Fault diagnosis of roller bearing using parameter evaluation technique and multi-class support vector machine

Cite as: AIP Conference Proceedings **1788**, 030081 (2017); https://doi.org/10.1063/1.4968334
Published Online: 03 January 2017

Didik Djoko Susilo, Achmad Widodo, Toni Prahasto, et al.

Fault Diagnosis of Roller Bearing Using Parameter Evaluation Technique and Multi-Class Support Vector Machine

Didik Djoko Susilo[1, a)], Achmad Widodo[2, b)], Toni Prahasto[2, c)], Muhammad Nizam[3, d)]

[1]*Department of Mechanical Engineering, Sebelas Maret University, Surakarta, Indonesia.*
[2]*Department of Mechanical Engineering, Diponegoro University, Semarang, Indonesia*
[3]*Department of Electrical Engineering, Sebelas Maret University, Surakarta, Indonesia*

[a)]Corresponding author: djoksus@uns.ac.id
[b)]awid@undip.ac.id
[c)]toni.prahasto@gmail.com
[d)]nizam_kh@ieee.org

Abstract. Roller bearing is one of the vital parts of a rotating machine. Bearing failure can result in serious damage of the machine. This paper aims to develop a bearing fault diagnosis method using parameter evaluation technique to improve the diagnosis accuracy. The parameter evaluation technique is used to select five features that are used as predictors in multi-class support vector machine (SVM) classification. The purpose of this feature reduction was to avoid the curse of dimensionality and to increase the accuracy of the diagnosis. The diagnosis process was performed by classification of bearing states using one-against-one method multi-class SVM. Three types of kernel functions i.e., linear, polynomial, and Gaussian RBF were used in the SVM classification. The bearing conditions which is diagnosed in this paper were normal bearing, inner race fault, and outer race fault conditions. As a result, the classification performance of multiclass SVM using five selected features as the parameter have excellent performance in predict the bearing conditions data for all types of kernel functions.

INTRODUCTION

Rotating machinery like a turbine, a pump, and a compressor is among the most important components in the industry. They are used in power plant, manufacture industry, textile industry, automotive industry, etc. It is important to maintain these machines in proper working conditions constantly. Ability to confidently determine the state of the system and predict failures would greatly increase the productivity of the plant. Rotating machinery usually is composed of different sub-systems interacting with each other in a nonlinear fashion; changes in any of these components can significantly affect the overall performance. Moreover, catastrophic failure of these machines can cause significant economic loss.

Bearings are the load-bearing members of rotating machines. They are the key to the effective functioning of the machine and often the cause of failure. Hence it is critical to be able to detect and examine a faulty bearing. Roller bearing is one of the most widely used machinery components. It directly influences the operation of the whole machinery. Unexpected roller bearing failures can interrupt the production, cause unscheduled downtime and economic losses. So the development of proper condition monitoring and fault diagnosis procedure to prevent malfunction and failure of roller bearings during operation is necessary. As a result, the fault diagnosis of rolling element bearings has been the subject of extensive research in recent years. According to former studies [1], most roller bearing faults occur on the surface of the outer race, inner race or rolling elements. To identify the most probable faults leading to failure, many methods are used for data collection, including vibration monitoring,

International Conference on Engineering, Science and Nanotechnology 2016 (ICESNANO 2016)
AIP Conf. Proc. 1788, 030081-1–030081-12; doi: 10.1063/1.4968334
Published by AIP Publishing. 978-0-7354-1452-5/$30.00

thermal imaging, oil particle analysis, etc. Vibration signal analysis proves its worthiness in costs, effective and convenience reasons, and these advantages make the justification why vibration signal analysis is widely studied and used as a state monitoring approach for roller bearing [2].

Generally, the key problem of vibration signal monitoring is feature extraction. Many techniques have been developed in this area such as conventional feature extraction techniques can be concluded as time domain analysis [3], frequency domain analysis [4] and time-frequency domain analysis [5]. However, in most cases, vibration signals are too noisy to be used in these techniques because the signal collected from roller bearing is mixed with many other signal sources. The features generated by the incipient fault are usually very weak and might be covered by other signals. This property makes the diagnosis methods based on vibration signal becomes unreliable. Thus, effective extraction of early fault symptom is still a critical challenge.

Modern signal processing methods for fault diagnosis can be divided into two categories. Both categories of signal processing methods are shaping the development trend in the fault diagnosis field. The first category is the direct signal analysis or decomposition, such as short-time fast Fourier transform (STFT), wavelet transform, principal component analysis (PCA), blind source separation.They are mainly used in the signal basis processing or preprocessing, for instance, the time-frequency domain analysis, feature extraction, noise reduction and others. Some applications of this technique in bearing fault detection are motor induction bearing fault detection using FFT of the motor current signal [6]. Furthermore, fault feature extraction of rolling bearing based on empirical mode decomposition (EMD) [7] and detection of roller bearing system using a wavelet denoising scheme and proper orthogonal value of an intrinsic mode function covariance matrix [8]. According to the study of [9], rolling element bearing fault diagnosis is best using high-frequency resonance technique (HFRT) combine with defect frequency analysis (DFA) or continuous wavelet transform(CWT) and hilbert huang transform (HHT) amplitude. The second is the processing based on intelligence algorithms, such as artificial neural networks (ANN), support vector machines (SVM), Bayes classification, decision tree, etc. Some applications of this technique are bearing fault detection using ANN and genetic algorithm [10], fault diagnosis of low speed bearing based on relevance vector machine (RVM) and SVM using independent component analysis (ICA) and PCA for feature extraction [11], bearing fault classification by ANN and Selfs Organizing Maps (SOM) using wavelets parameters [12], bearing fault diagnosis based on multiscale permutation entropy and support vector machine [13], feature extraction using wavelet transform and fault classification using multiclass SVM [14] and [15], usage of peak to average ratio of bearing fault spectrum as fault indicator and SVM for bearing fault classification [16]. The usage of the SVM method has been summarized for machine fault diagnosis [17]. It was used for fault diagnosis in rolling element bearing, induction motors, machine tools, rotating machines, HVAC machines, and other machines.

Figure 1 illustrates the general structure of a roller bearing. It is composed of six components: housing, outer race, inner race, rolling elements, cage and shaft [18]. The interaction of defects in rolling element bearings produces impulses of vibration. As these shocks excite the natural frequencies of the bearing elements, the analysis of the vibration signal in the frequency domain by means of the Fast Fourier Transform (FFT) has been an effective method for predicting the health condition of the bearings [19].

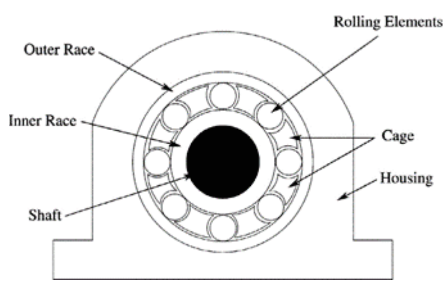


FIGURE 1. Typical roller bearing.

Each defective bearing component produces different frequencies, which allows for localizing different defects occurring simultaneously. Ball Pass Frequency on an Outer race defect (BPFO), ball pass frequency on an inner race defect (BPFI), fundamental train frequency (FTF) and ball spin frequency (BSF) – as well as their harmonics, modulating frequencies, and envelopes – are examples of frequency-domain indicators, calculated from kinematic considerations – that is, the geometry of the bearing and its rotational speed [20].

In addition to frequency, time domain indicators have been widely employed as input features to train a bearing fault diagnosis classifier. Time-domain indicators are scalar indicators that allow for representing the vibration signal through a single scalar value. For instance, the peak is the maximum amplitude value of the vibration signal, rms (root mean square) represents the effective value (magnitude) of the vibration signal and kurtosis describes the impulsive shape of the vibration signal.

With the development of the computer and information technologies, traditional data-based condition monitoring and fault diagnosis is gradually replaced by featured-based due to the fast transfer, small storage space and high accuracy. Here, feature means some value that can represent machines conditions. The fault diagnosis is performed by means of machine learning techniques such as: artificial neural network, genetic algorithm, fuzzy reasoning, and support vector machine. Machine learning refers to a system capable of autonomous acquisition and integration knowledge. This capacity to learn from experience, analytical observation, and other means, result in a system that can continuously self-improve and thereby offer increased efficiency and effectiveness [21].

An accurate diagnosis for complex machines needs several sensors to obtain details of information about condition, which result in plenty of raw data. Thereby, many numbers of features are calculated and stored to keep information of the machine condition at the highest level. However, a lot of additional features make the computational time and cost more and the ability of the diagnostics system to make an efficient diagnosis is decreased. Moreover, irrelevant features complicate the whole diagnostics system and in turn reducing its capability to make an effective analysis. Too many features can cause curse of dimensionality and peaking phenomena that substantially degrade classification accuracy. Also, many features still can bring traffic jam or storage problem in the use and maintenance of data.

To overcome the above situations, it is needed feature dimensionality reduction. There are two methods that can be applied, i.e. feature extraction and feature selection. Feature extraction is a method that creates new features based on transformations and combinations of the original features set. The term feature selection refers to algorithms that select the best feature subset from all features. Often feature extraction proceeds feature selection; firstly feature dimensionality is greatly reduced by extraction and then the significant features are selected from transformed features. Feature extraction leads to cost saving in computation time.

Feature selection contributes to monitoring and diagnosis accuracy. In the feature selection process, the number of selected features is decreasing to value that even acceptable to accomplish proper further diagnosis. Removing the irrelevant features is one of the most accepted methods in this field. Each feature should be ranked on priority and less important one to be removed and the others are applied. This paper presents parameter evaluation technique based on the distance evaluation technique to reduce the number of features feed into SVM classifier. Furthermore, the performance of fault classification using all features and selected feature is compared.

METHODOLOGY

The present study attempts to develop a bearing fault diagnosis using features reduction through parameter evaluation. The methodology of this study is as shown in Fig. 2.

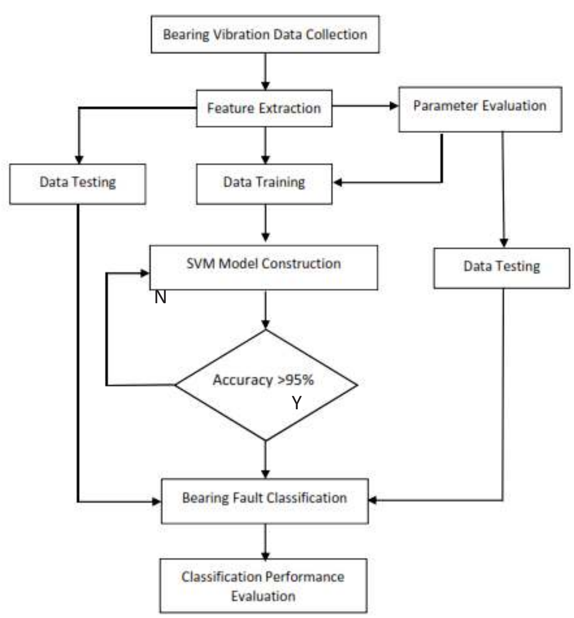


FIGURE 2. Proposed method of the present study.

Bearing Data Collection

Bearing vibration data was obtained from the MFPT bearing data set available at http://data-acoustics.com/. There were measurement of three bearing condition, i.e. normal bearing, faulty innerrace bearing, and faulty outerrace bearing. The bearing data were as follow:

- Bearing Type : NICE
- Number of balls : 8
- Ball Diameter : 5.97mm
- Pitch Diameter : 31.62mm
- Contact Angle : 0
- FTF : 0.5935 x (shaft speed)
- BPFO : 3.245 x (shaft speed)
- BPFI : 4.755 x (shaft speed)
- BSF : 2.5564 x (shaft speed)
- Nominal speed : 25 Hz

The vibration measurement was conducted for three operating conditions:

1. Baseline – no fault, sampling rate of 97,656 Hz, a load of 270 lbs, and record length of 6 seconds.
2. Outer race fault with the same load of 270 lbs, the sampling rate of 97,656 Hz, and record length of 6 seconds and outer race fault with variable load, the sampling rate of 48,828 Hz, and record length of 3 seconds.
3. Inner race fault with variable load, the sampling rate of 48,828, and record length of 3 seconds.

Thirty data were collected from each bearing condition. Each data contains 10,000 points. So, the total number of the bearing data were 90 set. The data for bearing fault condition were acquired from different loads.

Feature Extraction

Features are some representatives values which can indicate bearing conditions. The represented features include time domain features such as mean, root mean squares (RMS), variance, skewness, kurtosis, etc., frequency domain features such as content at the feature frequency, the amplitude of FFT spectrum, etc., and time-frequency domain features such as statistical characteristics of short time Fourier Transform (STFT), Wigner-Viller distribution, wavelet transform, etc.

In this study, there are 21 features extracted from the vibration signal. They are mean, rms, shape factor, skewness, kurtosis, crest factor, entropy estimation value, entropy estimation error, histogram upper bound, histogram lower bound, rms frequency, frequency center value, root variance frequency value, and first 8 order coefficients of auto-regression (AR) model.

All of the 21 features were computed from each bearing measurement using MATLAB code. Therefore it was obtained a 90 by 21 feature matrix. These features would be fed to the SVM model for bearing fault classification.

Parameter Evaluation

Parameter evaluation means a feature selection to choose the features that are connected to the classification model construction. This feature selection techniques are used for three reasons:

- Simplification of models for ease of interpretation,
- Shorter training time,
- Enhanced generalization by reducing overfitting.

Feature selection process directly reduces the number of original features by selecting a subset of them that still that still retains sufficient information for classification. Usually, a large number of features often include many garbage features. Such features are not only useless in classification, but also sometimes degrade the performance of a classifier which is designed by a finite number of training samples. In such a case short, removing the garbage features can improve the classification accuracy.

In this paper, parameter evaluation or feature selection was made using distance evaluation technique. This technique consists of four steps as follows:

Step 1: calculating the average distance of the same condition data (dij), followed by getting the average distance of the total conditions (dai). The equation can be defined as follows:

$$d_{i,j} = \frac{1}{Nx(N-1)} \sum_{m,n=1}^{N} \left| p_{i,j}(m) - p_{i,j}(n) \right|; \ (m,n = 1,2,\ldots,N, \ m \neq n) \tag{1}$$

where N is the number of the same condition, $p_{i,j}$ is the eigenvalue, $d_{i,j}$ is the average distance of the same condition, i and j represent the number of parameters and conditions respectively.

$$d_{ai} = \frac{1}{M} \sum_{j=1}^{M} d_{i,j} \tag{2}$$

where M is the number of different conditions.

Step 2: calculating the average distance between different condition data (d_{ai}).

$$d'_{ai} = \frac{1}{M\,x(M-1)} \sum_{m,n=1}^{M} \left| p_{ai,m} - p_{ai,n} \right|; \ (m,n = 1,2,\ldots,M; m \neq n) \tag{3}$$

Where d'_{ai} is the average distance of different conditions data, $p_{ai,j}$ is the average value of the same condition data.

$$p_{ai,j} = \frac{1}{N} \sum_{n=1}^{N} p_{i,j}(n); \ (n = 1,2,\ldots,N) \tag{4}$$

Step3: calculating the ratio d_{ai} / d'_{ai}

Step4: selecting the feature parameters, from large value to small value. The dai is the smaller, the better, whereas d'_{ai} is the bigger , the better. So, bigger represents the feature well.

$$\alpha_i = d'_{ai} / d_{ai} \tag{5}$$

Where α_i is the effectiveness factor of features.

Support Vector Machine (SVM)

Support vector machine (SVM) is a computational learning method based on statistical learning theory introduced by [22]. The basic characteristic of the SVM model is to map the original nonlinear data into a higher-dimensional feature space where a hyperplane is constructed to bisect two classes data and maximize the margin of separation between itself and those points lying nearest to the support vectors.

Given data input $\mathbf{x}_i$ (i = 1,2, ..., M), M is the number of sample. The samples are assumed two classes (binary classifier), namely positive class and negative class. Each of classes associates with labels be $y_i = 1$ for positive class and $y_i = -1$ for negative class, respectively. In the case of linearly separating data, the hyperplane $f(\mathbf{x}) = 0$ that separates the data is given by:

$$f(\mathbf{x}) = \mathbf{w}^{\mathbf{T}}\mathbf{x} + b = \sum_{j=1}^{M} w_j x_j + b = 0 \tag{6}$$

Where, **w** is M-dimensional vector and b is a scalar. The vector **w** and scalar b are used to define the position of separating hyperplane. The decision function is made using sign f(x) to create separating hyperplane that classifies input data in either positive class and negative class. A distinctly separating hyperplane should satisfy the constraints:

$$y_i f(\mathbf{x}_i) = y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 \ \ for \ i = 1, 2, \ldots, M \tag{7}$$

The separating hyperplane that creates the maximum distance between the plane and the nearest data, i.e., the maximum margin, is called the optimal separating hyperplane. The illustration of the optimal separating hyperplane is shown in Fig. 3 below.

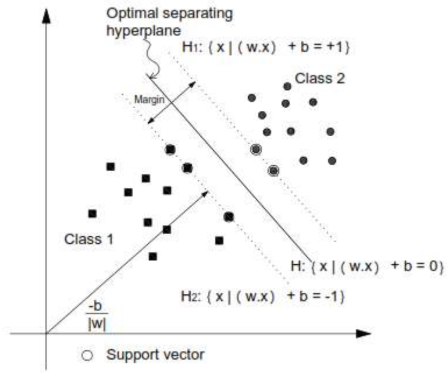


FIGURE 3. Optimal separating hyperplane.

From the geometry, the geometrical margin is found to be $\|\mathbf{w}\|^2$. Taking into account the noise with slack variable ξ_i and the error penalty C, the optimal hyperplane separating the data can be obtained as a solution to the following optimization problem.

$$\text{Minimize } \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{M}\xi_i \tag{8}$$

$$\text{Subject to } \begin{cases} y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 - \xi_i, & i = 1,\ldots,M \\ \xi_i \geq 0 & i = 1,\ldots,M \end{cases} \tag{9}$$

Where ξ_i is measuring the distance between the margin and the example $\mathbf{x}_i$ that lying on the wrong side of the margin. The calculation can be simplified by converting the problem with Kuhn-Tucker condition into the equivalent Lagrangian dual problem, which will be

$$\text{Minimize } L(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{M}\alpha_i y_i(\mathbf{w}.\mathbf{x}_i + b) + \sum_{i=1}^{M}\alpha_i \tag{10}$$

The task is minimizing Eq. (10) with respect to w and b, while requiring the derivatives of L to α to vanish. At optimal point, it has the following saddle point equations:

$$\frac{\partial L}{\partial \mathbf{w}} = 0, \qquad \frac{\partial L}{\partial b} = 0 \tag{11}$$

Which replace into form

$$\mathbf{w} = \sum_{i=1}^{M}\alpha_i y_i \mathbf{x}_i, \qquad \sum_{i=1}^{M}\alpha_i y_i = 0 \tag{12}$$

From Eq. (12), it can be seen that w is contained in the subspace spanned by the xi. Substitution eq. (12) into Eq. (10) will get dual quadratic optimization problem as follow:

$$\text{Maximize } L(\alpha) = \sum_{i=1}^{M}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{M}\alpha_i\alpha_j y_i y_j x_i x_j \tag{13}$$

$$\text{Subject to } \begin{array}{l} \alpha_i \geq 0, \quad i = 1,\ldots,M \\ \sum_{i=1}^{M}\alpha_i y_i = 0 \end{array} \tag{14}$$

Thus, by solving the dual optimization problem, one obtains the coefficients α_i which required to express the **w** to solve Eq. (8). This lead to non-linear decision function.

$$f(x) = sign\left(\sum_{i,j=1}^{M}\alpha_i y_i(\mathbf{x}_i\mathbf{x}_j) + b\right) \tag{15}$$

SVM can also be used in non-linear classification tasks with the application of kernel functions. The data to be classified is mapped onto a high-dimensional feature space where the linear classification is possible using nonlinear vector function $\Phi(\mathbf{x})$. In the non-linear classification, kernel function will return the dot product of the features space mappings of the original data points, stated as $K(\mathbf{x}_i, \mathbf{x}_j) = (\Phi^T(\mathbf{x}_i).\Phi_j(\mathbf{x}_j))$. When applying a kernel function, the learning in the feature space does not require explicit evaluation of Φ and the decision function will be:

$$f(x) = sign\left(\sum_{i,j=1}^{M} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}_j) + b\right) \tag{16}$$

There are different kernel functions used in SVM, such as linear, polynomial, and Gaussian RBF. The selection of the appropriate kernel function is very important, since the kernel defines the feature space in which the training set examples will be classified. The formulation of the kernel function is given in Table 1 below. In this study, the classification used these three type kernel functions and the results were compared.

TABLE 1. Formulation of kernel functions

Kernel	$K(\mathbf{x}_i, \mathbf{x}_j)$
Linear	$\mathbf{x}_i^T \mathbf{x}_j$
Polynomial	$(\gamma \mathbf{x}_i^T \mathbf{x}_j + r)^d,\ \gamma > 0$
Gaussian RBF	$e^{-\frac{\lVert \mathbf{x}_i - \mathbf{x}_j \rVert^2}{2\gamma^2}}$

In the real world problem, it is frequently found more than two class problems. Roller bearing fault case for instance, there are a normal bearing condition, inner race fault, outer race fault, and ball fault. To handle multiclass classification using SVM, there are two types approach. One is by constructing and combining several binary classifier or one-against-one, while the other by considering all the data in one optimization formulation or one-against all.

In the one-against-all method, it constructs *k* SVM models where *k* is the number of classes. The *i*th SVM is trained with all of examples in the *i*th class with positive labels, and all the other example with negative labels. Thus given l training data $(x_1, y_1), \ldots, (x_l, y_l)$, where $x_i \in R^n,\ i = 1, \ldots, l.$ and $y_i \in \{1, \ldots, k\}$ is the class of x_i. The *i*th SVM solves the following problem:

$$\text{Minimize:} \quad \frac{1}{2}\left\lVert \mathbf{w}^i \right\rVert^2 + C\sum_{i=1}^{l} \xi_j^i (\mathbf{w}^i)^T \tag{17}$$

$$\text{Subject to } (\mathbf{w}^i)^T \phi(x_j) + b' \geq 1 - \xi_j^i,\ \ if\ \ y_j = i \tag{18}$$

$$(\mathbf{w}^i)^T \phi(x_j) + b' \leq -1 + \xi_j^i,\ \ if\ \ y_j \neq i \tag{19}$$

$$\xi_j^i \geq 0, \quad j = 1, \ldots, l \tag{20}$$

where the training data $\mathbf{x}_i$ is mapped to a higher dimensional space by function ϕ and C is the penalty parameter.

Minimizing eq. (17) means we would like to maximize $2/\lVert \mathbf{w}_i \rVert$, the margin between two groups of data. When data is not separable, there is a penalty term $C\sum_{i=1}^{l} \xi_{i,j}$ which can reduce the number of training errors. After solving Eq. (17), there are k decision functions:

$$(\mathbf{w}^1)^T \phi(x) + b^1,$$
$$\vdots$$
$$(\mathbf{w}^k)^T \phi(x) + b^k$$

We say x is in the class which has the largest value of the decision function:

$$\text{class of } x \equiv \arg\max_{i=1,\ldots,k} (\mathrm{w}^i)^T \phi(x) + b^i) \tag{21}$$

The other method for multi-class SVM classification is called one-against-one method. This method constructs $k(k$-1)/2 classifiers where each one is trained on data from two classes. For training data from ith and the jth classes, we solve the following binary classification problem.

$$\text{Minimize:} \quad \frac{1}{2}\left\|\mathbf{w}^{ij}\right\|^2 + C\sum_t \xi_t^{ij} (\mathbf{w}^{ij})^T \tag{22}$$

$$\text{Subject to:} \quad (\mathbf{w}^{ij})^T \phi(\mathbf{x}_t) + b^{ij} \geq 1 - \xi_t^{ij}, \ \ if \ \ y_t = i \tag{23}$$

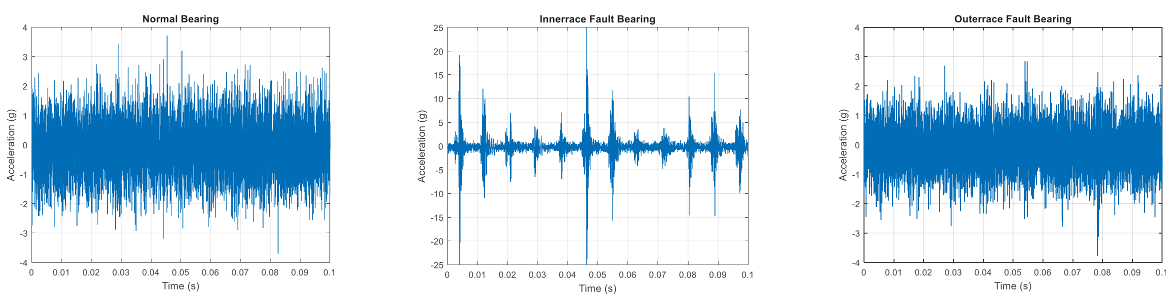
$$(\mathbf{w}^{ij})^T \phi(\mathbf{x}_t) + b^{ij} \leq -1 + \xi_t^{ij}, \ \ if \ \ y_t \neq i \tag{24}$$

$$\xi_t^{ij} \geq 0, \quad j = 1, \ldots, l \tag{25}$$

There are different methods for doing the future testing after all $k(k$-1)/2 classifiers are constructed. After some tests, we decide to use the following voting strategy: if the sign $((\mathbf{w}^{ij})^T \phi(\mathbf{x}) + b^{ij})$ says $\mathbf{x}$ is in the ith class, then the vote for the ith class is added by one. Otherwise, the jth is increased by one. Then we predict x is in the class with the largest vote. The voting approach described is called the “Max Wins” strategy.

RESULTS AND DISCUSSIONS

The vibration signals were collected from three different bearing conditions. Thirty data were collected from each bearing condition. Each data contains 10,000 sampling points. So, the total number of the bearing data were 90 set. The data for bearing fault conditions were acquired from different loads. The time-domain plots of the raw vibration signals are shown in Fig. 4.



(a) Normal bearing (b) Innerrace fault (c) Outerrace fault

FIGURE 4. Raw vibration signal of the roller bearings.

The statistical features were extracted from each vibration signals. Twenty-one features including those in time domain, frequency domain, and time-frequency domain were obtained. So, the dimension of the feature matrix was 90 by 21. These features were then divided into 60 training data and 30 testing data that were used in the construction of the SVM model dan testing scheme. Parameter evaluation technique using distance evaluation was used to select five features to reduce the dimension of the feature matrix. The features were ranked based on the effectiveness factor from the highest to the lowest. So that, the feature matrix dimension would decrease to 90 by 5. The result of the parameter selection is shown in Fig. 5. The features selected were feature number 8, 2, 1, 10, and 5. These features were entropy estimation error, rms, mean, histogram lower bound, and kurtosis respectively.

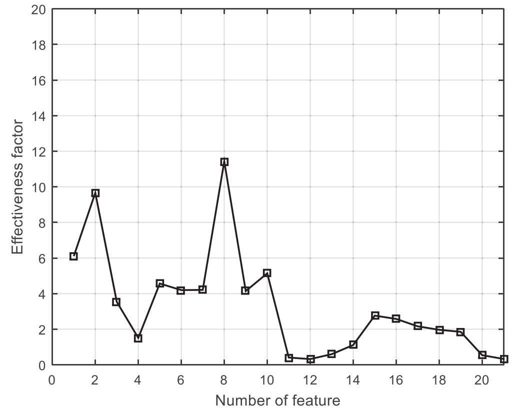


FIGURE 5. The effectiveness factor of the features.

The SVM model to classify bearing fault was constructed used the one-against-one method. Two SVM models were constructed. One used all features as their parameter, and the other one used the five selected features. Three kernel function types, i.e: linear, polynomial, and Gaussian RBF were used in the SVM models. After the SVM model constructed, the models were tested using the test data to predict the conditions of the bearings, and then the performance of the classification was evaluated. The results are shown in the Fig. 6 to Fig. 11.

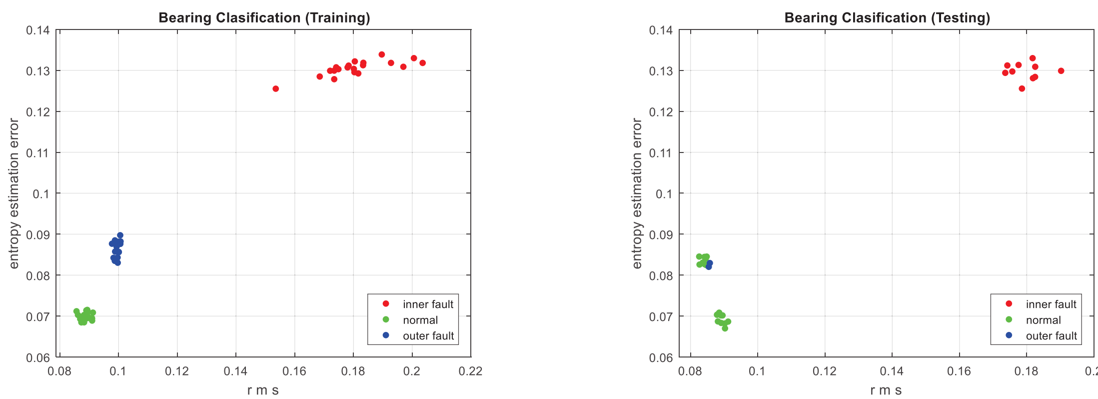


FIGURE 6. SVM classification result using all features and linear kernel function.

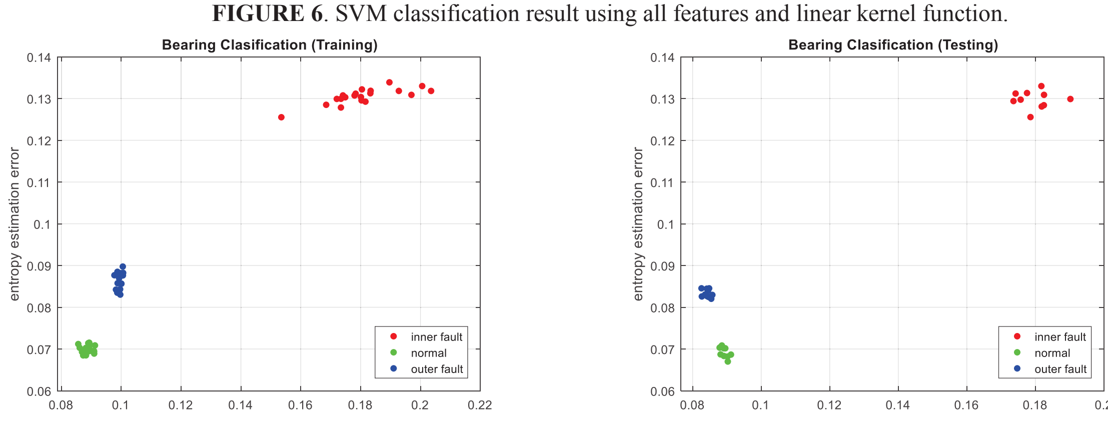


FIGURE 7. SVM classification results using all features and polynomial kernel function.

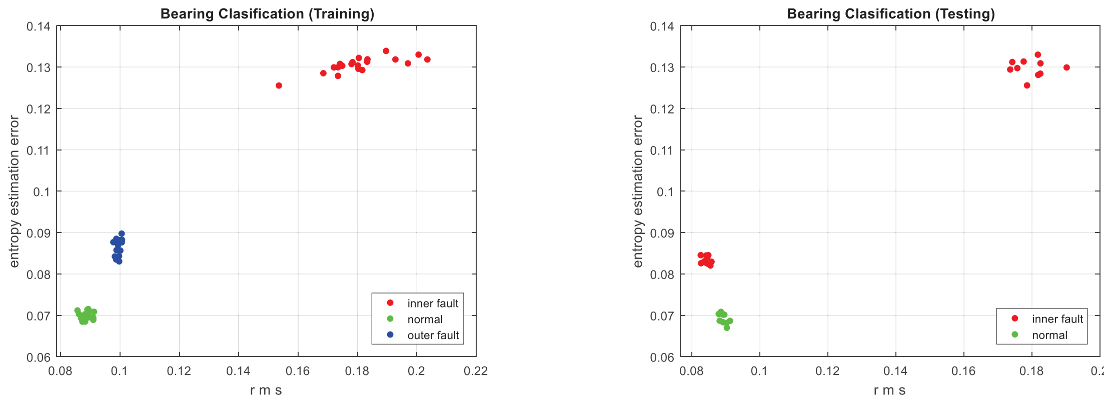


FIGURE 8. SVM classification results using all features and Gaussian RBF kernel function.

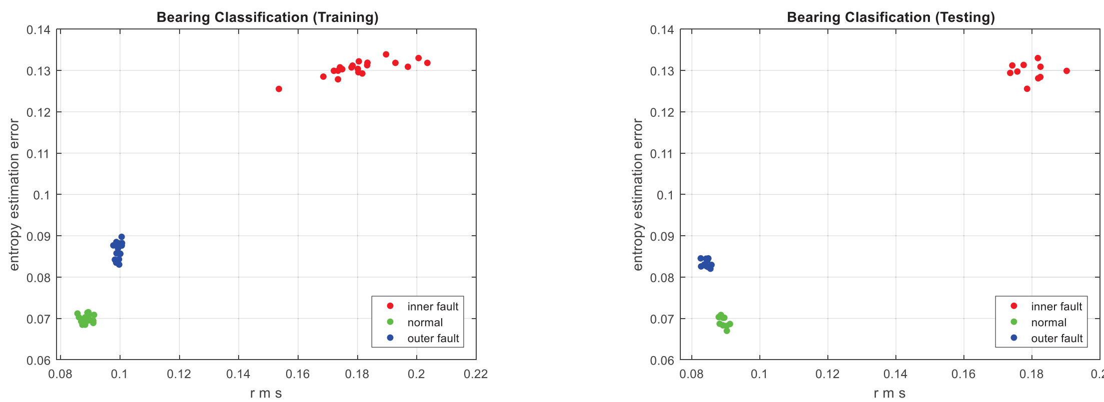


FIGURE 9. SVM classification result using 5 selected features and linear kernel function.

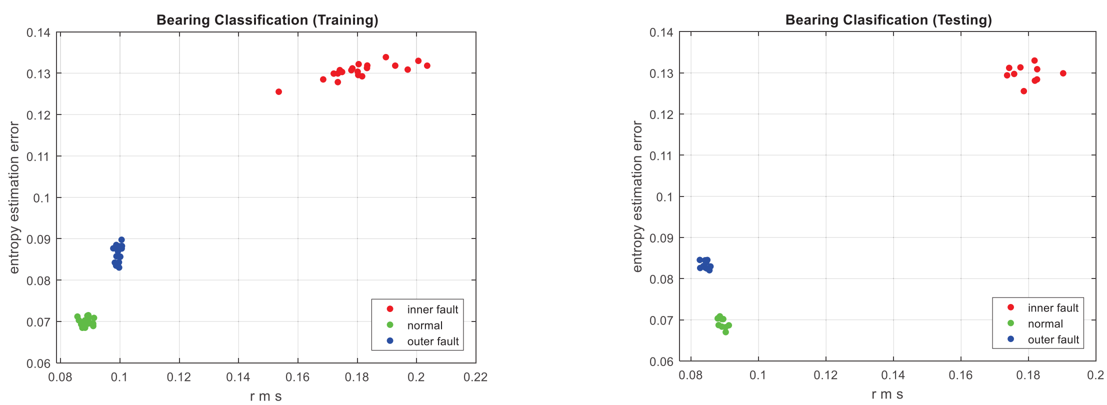


FIGURE 10. SVM classification results using 5 selected features and a polynomial kernel function.

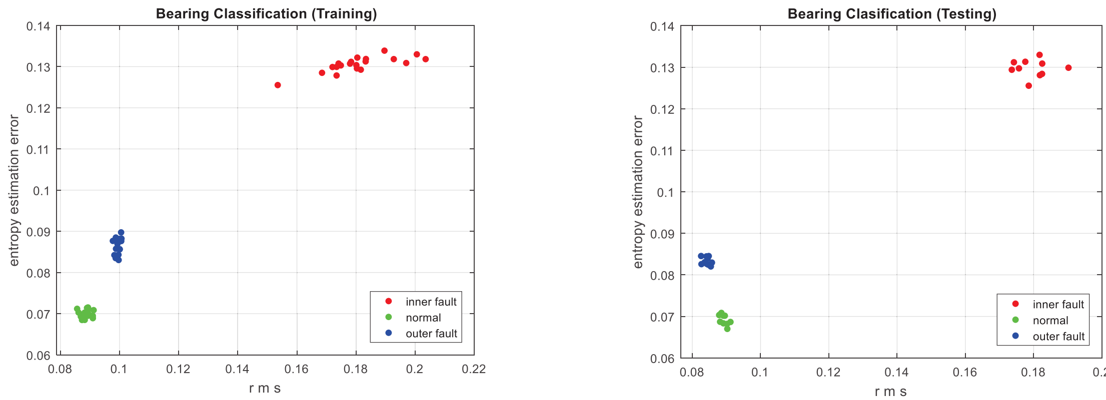


FIGURE 11. SVM classification result using 5 selected features and Gaussian RBF kernel function.

The figures above show that the accuracy of the SVM models using all features and five selected features are the same, i.e., 100% accuracy, but there is an observed difference of the performances in predicting new data. The performance is summarized in Table 2 below.

TABLE 2. The performance of the SVM Classifier.

	Linear		Polynomial		Gaussian RBF	
	Model	Testing	Model	Testing	Model	Testing
All Features	100 %	73 %	100 %	100 %	100%	67 %
Five selected features	100 %	100 %	100 %	100 %	100 %	100 %

TABLE 2 shows that the use of five selected features improves the bearing classification accuracy. The accuracy increases 17 % for linear kernel function and 13 % for Gaussian RBF kernel function. The accuracy in predicting new data reaches 100 %. On the other hand, when using all extracted features, high accuracy in prediction was found using polynomial kernel function and the accuracy are down when using linear and Gaussian RBF kernel function. The performance of this SVM classifier using parameter evaluation technique is slightly better than bearing faults diagnosis based on multiscale permutation entropy, and SVM used five features conducted by Wu et.al. in reference [13]. The accuracies of their work are in the range of 99.17 % to 99.77 %.

CONCLUSION

This paper has presented the parameter evaluation technique to select bearing fault features from the roller bearing vibration signals for reducing dimensions and increasing the performance of bearing fault diagnosis. As a result, the classification performance of multiclass SVM using five selected features as the parameter have excellent performance in predict the bearing conditions data for all types of kernel functions. The five selected features that well represent the bearing faults are entropy estimation error, rms, mean, histogram lower bound, and kurtosis.

REFERENCES

1. O. Sadettin, A. Nizami and C. Veli, NDT & E International **39**, pp. 293–298 (2006).
2. G. Wang, Y. He and K. He, Journal of Software **7**, pp. 1531-1538 (2012).
3. P. K. Kankar, C. S. Satish and S. P. Harsha, Expert Systems with Applications **38**, pp. 1876–1886 (2011).
4. N. Tandon and A. Choudhury, Journal of Sound and Vibration **205**, pp. 275–292 (1997).
5. H. Li, Journal of Computers **6**, pp. 1994-2000 (2011).
6. A. Singhal, and M.A., Khandekar, IJAREEIE **2**, pp. 3258—3264 (2013).
7. T. Han, D. Jiang and N. Wang, Shock and Vibration, Article ID 5957179, (2016).
8. J.H. Ahn, D.H., Kwak and B.H., Koh, Sensors **14**, pp.15022-15038 (2014).

9. P. Sakya, A.K. Darpe, M.S. Kulkarni, IJCM Journal **3(2)**, (2013).
10. B. Samanta, K.R. Al-Balushi, S.A. Al-Arami, EURASHIP Journal on Applied Signal Processing **3**, pp. 366-377 (2004).
11. A. Widodo, J.D. Son, B.S. Yang, Y.H. Kim, A.C.C. Tan, J. Mathew, D.S. Gu, B.K. Choi, "Fault Diagnosis of Low Speed Bearing Based on Acoustic Emission Signal and Multi-Class Relevance Vector Machine", in 15th International Congress on Sound and Vibration, 6-10 July 2008, Daejeon, pp. 1468-1475.
12. K.M. Bhavaraju, P.K. Kankar, S.C. Sharma. S.P. Harsha, IJEST **2 (5)**, pp. 1001-1008 (2010).
13. S. D. Wu, P. H. Wu, C. W. Wu, J.J., Ding, C.C. Wang, Entropy **14**, pp. 1343-1356 (2012).
14. C. Rajeswari, B. Sathiyabhama, S. Devendiran, K. Manivannan, IJMME-IJENS **15(1)**, (2015).
15. S.Vora, J.A. Gaikwad, J.V. Kulkarni, Advanced Research in Electrical and Electronic Engineering (AREEE) **2 (5)**, pp. 41-46 (2015).
16. D.H. Hwang, Y.W. Youn, J.H., Sun, K.H. Choi, J.H. Lee, Y.H. Kim, JEET **10**, pp. 30-40 (2015).
17. B.S. Yang and A. Widodo, Journal of System Design and Dynamics **2(1)**, pp. 12-23 (2008).
18. H. Guo, L. Jack, and A. Nandi, IEEE Transactions on Systems, Man, and Cybernetics, Part B, **35(1)**, pp. 89–99 (2005).
19. N. Tandon, and A. Choudhury, Tribology International **32(8)**, pp. 469–48 (1999).
20. S. Sassi, B. Badri, and M. Thomas, Journal of Vibration and Control **13(11)**, pp. 1603–1628 (2007).
21. A. Widodo, *Application of Intelligent System for Machine Fault Diagnosis and Prognosis* (Badan Penerbit Universitas Diponegoro, Semarang, 2009), pp. 21-22.
22. V.N. Vapnik, IEEE Transactions on Neural Network **10(5)**, pp. 988-999 (1999).