape factor, and Impulse factor. The mathematical formula of these features can be found in Goyal et al. (2016).

Spectral Features. In spectral domain, another feature subset is formed by calculating the Power Spectral Density (PSD) in different bands around the meshing frequency and its four harmonics. The width of each band is chosen equal to ten rotating frequency (426 Hz). Consequently the frequency bands are: [895–1321 Hz], [2004–2430 Hz], [3113–3539 Hz] [4222–4648 Hz], and [5331–5757 Hz].

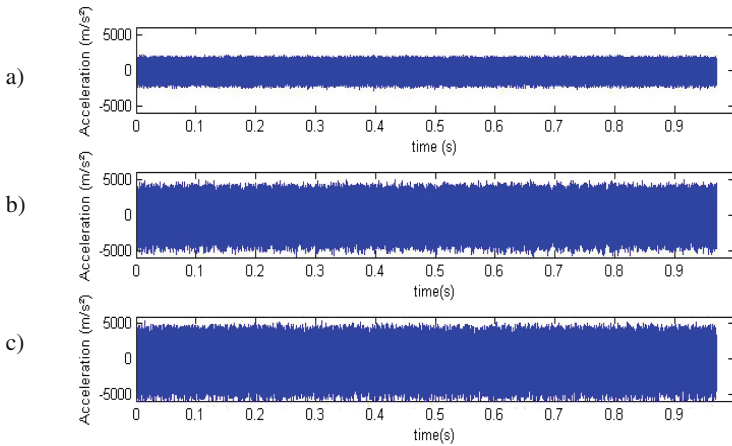


Fig. 3. Times domain signals acquired under torque of 45%. (a) Normal condition (b) defect level 1, (c) defect level 2

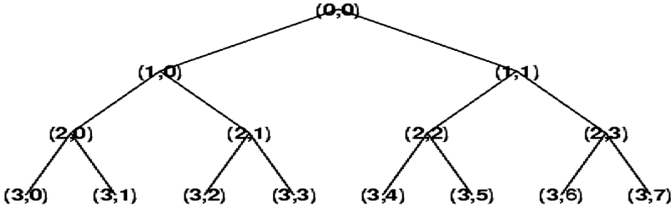


Fig. 4. Wavelet packet decomposition tree

Wavelet Packet Decomposition. In the last decade, Wavelet Packet Decomposition (WPD) has been proved to be a suitable tool for gear fault diagnosis, especially on vibration signal features extraction (Zhang et al. 2013). WPD shows good performance on both high and low frequency analysis. The selection of the mother wavelet is a crucial step in wavelet analysis. In (Rafiee et al. 2010) it has been shown that the Daubechies 44 wavelet is the most effective for both faulty gears and bearings. Hence, db44 is adopted in this paper. As shown in Fig. 4, Samples are firstly decomposed into forty coefficients at three depths, and then the kurtosis and energy of the 8 last coefficients (third depth of decomposition) are calculated. As result another feature set containing 16 features is obtained.

Empirical Mode Decomposition EMD. Empirical mode decomposition (EMD) is relatively new method of signal processing which was applied in bearings and gears fault diagnosis of rotating machinery (Liu et al. 2005; Mahgoun et al. 2016). It does not use a priori determined basis functions and can iteratively decompose a complex signal

Table 1. List of the extracted features

No	Feature	No	Feature
1	Mean	18	Kurtosis of coefficient 3.3
2	RMS	19	Kurtosis of coefficient 3.4
3	Skewness	20	Kurtosis of coefficient 3.5
4	Kurtosis	21	Kurtosis of coefficient 3.6
5	Peak factor	22	Kurtosis of coefficient 3.7
6	Peak to peak value	23	Energy of coefficient 3.0
7	Clearance factor	24	Energy of coefficient 3.1
8	Shape factor	25	Energy of coefficient 3.2
9	Impulse factor	26	Energy of coefficient 3.3
10	PSD in [895–1321 Hz]	27	Energy of coefficient 3.4
11	PSD in [2004–2430 Hz]	28	Energy of coefficient 3.5
12	PSD in [3113–3539 Hz]	29	Energy of coefficient 3.6
13	PSD in [4222–4648 Hz]	30	Energy of coefficient 3.7
14	PSD in [5331–5757 Hz]	31	Kurtosis of the 1st IMF
15	Kurtosis of coefficient 3.0	32	Kurtosis of the 2nd IMF
16	Kurtosis du coefficient 3.1	33	Kurtosis of the 3rd IMF
17	Kurtosis of coefficient 3.2		

into a finite number of zero mean oscillations named intrinsic mode functions (IMFs). Each resulting elementary component (IMF) can represent the local characteristic of the signal (Mahgoun et al. 2016).

Samples were decomposed into a number of IMFs, and then the kurtosis of the three first IMFs is calculated.

The feature extraction operation was repeated with all samples of the three operating modes (normal, with defect level 1, and defect level 2). Table 1 summarize the list of the extracted features.

4 Feature Selection

From the above section, one can understand that there will be thirty three (33) features extracted for classification of samples belonging to three different classes. However, the entire feature set will not be used for the classification. Some of the features contain redundant information which may unnecessarily increase the complexity. This problem is frequently found in almost all pattern recognition problems. The challenge is to find out the most pertinent features and eliminate the redundant features to increase the classification accuracy.

In this study, we propose a filter based feature selection method. First, features are ranked in decreasing order based on their evaluation with a selection criterion. Afterword Pareto method is used to select the optimal feature subset according to features evaluations, then the corresponding classification accuracies using SVMs are tabulated. Three different criterions are compared: Fisher criterion, correlation criterion, and Signal to noise ratio (SNR).

4.1 Pareto Based Feature Selection Method

Pareto is a technique used for decision making based on the Pareto Principle, known as the 80/20 rule (Kramp et al. 2016). It is a decision-making technique that statistically separates a limited number of input factors as having the greatest impact on an outcome, either desirable or undesirable. Pareto analysis is based on the idea that 80% of a project's benefit can be achieved by doing 20% of the work. This ratio is used in this study to select the optimal feature subset from the initial set. The selected features are those cumulating 80% of the selection criterion score. This can be realised as follow:

1. The first step is to evaluate the score of each feature using a selection criterion,
2. The second step is to rank features in decreasing order according to their scores,
3. Compute the cumulative percentage of each feature,
4. Plot a curve with features on x- and cumulative percentage on y-axis,
5. Plot a bar graph with features on x- and percent frequency on y-axis,
6. Draw a horizontal dotted line at 80% from the y-axis to intersect the curve. Then draw a vertical dotted line from the point of intersection to the x-axis. The vertical dotted line separates the important features (on the left) and trivial features (on the right).

4.2 Selection Criteria

In the proposed method, features are selected according to their evaluation using three different criteria: fisher score, correlation criterion, and Signal to noise criterion. Also, the effect of these criteria on classification accuracy will be discussed in Sect. 5.

Fisher score. The idea is that features with high quality should assign similar values to instances in the same class and different values to instances from different classes. With this intuition, the score for the i -th feature $S(i)$ will be calculated by Fisher Score as (Duda et al. 2000):

$$s(i) = \frac{\sum_{j=1}^c n_j (\bar{\mu}_{ij} - \bar{\mu}_i)^2}{\sum_{j=1}^c n_j \rho_{ij}^2} \quad (1)$$

where $\bar{\mu}_{ij}$ and ρ_{ij} are the mean and the variance of the i -th feature in the j -th class respectively, n_j is the number of instances in the j -th class, and $\bar{\mu}_i$ is the mean of the i -th feature, c is the number of classes.

Correlation Criterion. The Correlation criterion evaluates features on the basis of the hypothesis that good feature is highly correlated with the classification. This correlation is measured using ‘‘Bravais-Pearson’’ criterion given by the following equation (Dash and Liu 2003):

$$C(i) = \frac{\sum_{k=1}^m (\mu_{ik} - \bar{\mu}_i)(y_k - \bar{y})}{\sqrt{\sum_{k=1}^m (\mu_{ik} - \bar{\mu}_i)^2 \sum_{k=1}^m (y_k - \bar{y})^2}} \quad (2)$$

where $\bar{\mu}_i$ et \bar{y} are the mean of the i -th feature and class labels of data respectively. m is the number of all instances.

Signal to Noise Ratio (SNR). The signal to noise ratio (SNR) identifies the expression patterns with a maximal difference in mean expression between two classes and minimal variation of expression within each class (Mishra and Sahu 2011).

$$SNR(i) = \frac{|\bar{\mu}_{i1} - \bar{\mu}_{i2}|}{(\sigma_{i1} - \sigma_{i2})} \quad (3)$$

Where $\bar{\mu}_{i1}$ and $\bar{\mu}_{i2}$ denote the mean of the i -th feature in class 1 and class 2 respectively. σ_1 and σ_2 are the standard deviations for the i -th feature in each class.

5 Results and Discussion

From the initial feature set, the best features have been selected using Pareto-Fisher based feature selection algorithm, Pareto-correlation based feature selection algorithm, and Pareto-SNR based feature selection algorithm given in the above section.

Data composed of 240 samples was divided into two equally subsets. The first one is used for training SVMs, while the second is used for the test. SVMs accuracy is evaluated by the number of misclassified samples in the test. Based on results of previous work (Ziani et al. 2017), SVMs is trained with an RBF kernel and OAO strategy for multiclass SVM is adopted.

In the first time the SVMs classifier has been trained with the initial feature set composed of 33 features, then it has been trained with the optimal feature subset selected with different algorithms and the results are tabulated as follows:

Figure 5 shows the Pareto curve in the case of features selection using Fisher score. Features are ranked according to their scores, and then the selected features are those cumulating 80% of Fisher criterion. In this case, the optimal feature subset is composed of the following features: 1, 10, 6, 30, 2, 8, and 27.

From Tables 2 and 3, one can understand that three algorithms have selected the pertinent feature subset in different manner. However, looking at a problem in classification accuracy view point, it is clear that the classification accuracy was improved with the selected features in all cases. The Pareto-fisher gives 100% with only seven features, Pareto-correlation gives also 100% but with ten features, and finally Pareto-SNR gives 97.5% with 13 features.

Figure 6 shows 3D scatter plot of data with the entire feature set. This plot is performed using Principal Components Analysis (PCA) where data is projected on three Principal Components: PC1, PC2, and PC3. It is important to note that PCA is used here for data visualization but not for selection purpose. Figures 7, 8 and 9 show plots of data with pertinent features selected using the three criterions. It is clear that data is well separated using the selected features which explain the improvement of classification accuracy. The best data separation is obtained using the selected features by Pareto-fisher algorithm.

From Table 2, one can understand that the selected features are not the same in the three cases. This is logical since different criterions were used. However some features are selected by the three algorithms which confirm their discriminant ability. These features are: the mean, peak to peak, PSD calculated in the band [895–1321 Hz], and Energy of coefficient 3.7. The mean and Peak to peak values quantify the level of vibration. When any fault occurs in a gear, the level of vibration increase and the values of these features increase consequently. This can be confirmed in Fig. 3 where the level of vibration increases with the level of defect. PSD is a measure of the power of signal in frequency domain. When fault appear, PSD calculated around the meshing frequency increase significantly. This can be explained by the modulation phenomena characterized by the production of sidebands around the meshing frequency.

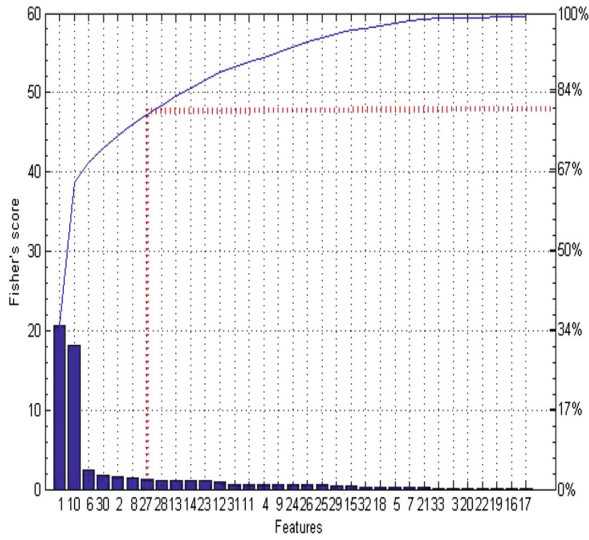


Fig. 5. Pareto curve with Fisher score

Table 2. Optimal features subsets

Selection method	Optimal feature subset
Pareto-Fisher	1, 10, 6, 30, 2, 8, 27
Pareto-correlation	10, 30, 4, 13, 15, 25, 27, 29, 18, 1
Pareto-SNR	1, 10, 11, 6, 2, 23, 28, 14, 12, 8, 24, 13, 26

Table 3. SVM classification accuracy

Inputs	Number of features	SVM accuracy (%)
The entire feature set	33	95.83
Features selected with Pareto-fisher	7	100
Features Selected with Pareto-correlation	10	100
Features Selected with Pareto-SNR	13	97.5

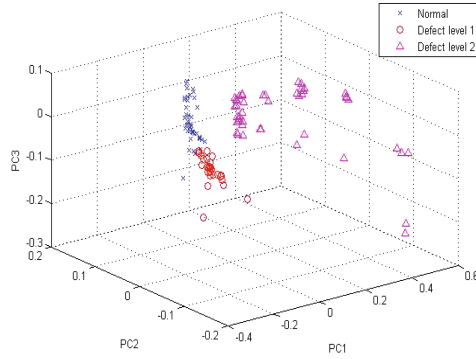


Fig. 6. 3D scatter plot of data with the entire feature set (33 features)

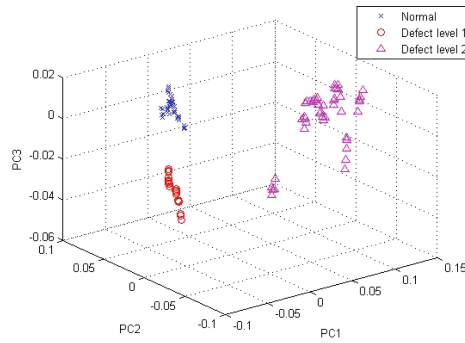


Fig. 7. 3D scatter plot of data with features selected by Pareto-Fisher algorithm (7 features)

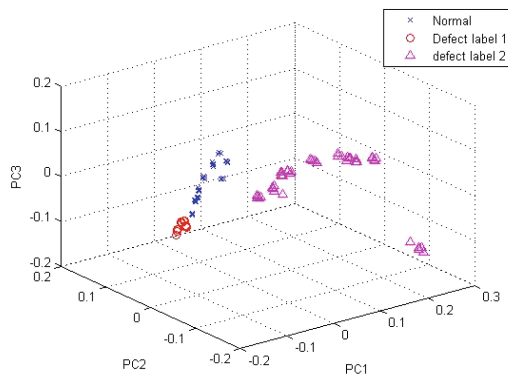


Fig. 8. 3D scatter plot of data with features selected by Pareto-correlation algorithm (10 features)

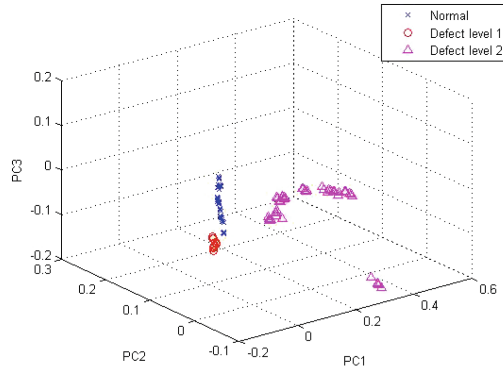


Fig. 9. 3D scatter plot of data with features selected by Pareto-SNR algorithm (13 features)

6 Conclusion

In this paper, an investigation has been made on different feature selection criterions and their effect on classification also studied. Different features were extracted from the vibration data using different signal processing methods. There were totally thirty three features out of which certain features may not be use for classification. The optimal feature subset was selected according three different criterions such as: Fisher score, correlation criterion, and Signal to Noise criterion. Their results and corresponding classification accuracies have been tabulated. Pareto method has been used to define the number of features to be selected. It can be concluded that Pareto-fisher based feature selection algorithm with SVMs classifier seem to perform better for this application. However, other algorithms also may suit for some other applications. Our future work will focus on a more comprehensive fault diagnosis of rotating machinery based on the unsupervised learning methods.

References

- Abdul Rahman, A.G., Chao, O.Z., Ismail, Z.: Effectiveness of impact-synchronous time averaging in determination of dynamic characteristics of a rotor dynamic system. *Measurement* **44**, 34–45 (2011)
- Bartkowiak, A., Zimroz, R.: Dimensionality reduction via variables selection - linear and nonlinear approaches with application to vibration-based condition monitoring of planetary gearbox. *Appl. Accoustics* **77**, 169–177 (2014)
- Baydar, N., Ball, A.: A comparative study of acoustic signals in detection of gear failures using Wigner-ville distribution. *Mech. Syst. Signal Process.* **15**, 1091–1107 (2001)
- Burges, C.A.: Tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Disc.* **2**, 955–974 (1998)
- Cameron, B.G.: Final report on CH-46 Aft transmission seeded fault testing, Research Paper RP907. Westland Helicopters Ltd, UK (1993)

- Chang, R.K.Y., Loo, C.K., Rao, M.V.C.: Enhanced probabilistic neural network with data imputation capabilities for machine-fault classification. *Neural Comput. Appl.* **18**, 791–800 (2009)
- Dash, M., Liu, H.: Consistency-based search in feature selection. *Artif. Intell.* **151**, 155–176 (2003)
- Duda, R., Hart, P., Stork, D.: *Pattern Classification*, 2nd edn. Wiley, Hoboken (2000)
- Goyal, D., Vanraj, B.S., Pabla, S., Dhama, S.: Condition monitoring parameters for fault diagnosis of fixed axis gearbox: a review. *Arch. Computat. Methods Eng.* **24**(3), 543–556 (2016). <https://doi.org/10.1007/s11831-016-9176-1>
- Gryllias, K.C., Antoniadis, I.A.: A support vector machine approach based on physical model training for rolling element bearing fault detection in industrial environments. *Eng. Appl. Artif. Intell.* **25**, 326–344 (2012)
- Konar, P., Chattopadhyay, P.: Bearing fault detection of induction motor using wavelet and support vector machines (SVMs). *Appl. Soft Comput.* **11**, 4203–4211 (2011)
- Kotsiantis, S.B.: Feature selection for machine learning classification problems: a recent overview. *Artif. Intell. Rev.* **42**(1), 157 (2011). <https://doi.org/10.1007/s10462-011-9230-1>
- Kramp, K.H., Van Det, M.J., Veeger, N.J.G.M.: The Pareto analysis for establishing content criteria in surgical training. *J. Surg. Educ.* **73**, 892–901 (2016)
- Kudo, M., Sklansky, J.: Comparison of algorithms that select features for pattern classifiers. *Pattern Recogn.* **33**, 25–41 (2000)
- Liu, B., Riemenschneider, S., Xub, Y.: Gearbox fault diagnosis using empirical mode decomposition and Hilbert spectrum. *Mech. Syst. Signal Process.* **17**, 1–17 (2005)
- Loughlin, P., Cakrak, F.: Conditional moments analysis of transients with application to the helicopter fault data». *Mech. Syst. Signal Process.* **14**, 515–522 (2000)
- Mahgoun, H., Chaari, F., Felkaoui, A., Haddar, M.: Early detection of gear faults in variable load and local defect size using ensemble empirical mode decomposition (EEMD). In: *Advances in Acoustic and Vibration, Proceeding of the International Conference on Acoustic and Vibration (ICAV 2016)*, Hammamet, Tunisia (2016)
- Mishra, D., Sahu, B.: Feature selection for cancer classification: a signal-to-noise ratio approach. *Int. J. Sci. Eng. Res.* **2**, 1–7 (2011)
- Nandi, A.K., Liu, C., Wong, M.L.D.: Intelligent vibration signal processing for condition monitoring. In: *International Conference Surveillance 7*, Institute of Technology of Chartres, France, 29–30 October 2013
- Rafiee, J., Arvani, F., Harifi, A., Sadeghi, M.-H.: Intelligent condition monitoring of a gearbox using artificial neural network. *Mech. Syst. Signal Process.* **21**, 1746–1754 (2007)
- Rafiee, J., Rafiee, M.A., Tse, P.W.: Application of mother wavelet functions for automatic gear and bearing fault diagnosis. *Expert Syst. Appl.* **37**, 4568–4579 (2010)
- Scholkopf, B.: SVMs—a practical consequence of learning theory. *IEEE Intell. Syst.* **13**, 18–19 (1998)
- Tang, J., Alelyani, S., Liu, H.: Feature selection for classification: a review. In: *Data Classification: Algorithms and Applications*, p. 37. CRC Press (2014)
- Vapnik, V.N.: *Statistical Learning Theory*. Wiley, New York (1998)
- Williams, W.J., Zalubas, E.J.: Helicopter transmission fault detection via time-frequency, scale and spectral methods. *Mech. Syst. Sig. Process.* **14**, 545–559 (2000)
- Worden, K., Staszewski, W.J., Hensman, J.J.: Natural computing for mechanical systems research: a tutorial overview. *Mech. Syst. Sig. Process.* **25**, 4–111 (2011)

- Yang, B.S., Hwang, W.W., Han, T.: Fault diagnosis of rotating machinery based on multi-class support vector machines. *J. Mech. Sci. Technol.* **19**, 846–859 (2005)
- Zhang, Z., Wang, Y., Wang, K.: Fault diagnosis and prognosis using wavelet packet decomposition, Fourier transform and artificial neural network. *J. Intell. Manuf.* **24**, 1213–1227 (2013)
- Ziani, R., Felkaoui, A., Zegadi, R.: Bearing fault diagnosis using multiclass support vector machines with binary particle swarm optimization and regularized Fisher's criterion. *J. Intell. Manuf.* **28**(2), 405–417 (2017)