Bearing Fault Diagnosis using Multiclass Support Vector Machine with efficient Feature Selection Methods

C.Rajeswari¹, B.Sathiabaham², S.Devendiran³ & K.Manivannan⁴

¹Research Scholar, Faculty: Information and Communication Engineering, Anna University, Chennai, India
raji.chen@gmail.com
²Department of Computer Science and Engineering, Sona College of Technology, Salem, Tamil Nadu, India
sathiya674@yahoo.co.in
³,⁴School of Mechanical and Building Sciences, VIT University, Vellore, Tamil Nadu, India
devendiran@vit.ac.in
manivannan.k@vit.ac.in

Abstract— Fault diagnosis in bearings has been the subject of intensive research as bearings are critical component in many rotating machinery applications and an unexpected failure of the bearing may cause substantial economic losses. Vibration analysis has been used as a predictive maintenance procedure and as a support for machinery maintenance decisions. In this study, we used Multiclass Support Vector Machines (MSVM) for classification task. SVM is a powerful classification tool that is becoming increasingly popular in various machine-learning applications. In data pre-processing step advance signal processing methods such as wavelet transform is used to extract the features which reveals the characteristics of the actual conditions of the bearing. To reduce the dimensionality of the feature SVM–Recursive Feature Elimination (SVM-RFE),Wrapper subset method, Relieff method and Principle component analysis (PCA). Moreover, investigation is done performance of MSVM with mentioned different feature selection strategies as well as MSVM alone on basis of classification accuracy. To validate the proposed methodology, four kinds of running states are simulated with artificially created faults in bearing that is accommodated in the test rig. By comparing classification accuracy along with computation timing the best scheme is selected for diagnostic prediction.

Index Term— Test rig, fault diagnosis, wavelet, SVM–Recursive Feature Elimination, Wrapper subset method, Relieff method and Principle component analysis(PCA)

I. INTRODUCTION

The performance of bearing is highly influential on performance of any rotating machine and it is consider as a heart of rotating machinery. Monitoring the dynamic behaviour and health condition of significant machine element bearing is in evitable and thereby provide an alarm and instructions for preventive maintenance by means of advanced sensors, data collection systems, data storage/transfer capabilities and data analytic tools developed for such purpose. Fault prediction is one of the important undertakings during the maintenance work. Vibration signal based bearing fault diagnosing method is prominently used over the past two decades [1,2]. When a fault occurs in a bearing, periodic impulses appear in the time domain of the vibration signal, while the corresponding bearing characteristic frequencies (BCFs) and their harmonics emerge in the frequency domain [3]. However, in the early it get suppressed by severe noise and higher-level vibrations. Consequently an effective signal processing method is of utmost importance for the extraction of damage sensitive features during the condition monitoring of bearings. Generally, the failure of a mechanical system is always accompanied with the changes of vibration characteristics from linear or weak nonlinear to strong nonlinear dynamics [4,5]. Until now, in the field of bearing fault detection, a variety of approaches and time–frequency signal processing tools have been utilized. Wavelet transform (WT) has been widely used as a de-noising tool as well as for feature extraction in rotating machinery diagnostics [6]. The Wavelet Transform (WT) provides powerful multi-resolution analysis in both time and frequency domains. The fault diagnosis of rolling bearing in early stage using wavelet packet transform and empirical mode decomposition were combined and extracted features are given as input to neural network after analysing the shortcomings of current feature extraction and fault diagnosis technologies[7]. In past decades, sometime-frequency analysis methods, including short time Fourier transform (STFT) [8], Wigner–Ville distribution (WVD) [9,10], and local mean decomposition(LMD) [11], have been formulated to probe non-stationary data and applied to feature extraction of defective rotary machinery. However, each of these methods leaves something to be desired and some even perform badly in analysing non-stationary and nonlinear data [12].

The early stage rolling failure diagnostic methods include the use of hearing, shock pulse and frequency demodulation technologies with lower industry standard accuracy and efficiency in diagnosis. With the continuous development of diagnostic techniques, artificial intelligence, such as expert system, artificial neural network (ANN), fuzzy logic, Roughset, genetic algorithm have been widely used in machine fault diagnosis [13-16] based on the empirical risk minimization principles. Unfortunately, because of the bottleneck of knowledge acquisition, the application of expert system is limited. Also, for the complexity of machinery and knowing little of the fault mechanism, the diagnosis methods based on neural network and soft-computing technology need to be studied further to improve the diagnosis performance, such as increasing diagnosis accuracy and decreasing running time, etc. In intelligent fault diagnosis method, statistical characteristics were calculated.
after signal processing and it has vast number of features within variety domains poses challenges to data mining. In order to achieve successful classification process in terms and prediction accuracy, feature selection is considered essential. Feature selection methods such as principal component analysis (PCA) [7] and Genetic algorithm (GA) [8] and J48 algorithm [9] are widely used to decrease dimensions of features proved for its meaningful diagnostics process. In the present work, most suitable signal processing technique based on sub-band coding and known for its fast computation among the wavelet family called discrete wavelet decomposition is used for extracting wavelet coefficients and statistical features are extracted for different levels of decomposition and fed into MSVM for classification process. For comparison of the prediction accuracy C4.5 is introduced and applied to bearing fault diagnosis .To reduce the response time of classification process the most significant feature selection methods were introduced, analysed and discussed about the importance of the feature selection method. 

In Section 1 we briefly reviewed the works which are strictly connected to the subject of this paper. In Section 2 contains theoretical background of wavelet transform, statistical features and feature reduction methods were discussed. In Section 3 Experimental setup and experimental procedure is presented. In Section 4 Analysis of simulated data according to the presented procedure is presented. In Section 5 the results for the same is presented and Last section contains conclusions. The methodology of the present work is illustrated in the Fig.1

II. THEORETICAL BACKGROUND OF FEATURE EXTRACTION AND FEATURE SELECTION

A. Wavelet transform

The aim of signal processing is to find the health condition of the rotating component using vibration signal. Recently, wavelet transform (WT), which is capable of providing both the time- domain and frequency-domain information simultaneously, has been successfully used in non-stationary vibration signal processing and fault diagnosis [20]. In this present work wavelet transform is used in feature extraction .The two type of wavelet Continuous wavelet transform(CWT) and Discrete wavelet transform (DWT) are discussed below. A series of oscillating functions with different frequencies such as window function $\Psi_{\alpha,\beta}(t)$ for scanning, $x(t)$ for translating the signal are used in wavelet transform, where $\alpha$ is the oscillating frequency changing parameter and $\beta$ is the translation parameter. A high frequency resolution is obtained when the frequency is low at low time resolution and a low frequency resolution is obtained when the frequency is at low time resolution.

The function with the parameter $\alpha$ and $\beta$ with mother wavelet is given as:

$$\Psi_{\alpha,\beta}(t) = \frac{1}{\sqrt{\alpha}} \Psi \left( \frac{t-\beta}{\alpha} \right)$$

(1)

If all signals $x(t)$ satisfy the condition then,

$$\int_{-\infty}^{\infty} x(t)^2 dt < \infty$$

(2)

that gives $x(t)$ decays to zero.

The wavelet transform, $CWT (\alpha, \beta)$, of a time signal $x(t)$ can be given as

$$CWT (\alpha, \beta) = \frac{1}{\sqrt{\alpha}} \int_{-\infty}^{\infty} x(t) \Psi^{*} \left( \frac{t-\beta}{\alpha} \right) dt$$

(3)

Where an analyzing wavelet is $\Psi(t)$ and $\Psi^{*}(t)$ will be the complex conjugate of $\Psi(t)$. For analyzing the different wavelet, real and complex valued function can be used.A lot of redundant data is created by calculating the wavelet coefficient at possible scale. Thus the analysis can be made sufficient and accurate by giving the limit of choice of $\alpha$ and $\beta$ to discrete numbers.

By selecting the fixed values $\alpha = \alpha_k^j$ and $\beta = k \beta_j \alpha_k^j$, $j, k = 0, \pm 1, \pm 2, \ldots$, we get the DWT

$$DWT (i, j) = \alpha_k^j \int_{-\infty}^{\infty} x(t) \Psi^* \left( \frac{t-k \beta_j}{\alpha_k^j} \right) dt$$

(4)

If $\alpha$ and $\beta$ are replaced by $2^j$ and $2^j k$, then the DWT is derived as

$$DWT (i, j) = \frac{1}{\sqrt{2^j}} \int_{-\infty}^{\infty} x(t) \Psi^* \left( \frac{t-2^j k}{2^j} \right) dt$$

(5)

Thus the high and low frequency signal is obtained from the original signals by decomposition process. The daubechies wavelet transform is used in this process because of its less
process consuming time. As well it is in the family of orthogonal wavelet. This is a type of multi resolution analysis and mostly used for solving a wide range of problems with large dataset. The scaling function of daubechies wavelet is given as db1-db15. It is noted that the highest level of efficiency is taken for the analysis by the Daubechies wavelet.

B. Feature extraction

The bearing fault diagnostic task is actually a problem of pattern classification and pattern recognition, of which the crucial step is feature extraction. In the machinery fault diagnosis field, features are often extracted in time domain and frequency domain. But In this paper, we extract features from wavelet decomposition data. There are many feature parameters are available and the features that taken for this proposed work is listed in Table I.

<table>
<thead>
<tr>
<th>Statistical feature</th>
<th>Equation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kurtosis</td>
<td>$k_{kur} = \frac{1}{n} \sum_{i=1}^{n} (k(t) - \bar{k})^4$</td>
<td>Fourth central moment of $X^4$, divided by fourth power of its standard deviation.</td>
</tr>
<tr>
<td>Skewness</td>
<td>$k_{skew} = \frac{1}{n} \sum_{i=1}^{n} (k(t) - \bar{k})^3$</td>
<td>Third central moment of the value, divided by the cube of its standard deviation.</td>
</tr>
<tr>
<td>Variance</td>
<td>$k_{var} = \frac{1}{n} \sum_{i=1}^{n} (k(t) - \bar{k})^2$</td>
<td>Measures how far a set of numbers is spread out</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>$k_{std} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (k_i - \mu)^2}$</td>
<td>Square root of an unbiased estimator of the variance of the population</td>
</tr>
<tr>
<td>Root mean square</td>
<td>$k_{rms} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} k_i^2}$</td>
<td>Root of sum of squared values</td>
</tr>
</tbody>
</table>

C. Support Vector Machine – Recursive Feature Extraction (SVM-RFE)

SVM-RFE is an iterative process that is used to eliminate redundant features and yield better subsets. SVM-RFE uses weight magnitude as the ranking criterion. This process involves three steps; training the SVM on the given training data set, ranking the features based on their weights, and eliminating the feature with the lowest weights. The inputs given for the SVM-RFE process are the training samples,

$$X = [x_1, x_2 ... x_m ... x_n]^T$$ \hspace{1cm} (6)

and class labels, $Y = [y_1, y_2 ... y_m ... y_n]^T$ \hspace{1cm} (7)

The subset of the persisting features are given as

$$S_b = \{1, 2, ... f\}$$ \hspace{1cm} (8)

From the given training data set, we build the SVM model, and is given as

$$M = SVM_{train}(X, Y)$$ \hspace{1cm} (9)

after training the SVM model the weight vector of dimension length $S_b$ is computed and is given as

$$W = \sum M_{m} x_{m} y_{m}$$ \hspace{1cm} (10)

the ranking criteria is computed for the weighted vectors, the persisting feature ranked list $S_b$ is updated from which the feature with smallest ranking criterion is eliminated. The accuracy of the classifier is evaluated by using 10 fold cross validation. Ranker is used as the search method. The ranker search method uses information gain measure to select the relevant attributes. More details about SVM – RFE can be found in [21].

D. Wrapper Subset Evaluation:

This method calculates the feasible attribute subsets using target learning algorithm. The wrapper method uses Greedy stepwise method as the search method. Survey shows that wrapper method provides better results than filter method because of the interaction between the search and the learning process. But this performance is achieved at the expense of computational cost incurred during invoking the learning algorithm for every feature subset. The greedy stepwise method ranks the attributes either in the stepwise forward or backward selection process. 10 fold cross validation is used to estimate the accuracy of a classifier. Cross validation is repeated as long as the standard deviation is greater than the
mean accuracy or until 10 repetitions are reached. More details about wrapper subset can be found in [22].

E. ReliefF Attribute Evaluation:
ReliefF (RF) method evaluates the worth of an attribute by repeatedly sampling an instance and considering the value of the given attribute for the nearest instance of the same and different class. It can operate on both discrete and continuous class data. The RF is not limited to two class problems and is more robust and can deal with incomplete and noisy data. RF selection process is based on ranking the features with higher weight in order to decrease the chance of selecting features with lower weight. For more information about ReliefF refer [23,24].

F. Principal Component Analysis (PCA)
PCA is a statistical technique that is used to reduce the dimensionality of the data. In order to reduce the number of features and dimensions of the features without affecting the usefulness of the features, feature reduction method is used. Among all the feature reduction techniques PCA is the only unsupervised technique that is it does not bother about the values of the class labels. From the given set of n dimension feature vectors \( x_i \) (i=1, 2, 3, \ldots, m), generally n<m, the mean is calculated as,

\[
\mu = \frac{1}{m} \sum_{i=1}^{m} x_i
\]

then by using the formula (7), the covariance matrix of feature vector is found

\[
C = \frac{1}{m} \sum_{i=1}^{m} (x_i - \mu) (x_i - \mu)^t
\]

by solving the eigen value problem of covariance matrix C, The principal components (PCs) is obtained

\[
C v_i = \lambda_i v_i
\]

Where \( \lambda \) (i=1, 2, \ldots, n) are the eigenvalues and they are sorted in descending order. \( v_j \) (i=1, 2, \ldots, n) are the corresponding eigenvectors. For the raw feature vectors with low-dimensional ones, compute the first k eigenvectors (k \leq n) which correspond to the biggest k eigen values. In order to select the number k, a threshold \( \theta \) is introduced to denote the biggest k eigenvectors,

\[
\frac{v_i^k}{\sum_{i=1}^{n} v_i^k} \geq \theta
\]

Given the precision parameter \( \theta \), the number of eigenvectors k can be decided.

For any given dataset, the eigen vectors are

\[
v = [v_1, v_2, \ldots, v_k]
\]

and eigen values are

\[
A = \text{diagonal}[\lambda_1, \lambda_2, \ldots, \lambda_k]
\]

once the matrix V is determined, the low-dimensional feature vectors, named principle component (PC), of raw ones are obtained as follows,

\[
P = V^T x_t
\]

The Principal Component’s (PCs) of PCA have three properties according to [25]

1) They are uncorrelated; 2) they have sequentially maximum variances; 3) the mean-squared approximation error in the representation of the original vectors by the first several PCs is minimal.

III. EXPERIMENTAL SETUP AND EXPERIMENTAL PROCEDURE
The proposed methodology is verified by performing tests on the designed experimental setup. The experimental setup shown in Fig.2 comprises of components such as variable frequency drive (VFD), three phase 0.5 hp AC motor, bearing, belt drive, gear box and brake drum dynamometer with scale. This experiment utilizes a standard deep groove ball bearing (No. 6005). Measurements of the vibration acceleration signals are captured by a tri-axial type accelerometer that is fixed over the bearing block. A 24 bit data acquisition system was used and the signals were collected at a sampling frequency of 12800 Hz. The bearing was maintained at a constant rotating speed of 1700 r/min. The brake drum dynamometer applied a constant load and the speed was monitored by a tachometer. Fig.2 also depicts normal, outer race, inner race and ball fault (1mm crack depth) conditions were formed using the EDM process. Initially the vibration data is collected for each bearing condition for around 30 seconds in the sampling rate of 12800Hz and it is separated into many data sets for classification purpose. Each sample set contains 6000 data points. In general analysis of the signal is done to find out various conditions of bearing component using time domain and frequency domain by varying amplitudes. Time domain plots are illustrated in Fig.3(a). One of the most basic approach for bearing conditioning monitoring is Frequency analysis.
Fast Fourier transform (FFT), is used to transform the time series data to frequency domain, where the signal is used to deduce the sine and cosine waves from the sample. In practice, analysing those characteristic frequencies such as BPFO- Ball Pass Frequency Outer Race, BPFI- Ball Pass Frequency Inner Race, FTF- Fundamental Train Frequency and BSF- Ball Spin Frequency and measuring the amplitude variations in the characteristic frequency and its side bands as well the harmonics of those frequencies will provide information regarding the health of the bearing. Even though the bearing conditions are difficult to be differentiated by their FFT spectral shown in Fig.3 (b).
IV. FAULT DIAGNOSIS AND RESULT ANALYSIS

The bearing signals are extracted in the sampling rate of 12800 Hz (6400 data points per second). For 20 seconds 128000 data points were collected through accelerometer for each condition of bearing. Then the raw signals are divided into approximately around 20 sample datasets and each sample contain 6000 data points were given as input to the wavelet transform and 6000 and decomposed using daubechies wavelet for 8 level and 15 scale as show in Fig.4(a), Fig.4(b), Fig.4(c)&Fig.4(d) respectively. From the decomposed data the statistical time domain and frequency domain features such as Root mean square(RMS), kurtosis(KUR), skew ness (SKW), variance(VAR), standard deviation(STD) are extracted. These features are used as input MSVM and C4.5. These features are useful in a fault prediction and are further processed and normalized sample feature (STD) values extracted from wavelet coefficients 8 level decomposition in db15 scale for 4 different states of bearing is shown in Table 2. For feature extraction and signal processing wavelet platform is used. Subsequent Feature selection and classification process is implemented using WEKA data mining tool.
Two classifier learners were employed for the classification of the data, Weka incorporated C4.5 (J48) and MSVM (SMO) and is used and for theoretical background of C4.5 refer [25] and For more information on the SMO algorithm, see [26]. The Waikato Environment for Knowledge Analysis (Weka) is a comprehensive suite of Java class libraries that implement many state-of-the-art machine learning and data mining algorithms. Tools are provided for pre-processing data, feeding it into a variety of learning schemes, and analysing the resulting classifiers and their performance [27-28]. The experiment was carried out with and without feature selection schemes. The tool is used to perform benchmark experiment.

As per the diagnosis methodology, features are selected using 4 different feature selection methods SVM–Recursive Feature Elimination , Wrapper subset method, ReliefF method and Principle component analysis (PCA) and It is compared (Table 4) based on the execution or elapsed time period shown in Fig.5. Among the selection methods SVM–Recursive Feature Elimination taken a less execution timing compare to all other methods but when compare to prediction accuracy it is not on the top of the list (shown in Table 5) along with both classifiers MSVM and C4.5 .The prediction results for single (sample) dataset are clearly depicted in the Table 5 , it contains number of features selected for each scheme, each bearing element individual accuracy, average accuracy and mean absolute error. Among all the schemes MSVM-S3 gives the best result of 95% of classification accuracy and less 0.0129 mean absolute error even though with less number of features(14 Nos.)

<table>
<thead>
<tr>
<th>Feature</th>
<th>d1</th>
<th>d2</th>
<th>d3</th>
<th>d4</th>
<th>d5</th>
<th>d6</th>
<th>d7</th>
<th>d8</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMS</td>
<td>F1</td>
<td>F2</td>
<td>F3</td>
<td>F4</td>
<td>F5</td>
<td>F6</td>
<td>F7</td>
<td>F8</td>
</tr>
<tr>
<td>KUR</td>
<td>F9</td>
<td>F10</td>
<td>F11</td>
<td>F12</td>
<td>F13</td>
<td>F14</td>
<td>F15</td>
<td>F16</td>
</tr>
<tr>
<td>SKW</td>
<td>F17</td>
<td>F18</td>
<td>F19</td>
<td>F20</td>
<td>F21</td>
<td>F22</td>
<td>F23</td>
<td>F24</td>
</tr>
<tr>
<td>VAR</td>
<td>F25</td>
<td>F26</td>
<td>F27</td>
<td>F28</td>
<td>F29</td>
<td>F30</td>
<td>F31</td>
<td>F32</td>
</tr>
<tr>
<td>STD</td>
<td>F33</td>
<td>F34</td>
<td>F35</td>
<td>F36</td>
<td>F37</td>
<td>F38</td>
<td>F39</td>
<td>F40</td>
</tr>
</tbody>
</table>

Fig. 4(c). Decomposed signal of ball fault bearing

Fig. 4(d). Decomposed signal of normal bearing
TABLE III
SAMPLE STATISTICAL FEATURES (NORMALIZED) VALUES OF STD EXTRACTED FROM WAVELET COEFFICIENTS 8 LEVEL DECOMPOSITION IN DB15 SCALE FOR 4 DIFFERENT STATES OF BEARING.

<table>
<thead>
<tr>
<th>States</th>
<th>d1</th>
<th>d2</th>
<th>d3</th>
<th>d4</th>
<th>d5</th>
<th>d6</th>
<th>d7</th>
<th>d8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>0.283462</td>
<td>0.053757</td>
<td>0</td>
<td>0.027785</td>
<td>0.002573</td>
<td>0.017349</td>
<td>0.012132</td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>0.673831</td>
<td>0.103484</td>
<td>0.931932</td>
<td>1</td>
<td>0.759407</td>
<td>0.8299</td>
<td>0.92486</td>
<td>0.955869</td>
</tr>
<tr>
<td>Normal</td>
<td>0.801204</td>
<td>0.007884</td>
<td>0.903114</td>
<td>0.956812</td>
<td>0.735277</td>
<td>0.748216</td>
<td>0.937741</td>
<td>0.985658</td>
</tr>
<tr>
<td>Ball fault</td>
<td>0.073152</td>
<td>0.933981</td>
<td>0.083051</td>
<td>0.051354</td>
<td>0.120538</td>
<td>0.005602</td>
<td>0.006983</td>
<td>0.000759</td>
</tr>
<tr>
<td>Ball fault</td>
<td>0.098997</td>
<td>0.926975</td>
<td>0.097925</td>
<td>0.047301</td>
<td>0.103161</td>
<td>0.013037</td>
<td>0.000451</td>
<td>0.000175</td>
</tr>
<tr>
<td>Ball fault</td>
<td>0.390674</td>
<td>0.707986</td>
<td>0.138631</td>
<td>0.041442</td>
<td>0.070923</td>
<td>0.015776</td>
<td>0.001171</td>
<td>0</td>
</tr>
</tbody>
</table>

Bearing fault diagnosis results are also evaluated and compared based on standard parameters such as classification accuracy, sensitivity and specificity for various classifiers are proposed in this work and it is depicted in Table 6. It is interesting to note that an increase in classification accuracy is recorded for the proposed feature reduction methods, with respect to the unreduced data in most of the cases. Also, when comparing classification results, the feature reduction methods have the more or less same classification accuracy values which are recorded in most of the cases.

They are demonstrated in Fig.6 & Fig.7. Test carried out for different data sets and classification accuracy is depicted in Table VII and it shows the MSVM-S3 scheme comprise of wrapper feature selection method and support vector machine classifier with 14 selected features gives the maximum accuracy of 96% overall for all four conditions of bearings with mean absolute error of 0.0137. On an average compare to other schemes MSVM-S3 gives better result.

TABLE IV
ELAPSED TIME FOR DIFFERENT FEATURE SELECTION METHODS IN SECONDS

<table>
<thead>
<tr>
<th>Sl.No</th>
<th>Feature selection scheme</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>No feature selection</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>2</td>
<td>SVM-RFE</td>
<td>0.093</td>
<td>0.082</td>
<td>0.073</td>
<td>0.089</td>
<td>0.095</td>
<td>0.087</td>
<td>0.090</td>
<td>0.079</td>
<td>0.073</td>
<td>0.078</td>
</tr>
<tr>
<td>3</td>
<td>Wrapper subset</td>
<td>1.762</td>
<td>1.521</td>
<td>1.020</td>
<td>1.890</td>
<td>1.678</td>
<td>1.569</td>
<td>1.700</td>
<td>1.432</td>
<td>1.178</td>
<td>1.281</td>
</tr>
<tr>
<td>4</td>
<td>ReliefF</td>
<td>0.385</td>
<td>0.207</td>
<td>0.359</td>
<td>0.289</td>
<td>0.217</td>
<td>0.223</td>
<td>0.323</td>
<td>0.345</td>
<td>0.268</td>
<td>0.290</td>
</tr>
<tr>
<td>5</td>
<td>PCA</td>
<td>2.261</td>
<td>2.011</td>
<td>1.721</td>
<td>2.523</td>
<td>2.610</td>
<td>1.890</td>
<td>2.200</td>
<td>2.412</td>
<td>1.891</td>
<td>1.359</td>
</tr>
</tbody>
</table>
Fig. 5. Different data set Vs Execution timing of feature selection methods

Fig. 6. Classification accuracy of MSVM with different feature reduction methods for different datasets

Fig. 7. Classification accuracy of C4.5 with different feature reduction methods for different datasets
Table V
Prediction results of different schemes of classifiers

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Feature selection method</th>
<th>Scheme name</th>
<th>No. of selected features</th>
<th>Feature description</th>
<th>Prediction accuracy of test data in (%)</th>
<th>Average accuracy on testing data (%)</th>
<th>Mean absolute error</th>
</tr>
</thead>
<tbody>
<tr>
<td>C4.5</td>
<td>SVM-RFE</td>
<td>C4.5-S1</td>
<td>All(40)</td>
<td>F1-F40</td>
<td>80 83 81 80</td>
<td>81.00 0.1082</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>C4.5-S2</td>
<td>18</td>
<td>F1,F7,F13,F18,F20,F21,F23,F24,F25,F27,F28,F29,F31,F34,F35,F37,F38,F40,F1,F2,F4,F5,F10,F13,F17,F21,F25,F28,F30,F33,F36</td>
<td>93 86 88 90 89.25 0.1954</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wrapper</td>
<td>C4.5-S3</td>
<td>14</td>
<td>F1,F2,F5,F7,F8,F11,F13,F15,F16,F17,F18,F21,F24,F27,F28,F29,F31,F34,F36,F37,F1,F2,F4,F5,F8,F9,F10,F13,F17,F21,F25,F28,F30,F33,F36</td>
<td>95 97 94 94 95.00 0.0129</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ReliefF</td>
<td>C4.5-S4</td>
<td>20</td>
<td>F1,F2,F5,F7,F8,F11,F13,F15,F16,F17,F18,F21,F24,F27,F28,F29,F31,F34,F36,F37,F1,F2,F4,F5,F8,F9,F10,F13,F17,F21,F25,F28,F30,F33,F36</td>
<td>95 96 90 87 92.00 0.0157</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PCA</td>
<td>C4.5-S5</td>
<td>11</td>
<td>F1,F2,F5,F7,F8,F11,F13,F15,F16,F17,F18,F21,F24,F27,F28,F29,F30,F33,F36</td>
<td>92 90 93 94 92.25 0.0198</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSVM</td>
<td>No feature selection</td>
<td>MSVM-S1</td>
<td>All(40)</td>
<td>F1-F40</td>
<td>85 87 83 80</td>
<td>83.75 0.2942</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SVM-RFE</td>
<td>MSVM-S2</td>
<td>18</td>
<td>F1,F7,F13,F18,F20,F21,F23,F24,F25,F27,F28,F29,F31,F34,F35,F37,F38,F40</td>
<td>90 89 93 91 90.75 0.1848</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wrapper</td>
<td>MSVM-S3</td>
<td>14</td>
<td>F1,F2,F4,F5,F10,F13,F17,F21,F25,F28,F30,F33,F36</td>
<td>96 97 94 97 96.00 0.0137</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ReliefF</td>
<td>MSVM-S4</td>
<td>20</td>
<td>F1,F2,F5,F7,F8,F11,F13,F15,F16,F17,F18,F21,F24,F27,F28,F29,F31,F34,F36,F37,F1,F2,F4,F5,F8,F9,F10,F13,F17,F21,F25,F28,F30,F33,F36</td>
<td>94 96 91 97 94.50 0.0145</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PCA</td>
<td>MSVM-S5</td>
<td>18</td>
<td>F1,F2,F5,F7,F8,F11,F13,F15,F16,F17,F18,F21,F24,F27,F28,F29,F30,F33,F36</td>
<td>91 93 93 95 93.00 0.0724</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table VI
Results of different prediction parameters for single data set

<table>
<thead>
<tr>
<th>Fault type</th>
<th>MSVM</th>
<th>C4.5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accuracy</td>
<td>Sensitivity</td>
</tr>
<tr>
<td>Normal</td>
<td>90.14193</td>
<td>56.40312</td>
</tr>
<tr>
<td>Normal</td>
<td>93.56443</td>
<td>54.79883</td>
</tr>
<tr>
<td>Normal</td>
<td>90.76421</td>
<td>58.02686</td>
</tr>
<tr>
<td>Inner race fault</td>
<td>85.79575</td>
<td>63.21895</td>
</tr>
<tr>
<td>Inner race fault</td>
<td>91.70734</td>
<td>65.34828</td>
</tr>
<tr>
<td>Inner race fault</td>
<td>88.53764</td>
<td>65.12465</td>
</tr>
<tr>
<td>outer race fault</td>
<td>93.65194</td>
<td>79.07716</td>
</tr>
<tr>
<td>outer race fault</td>
<td>90.23916</td>
<td>78.13403</td>
</tr>
<tr>
<td>outer race fault</td>
<td>93.50609</td>
<td>79.223</td>
</tr>
<tr>
<td>Ball fault</td>
<td>85.76658</td>
<td>74.274</td>
</tr>
<tr>
<td>Ball fault</td>
<td>86.93334</td>
<td>73.47671</td>
</tr>
<tr>
<td>Ball fault</td>
<td>83.14137</td>
<td>72.88361</td>
</tr>
</tbody>
</table>
V. CONCLUSION

In this study, we used ten schemes for automatic fault classification problem. Four bearing conditions states including normal, Inner race fault, outer race fault and ball fault are simulated on experimental set up and the sample data are used to make fault diagnosis test. It is proved by the experiment that MSVM along with Wrapper subset feature selection is a good scheme and it can diagnose bearing faults accurately. As our study revealed that data pre-processing, feature extraction and appropriate feature selection process are important steps in machine learning process since they increase the performance of the classifiers. The WEKA tool is used to feature selection and for classify the data. The classification performance is evaluated by using classification accuracy and mean absolute error. The results of classification algorithms MSVM and C4.5 are discussed along with feature reduction and without feature reduction process.

REFERENCES


