



Separability Index-Based Feature Selection and a Two-Tier Classifier for Improving Diagnostic Performance in Bearings

Viet Tra, Bach Phi Duong, and Jong-Myon Kim^(✉)

School of IT Convergence, University of Ulsan, Ulsan 680-749, South Korea
traviet.vt@gmail.com, duongbachphi@gmail.com,
jongmyon.kim@gmail.com

Abstract. Faults in bearings have traditionally been diagnosed by identifying the characteristic defect frequencies in the envelope power spectrum (EPS). Although EPS-based methods are effective for constant operating conditions, it makes difficult to detect a fault when the machine operates at variable shaft speeds. To address this issue, we propose a method that performs as follows: (a) a feature extraction technique is used to extract as many information as possible from different sub-bands using a non-stationary signal analysis method (*e.g.*, discrete wavelet pack transform (DWPT)), (b) a novel feature selection method is then applied to select the most informative features from the extracted high-dimensional feature pool, and (c) the selected features are further fed to a two tiers classifiers to identify the condition of a bearing. This two tiers classifier includes the combination of support vector machine groups in the first tier and the combination multilayer perceptron in the second tier. The proposed method is tested for a bearing fault diagnosis application using acoustic emission (AE) signal under various conditions.

Keywords: Bearing fault diagnosis · Support vector machine
Multilayer perceptron · Feature selection

1 Introduction

Bearing elements are crucial components and the leading cause of failures in induction motors. If timely detection of defects in a bearing is not performed, it could result in significant damage to equipment, interruption of production, and costly repairs. Therefore, the early detection of bearing faults is an integral part of condition-based maintenance that helps operators to provide a timely maintenance solution for the equipment.

There are many techniques that have been proposed for diagnosing bearing faults based on vibration signal. On the other hand, fault detection and diagnosis (FDD) of bearing defects has shown significant performance improvement using acoustic

emission (AE) for low-speed bearings since AE-based methods are highly sensitive to the low energy acoustic emissions released by developing cracks in a bearing [1, 2]. Therefore, this study employs the AE signal to diagnose defects in the bearing.

One of the most common methods for diagnosing localized defects in a bearing is an envelope power spectrum (EPS) analysis method. In the EPS-based method, detecting peaks are strongly correlated with characteristic frequencies in the power spectrum of the envelope signal. However, these frequencies depend on the rotational speed of the device as well as the bearing geometry. Therefore, this method may not be suitable for the use in cases when the rotational speed of an equipment changes during the operation process. Another widely used method is the data-driven method. In the data-driven approach, as many features of the acquired signals are extracted as possible to define each fault distinctly and then used to develop discriminatory models for classifying different bearing faults. As bearing fault signals (*e.g.*, AE in this study) are inherently non-stationary and non-linear, frequency domain analyses are often ineffective in helping to extract useful information to detect faults. To address this issue, we propose a method that divides a signal into four sub-bands by using discrete wavelet packet transform (DWPT) which is more efficient for analyzing non-stationary signals [3]. A set of four feature vectors are then extracted from four sub-bands instead of using one feature vector of other methods [4]. This study ensures a high-dimensional feature vector so that no information is missed, but this high-dimensional feature vector can be redundant and irrelevant. Therefore, a feature selection technique is applied based on the separability index (SI) of features, which is an evaluation metric that ensures the quality of the selected features. These selected features are only discriminative in a classifier point of view, but also provide a simpler model (*i.e.*, a low-dimensional feature vector) for the classifier. Finally, a two-tier classifier (TTC) is applied to classify each fault type, which consists of two tiers; support vector machine groups (SVM-Groups) and a multilayer perceptron (MLP).

2 The Proposed Methodology for Bearing Fault Diagnosis

A block diagram of the proposed method is shown in Fig. 1. In this method, we use the discrete wavelet packet transform (DWPT) with Daubechies 15 mother wavelet to decompose the original signal into four sub-bands. In DWPT, each sub-band represents a resolution level of the original signal, and the number of sub-bands is determined based on empirical tests by evaluating sufficient extraction of discriminating fault information. For each sub-band signal, a total of twenty-two features are extracted including sixteen time-domain features such as peak value (PV), root mean square (RMS), kurtosis value (KV), crest factor (CF), clearance factor (CRF), impulse factor (IF), shape factor (SF), entropy value (EV), skewness value (SV), square mean root

(SMR), 5th normalized moment, 6th normalized moment, mean value (MV), peak-to-peak (PPV), margin factor (MF), kurtosis factor (KF), and six frequency-domain features, such as frequency center (FC), RMS frequency (RMSF), root variance frequency (RVF), kurtosis frequency (KF), entropy frequency (EF), sum square frequency (SSF).

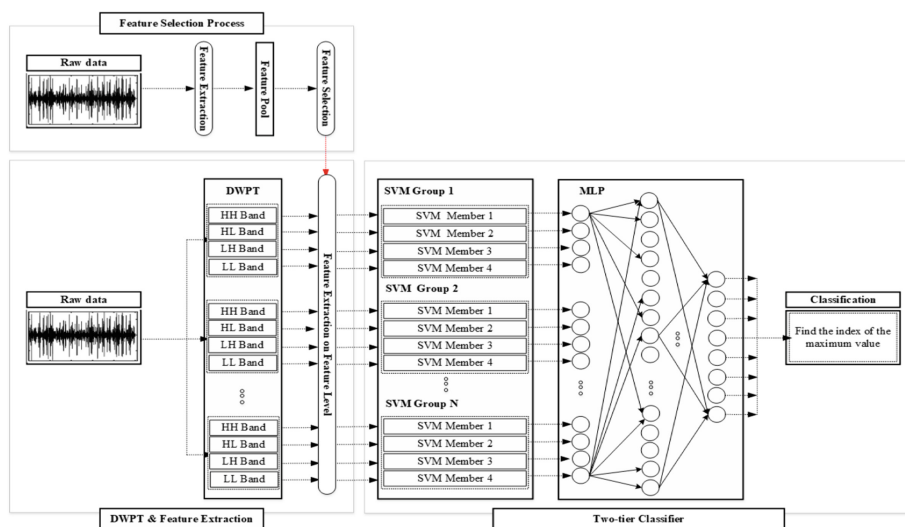


Fig. 1. The proposed method for bearing fault diagnosis using DWPT analysis and a two-tier classifier

Feature selection is an essential technique in such feature-based methods. It aims to remove non-useful features and select a subset of relevant features, thereby reducing the dimension of the feature space, the complexity of the system as well as improving system performance parameters such as accuracy [3]. In this study, we propose a new method for selecting informative features based on the separability index (SI) of features. These selected features then form a feature vector that used as input for a classifier. Finally, a two-tier classifier (TTC) is used, in which the first tier is a combination of support vector machine groups (SVM-Groups) as the primary classifier, and the second tier is a multilayer neural network (MLP) that uses output of the SVM-Groups in the first tier as input. Therefore, this MLP makes a final decision for classifying bearing faults. In the following sub-sections, details of feature selection and the two-tier classifier are given.

2.1 Feature Selection Method Base on a Separability Index

A feature is regarded to be useful, if it represents the samples of different fault classes as distinct clusters in its feature space. In Fig. 2(a), the shape factor (SF) and square mean root (SMR) features describe the pattern of each class as a dense cluster that is entirely separated from other clusters. This means that in this feature space, the classifier can easily classify groups of faulty signals. Therefore, these features are considered as useful features, while the peak and the impulse factor features are redundant ones, as shown in Fig. 2(b). In addition, a feature is determined to be good, if the variance between samples in the same class is small, and between samples that belong to different categories is large. Figure 2(c) and (d) are examples of two cases in which the features are considered to be good and redundant.

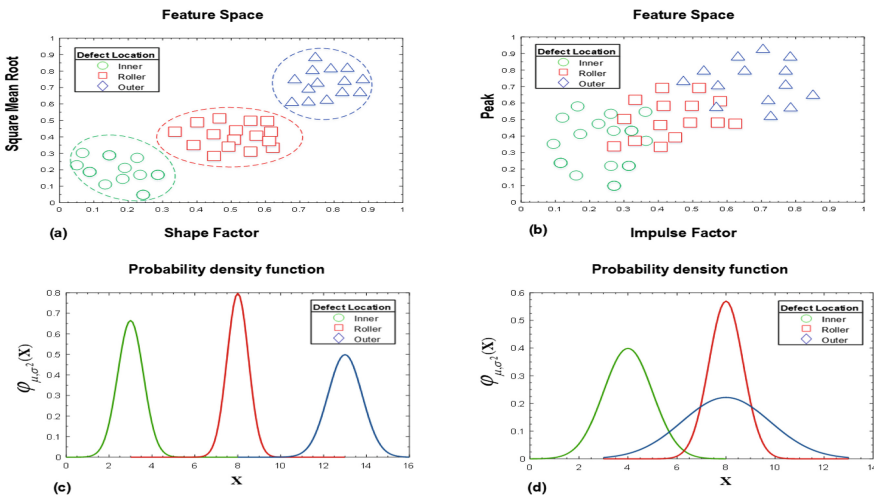


Fig. 2. Representation of fault classes in different feature spaces

In this study, we use a separability index (SI) as a tool for evaluating the discrimination of feature as well as comparing it with other features. This index is the ratio between the degree of separation (DoS) between the samples of different fault classes, and the average sparseness of the fault classes (SoC) in the same class feature space. The detail of this algorithm is described as follows:

Separability Index - Algorithm

Input

- (1) Training dataset D_w with m samples $\{x_i, y_i\}, i = 1, \dots, m$ of $n_{classes}$ which is the total number of defect types, where x_i is an instance in the k -dimensional feature space X and $y_i \in Y = \{1, \dots, n_{classes}\}$ is the identity label associated with x_i .
- (2) To consistently compare features, we scale features to make sure they are on a similar scale. Get every feature into approximately $(-1 \leq x_i \leq 1)$ range and has almost zero mean by replacing x_i with $(x_i - \mu_i) / s_i$, where μ_i is a mean value of x_i in training set, and s_i is a range (*max - min*) or standard deviation.

Procedure

For each feature, we perform the following steps to calculate the separation index:

- (1) Calculate the average sparseness of classes (SoC): First, the standard deviation values of corresponding classes are computed. Mean of these values describes an average sparseness of the clusters represented by the feature.

$$SoC = \frac{\sum_{n=1}^{n_{classes}} std(x_i \in class(n))}{n_{classes}} \quad (1)$$

- (2) Calculate the degree of separation (DoS): we first determine the mean value of each class and then calculate the standard deviation of the mean. This value represents the distinction between clusters that are described by the feature.

$$DoS = std(\bar{x}_n, \bar{x}_n = mean(x_i \in class(n))) \quad (2)$$

- (3) Calculate a separability index (SI): SI is the ratio between the degree of separation (DoS) and the sparseness of the classes (SoC). If SI is high, it means the feature is considered as the discriminative feature.

$$SI = DoS / SoC \quad (3)$$

After performing the steps of the SI algorithm, we choose ten features with the highest separability index, as the most distinguishable features in the feature pool. These features create a feature vector as input of a classifier.

2.2 The Architecture of Two-Tier Classifier

In this study, we use a two-tier classifier for classifying faults. The first tier is a combination of 8 SVM-Groups, corresponding to the eight types of faults which need to be classified. Each SVM-Group consists of four SVM members which trained by a feature vector extracted from the corresponding sub-band. Eight SVMs in the 8 SVM-Groups use the feature vector as an input that are extracted from the first sub-band for training using the one-against-all multi-class (OAA MC) SVM framework. We then carry out the above procedure for SVMs that use feature vectors sequentially derived from the second, third and fourth sub-bands.

SVM is a supervised binary classification algorithm that finds the hyperplane with a maximum width in the feature space of the two classes. Nonlinear classification problems can be solved by mapping samples in high-dimensional feature spaces by kernel functions [5]. In this study, we use the Gaussian radial basis function (RBF) as a kernel function and using the Lagrange multipliers to solve the optimization problem. The corresponding classification function for the SVM is as follows:

$$F(x) = \text{sgn}\{f(x)\}$$

$$\text{where } f(x) = \sum_{i=1}^n \alpha_i^* y_i^* K(x_i^*, x) + b \tag{4}$$

Here, α_i^* represents the Lagrange multiplier corresponding to the support vector (x_i^*, y_i^*) . The decision value $f(x)$ represents the signed distance of an unknown observation x from the decision boundary and has ranged between $-\infty$ to $+\infty$. Figure 3 shows outputs of SVMs at the first stage, corresponding to 240 signal samples of 8 fault classes. As can be shown in the figure, the SVM-Groups reduces the ten original features into decision values that are more discriminative.

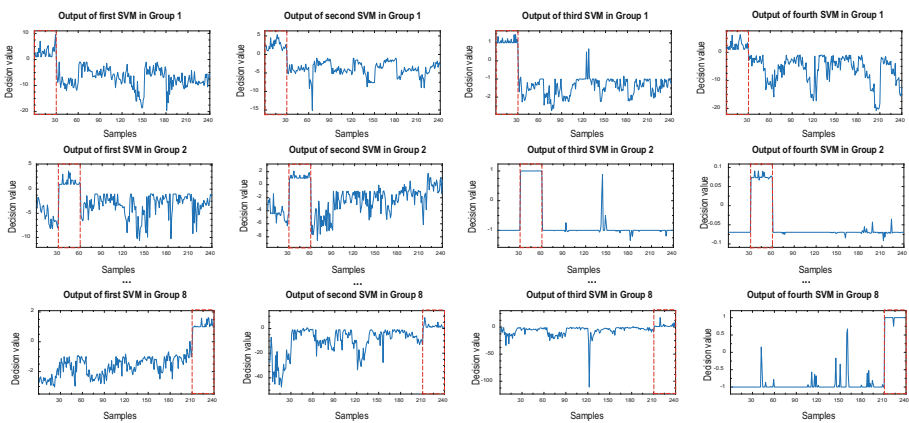


Fig. 3. The decision values of SVMs in SVM-groups

The second tier of the TTC is a multilayer perceptron (MLP) that uses the decision values of the SVMs as input. In this study, we use the MLP with one input layer, one hidden layer, and one output layer. The size of the input layer depends on the number of SVMs used at the first tier, and the size of the output layer depends on the number of classes that MLP needs to distinguish. We also use the Bayesian regularization backpropagation algorithm that based on Levenberg-Marquardt optimization to train MLP. Instead of using the standard cost function, this algorithm uses the new cost function, which is a combination of squared errors and weights that make the neural network less likely to overfitting by forcing it to have smaller weights and biases, thereby producing a well-generalized network.

3 Experimental Results and Discussion

We evaluate the effectiveness of the proposed method by using acoustic emission (AE) signals obtained from a bearing in both free-defect state (BFD) and seven different defects. These localized defects include single defects such as a crack in the outer raceway (BCO), crack on the inner raceway (BCI), crack on the roller (BCR); and combine-defects such as cracks on the inner and outer raceways (BCIO), cracks on the outer raceway and roller (BCOR), cracks on the inner raceway and roller (BCIR), and cracks on the inner raceway, outer raceway, and roller (BCIOR). To simulate the cases when bearing shaft changes during operation as in reality, we measure data at six different rotational speeds from 250 revolutions per minute (r/min) to 500 (r/min) with an increase of 50 (r/m). For each condition and at each rotational rate, we use 90 data samples with a signal length of 5 s to construct dataset. Thus, each dataset consists of $N_{RPM} \times N_{Classes} \times N_{Signals}$ or 4320 samples, where N_{RPM} is the number of operating speeds for which the AE signals are recorded ($N_{RPM} = 6$), $N_{Classes}$ is the total number of defect types or bearing conditions ($N_{Classes} = 8$), and $N_{Signals}$ is the total number of AE signals recorded for each bearing condition at each shaft speed ($N_{Signals} = 90$). We use two datasets in this study; each dataset corresponds to a crack size of 3 mm or 12 mm. The datasets are presented in detail as given in Table 1.

Table 1. Description of the datasets for two different crack sizes.

Dataset with single and compound seeded bearing failures		Revolutions per minute (r/min)	Crack size		
			Length	Width	Depth
Dataset 1	Training set	AE signals for 6 different rates*	3 mm	0.35 mm	0.30 mm
	Test set	AE signals for 6 different rates*		0.35 mm	0.30 mm
Dataset 2	Training set	AE signals for 6 different rates*	12 mm	0.49 mm	0.50 mm
	Test set	AE signals for 6 different rates*		0.49 mm	0.50 mm

*Each dataset contains AE signals recorded at six different rotational speeds of 250 r/min, 300 r/min, 350 r/min, 400 r/min, 450 r/min, and 500 r/min

To demonstrate the effectiveness of the proposed method, we compare the proposed method with envelope analysis-based methods in [4] and [4]* in terms of the average classification accuracy and average sensitive values. In the [4]*, instead of using all 22 features as the original method, we use only 9 features which are the narrow-band RMSF values of bands around each BPFI, BPFO, 2xBSF, up to third harmonics. These features are then evaluated and selected using outlier-insensitive hybrid feature selection algorithm. Finally, selected features are used to detect bearing faults using the k -NN classifier. Table 2 presents the result of the experiment. It is clear that the proposed method yields a better result than the method in [4]*, as this method only uses information about defect frequencies to diagnose bearing faults. The average classification accuracies of the proposed method are 99.1% and 99.6% for dataset 1 and dataset 2, respectively, and represent a significant improvement in comparison to methods in [4] and [4]*.

Table 2. Average classification accuracies and sensitivities for single and multiple combined bearing of variant bearing speed datasets

Datasets	Methodologies	Average sensitivity of each fault types								ACA (%)
		BCI	BCO	BCR	BCIO	BCIR	BCOR	BCIOR	BNC	
Dataset 1	[4]*	87.0	60.3	84.4	40.7	63.7	63.7	61.4	60.3	65.2
	[4]	96.6	80.3	98.8	75.9	80.7	93.3	86.2	85.1	87.1
	Proposed method	100	99.6	100	99.2	100	97.7	96.6	99.6	99.1
Dataset 2	[4]*	96.2	88.8	58.5	76.6	69.2	50.3	80.7	97.0	77.2
	[4]	90.7	90.3	80.7	95.9	85.5	93.7	98.1	95.1	91.2
	Proposed method	98.8	100	100	100	98.8	99.6	99.6	100	99.6

*Method in [4] but only use 9 features in the envelope spectrum domain.

4 Conclusions

In this paper, a new method was presented to diagnose bearing faults in case of variable operating conditions such as 6 different shaft speeds. In the proposed method, the original signal was split into multiple sub-bands, where each sub-band was analyzed to select the most discriminative features to form a feature vector for training SVMs and MLPs in the TTC classifier framework. Our experimental result showed that the proposed method could efficiently diagnose both single and combined bearing defects under variable rotational speed conditions with a high classification accuracy.

Acknowledgements. This work was supported by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, Industry & Energy (MOTIE) of the Republic of Korea (Nos. 20162220100050, 20161120100350, 20172510102130). It was also funded in part by The Leading Human Resource Training Program of Regional Neo industry through the National Research Foundation of Korea (NRF) funded by the Ministry of Science,

ICT and future Planning (NRF-2016H1D5A1910564), and in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2016R1D1A3B03931927).

The authors declare that there is no conflict of interest regarding the publication of this manuscript.

References

1. Tra, V., Kim, J., Khan, S.A., Kim, J.-M.: Bearing fault diagnosis under variable speed using convolutional neural networks and the stochastic diagonal levenberg-marquardt algorithm. *Sensors* **17**, 2834 (2017)
2. Tra, V., Kim, J., Khan, S.A., Kim, J.-M.: Incipient fault diagnosis in bearings under variable speed conditions using multiresolution analysis and a weighted committee machine. *J. Acoust. Soc. Am.* **142**, EL35–EL41 (2017)
3. Rauber, T.W., de Assis Boldt, F., Varejão, F.M.: Heterogeneous feature models and feature selection applied to bearing fault diagnosis. *IEEE Trans. Industr. Electron.* **62**, 637–646 (2015)
4. Kang, M., Islam, M.R., Kim, J., Kim, J.-M., Pecht, M.: A hybrid feature selection scheme for reducing diagnostic performance deterioration caused by outliers in data-driven diagnostics. *IEEE Trans. Industr. Electron.* **63**, 3299–3310 (2016)
5. Ma, H., Xiong, Y., Fang, H., Gu, L.: Fault diagnosis of bearing based on fuzzy support vector machine. In: *Prognostics and System Health Management Conference (PHM)*, pp. 1–4. IEEE (2015)