Comparison of Conventional Method of Fault Determination with Data-Driven Approach for Ball Bearings in a Wind Turbine Gearbox

V. N. Balavignesh, B. Gundepudi, G. R. Sabareesh*, and I. Vamsi

Department of Mechanical Engineering, BITS Pilani – Hyderabad campus, Jawahar Nagar, Shameerpet, Rangareddy District, 500078, India

Abstract

The presented investigation on fault diagnosis of ball bearings compares the conventional method using FFT spectra with a data-driven approach using Support Vector Machines (SVMs). Three different cases of bearings (one healthy and two faulty bearings with different crack thickness) were used as experimental cases. The experimentally obtained time-domain acceleration signals were converted to the frequency-domain and de-noised using optimal wavelets selected based on relative magnitudes of Shannon entropy and energy values. The dominant peak was identified for each case and was subsequently compared with the characteristic bearing frequencies evaluated theoretically. The wavelet transformed time-domain experimental data was also used to train the SVM classifier. Also, the effect of statistical tools such as Principal Component Analysis (PCA) and Zero-phase Component Analysis (ZCA) on the classification accuracy of normal SVM and wavelet feature extraction-based SVM have been investigated.

Keywords: Support Vector Machines; fault diagnosis; wavelet transform; wind turbine; principal component analysis

(Submitted on June 24, 2017; Revised on December 26, 2017; Accepted on December 29, 2017)

© 2018 Totem Publisher, Inc. All rights reserved.

1. Introduction

Bearings are vital components in a majority of rotating industrial machinery, such as wind turbine gearboxes, which are subjected to faults frequently. Such ubiquitous components are expected to develop faults over a period of time of their use, leading to machine malfunctioning and downtime. As such, fault diagnosis of rolling element bearings has always been a primary area of interest for researchers. Condition monitoring of bearings is a viable solution to this problem; its ultimate goal being the design of a system that monitors the working condition of machines and also detects the faults developed. Different condition monitoring techniques such as acoustic measurement, electrical effects monitoring, temperature monitoring, wear debris analysis and vibration analysis are used in industries to identify machinery faults [6]. The most common method – the vibration analysis – uses the vibration data of bearings to diagnose faults in bearings. The information from this data is either extracted directly or is used after processing the data using supplemental techniques. Data-driven approaches further use classifiers based on Machine Learning methods such as SVMs and Artificial Neural Networks (ANNs).

Sugumaran et al. focused on the feature selection process for classification using a decision-tree based approach and have used a kernel based neighborhood score Multi-class SVM [23]. Chen et al. proposed a fault diagnosis method based on Dependent Feature Vector and Probability Neural Network (DFV-PNN) for rolling element bearing fault diagnosis [5]. Unal et al. proposed a Genetic Algorithm (GA) optimized neural network for fault diagnosis of roller element bearings, which was proven to be better than the traditional ANN classifier [25]. They have also suggested methods to extract features using envelope analysis accompanied by Hilbert transform and Fast Fourier Transform, which when combined with the GA optimized ANN proved to be efficient. It is obvious that the parameters from various sensors that are mounted will have contrasting early fault detection capabilities. Lin and Makis utilized multivariate vibration data in the fault diagnosis of gears operating under varying load conditions [14]. The authors divided the obtained vibration data into four groups by considering the load variations and vector auto-regression, which are applied to calculate the residuals. These residuals are further

* Corresponding author.
E-mail address: sabareesh@hyderabad.bits-pilani.ac.in
processed to obtain the statistical features. Samantha and Al-Balushi applied the traditional ANN classifier using pre-processing tools such as high-pass, band-pass filtering, envelope detection (demodulation) and wavelet transforms for fault diagnosis of roller element bearings [22]. Samanta also studied the diagnosis of gear condition in a comparative study between the performance of SVMs and ANNs, both without and with automatic selection of features and classifier parameters [21]. Li et al. detected the local gear faults that are presented in the planetary gearbox by de-noising the original signal by means of Empirical Mode Decomposition (EMD) [13]. Jaramillo et al. developed a holistic approach to provide a more complete and definitive diagnostic picture of the health of the different components using a Two-Stage Bayesian Inference approach [8]. Zhang and Zhou performed a multi-fault diagnosis for roller element bearings based on Ensemble Empirical Mode Decomposition (EEMD) and optimized Support Vector Machine (SVM). EEMD adaptively decomposes the vibration signals into Intrinsic Mode Functions (IMFs) [28]. The EEMD energy entropy and singular values of matrix with its rows as IMFs were then used as features. The existence of faults in a bearing has been decided based on the values of the EEMD energy entropy, whereas, the classification of the faulty bearings has been performed by using singular values of matrix as inputs for the multi-class SVM optimized by Inter-Cluster Distance (ICD) in the feature space. Zhang et al. make use of the noise presented in the signal and detected the bearing faults based on stochastic resonance optimized by the Levenberg-Marquardt algorithm [27]. Gryllias and Antoniadis proposed a hybrid two-stage One Vs All Support Vector Machine (SVM) approach for the automated diagnosis of defective roller element bearings [7].

The key aspect of this approach is that the raw experimental vibration signals can directly be used for classification, thus eliminating the need for pre-processing. Sometimes post-processing techniques, like adaptive boosting and weighted voting, can be used to boost the efficiency of the adopted classifier. Zhang et al. proposed a multi-variable ensemble-based incremental Support Vector Machine (SVM) for fault diagnosis of roller bearings [29]. Mba et al. have proposed a novel time-domain feature extraction technique for the fault diagnosis of spacecraft flywheel roller element bearings [15]. The suggested method makes use of the inherent noise that is presented in the signal. Kannatey-Asibu et al. presented the results for a monitoring system based on the concept of classifier fusion following class-weighted voting, which is used to further enhance the system performance [10]. Apart from the data-driven approach to fault diagnosis of rolling element bearings, the conventional approach involving signal processing techniques such as wavelet de-noising and associated mother wavelet selection procedure has also remained popular among many researchers. Regarding the selection of mother wavelet procedure, GA-based approaches and Maximum Energy to Shannon Entropy ratio-based approaches have been developed [20,26]. These mother wavelets are further used for de-noising the experimental vibration signals. Chen et al. developed a customized maximal-overlap multi-wavelet de-noising technique for condition monitoring of a rolling mill driven in which customized multi-wavelet basis function is constructed via symmetric lifting scheme and then vibration signal is processed by maximal-overlap multi-wavelet transform [4]. Many researchers have also worked using a conventional method of fault diagnosis by using signal processing techniques. Peng et al. performed a comparative study of an improved Hilbert-Huang Transform (HHT) against the conventional wavelet transforms and proved it to be more effective than the wavelet transforms [17]. Prabhakar et al. used a discrete wavelet transform for detecting bearing race faults on bearings having both single and multiple point defects on inner and outer races as well as combination faults [18]. Kar and Mohanty developed a fault diagnosis based on motor current signature analysis and wavelet transform to substitute the traditional conventional approach [11]. Abbasion et al. performed a roller element bearing multi-fault classification based on the wavelet de-noising and SVMs [1]. They have demonstrated the capability of wavelet transform as a tool for multi-fault diagnosis of roller bearings. Kankar et al. have studied the fault diagnosis of ball bearings with localized defects using wavelet-based feature extraction [9]. They have used the Minimum Shannon Entropy Criterion (MSEC) for choosing the mother wavelet. Further, they have also compared the classification accuracies of three different classifiers – SVMs, learning vector quantization and self-organizing maps – and concluded SVMs to be the better choice. Mishra et al. have proposed a method to study the variation of speed in extremely slow speed operation by using instantaneous angular position as the base variable for signal processing instead of time [16]. Radhika et al. have worked on the fault diagnostics of induction motor using current signature analysis, with wavelet transform that is treated as a pattern classification problem [19]. They extracted features using wavelet transforms and then used the extracted features to classify the faults using the Support Vector Machine model.

In spite of many works discussed above, there hasn’t been much emphasis on comparison between the conventional and data-driven approaches. Also, there hasn’t been much emphasis on why the data-driven approaches are more efficient than the conventional method for condition monitoring of industrial machinery. The present research work is an attempt in which fault diagnosis of bearings in a wind turbine test rig using both conventional and data-driven approaches. Fast Fourier Transform (FFT) was used to convert the time-domain signal to a representation in the frequency-domain. Wavelet transforms have been used for both de-noising the experimental acceleration signals as well as for feature extraction. The choice of Mother wavelet has been made based on three different procedures, namely Minimum Shannon Entropy Criteria (MSEC), Maximum Energy Criteria (MEC) and Maximum Energy to Shannon Entropy Ratio Criteria (R(s)). Furthermore, a one-Vs-all SVM classifier has been used for classification of the faults. Effects of dimension-reduction techniques such as Principal Component Analysis (PCA) and Zero-phase Component Analysis (ZCA) on the classification accuracy of the SVM classifier have also been
investigated by varying the number of features in the transformed space to reduce the redundancy. This entire work is summarized in Figure 1.

2. Experimental setup

The experimental setup used in the present investigation is a bench-top wind turbine test rig developed at BITS Pilani, Hyderabad Campus as shown in Figure 2. Details regarding the experimental setup have been given in Biswal et al. and Biswal and Sabareesh [2,3]. The test rig includes a 3-stage spur gear train that has an overall speed ratio of 48:1. A 1hp, 3-phase AC motor replaces the generator to perform the experiment and the speed of the motor is regulated by programmed variable frequency drive.

<table>
<thead>
<tr>
<th>S.No</th>
<th>Nature/Condition of the bearing</th>
<th>Dimensions of the fault</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Healthy</td>
<td>NA</td>
</tr>
<tr>
<td>2</td>
<td>Faulty (Outer race)</td>
<td>0.25 mm</td>
</tr>
<tr>
<td>3</td>
<td>Faulty (Outer race)</td>
<td>0.50 mm</td>
</tr>
</tbody>
</table>
Two different fault dimensions with only the outer race fault were used in the present study, as specified in Table 1 and shown in Figure 3. The bearings were housed in an experimental wind turbine gearbox and acceleration data was recorded using a piezo-electric accelerometer. Experiments were performed for two different speeds of rotation – 800 rpm and 1000rpm. The deep groove ball bearing used in the present study is SKF SY 25 FM. The salient characteristics of the bearing are as follows.

- Bore diameter = 25 mm
- Pitch diameter (D) = 31.102 mm
- Rolling element diameter (d) = 7.938 mm
- Contact angle (φ) = 0 degrees
- Number of rolling elements (n) = 9

The data was acquired at a sampling rate of 1 kHz and the sampling length was 10000 for each set of reading. In turn, 100 data sets were taken for each case of bearing considered in the study. For the SVM model, the training set comprised of 80% of the sample sets and the remaining were used for testing the data.

3. Conventional approach to fault diagnosis

The conventional approach to fault diagnosis (with additional wavelet based de-noising techniques) that was used in the present study is explained in the following steps:

- Acceleration signals obtained were transformed to its frequency-domain for the various cases of healthy and faulty bearings.
- Using mother wavelets experimental signals were de-noised for conventional fault diagnosis. The selection of mother wavelet was done using Minimum Shannon Entropy Criteria (MSEC), Maximum Energy Criteria (MEC) and Maximum Energy to Shannon Entropy Ratio Criteria (R(s)). Kankar et al. have used MSEC for wavelet selection [9]. However, Yan [26] suggested the use of R(s) criteria to be the best for wavelet classification as R(s) is the ratio of energy and Shannon entropy, and thus takes into account the effects of both MSEC as well as MEC. Three different criteria were used to ensure that the best wavelet was chosen for the current investigation. Values of the parameters, namely, Shannon entropy, energy and the energy to Shannon entropy ratio have been evaluated for the acceleration data corresponding to the three cases – namely healthy bearing condition, outer race fault of crack thickness 0.25 mm and outer race fault of crack thickness 0.50 mm. These parameters have been evaluated for a set of wavelets for each of the faults corresponding to different RPMs, and the choice of mother wavelet has been made.
- Common defect frequencies of the bearing were determined using theoretical equations. These peaks were identified from the de-noised plots.

For de-noising, the interval-dependent de-noising procedure was employed. The input experimental signal was decomposed into a number of intervals based on variance change points corresponding to the first level detail coefficients, and each interval was subjected to a threshold separately. Thresholds were obtained using a minimax threshold rule and soft thresholding was used to modify the wavelet coefficients. Minimax threshold rule is based on the minimax decision making principle by which, when presented with two various and conflicting strategies, the strategy that will minimize the maximum losses that could occur is chosen.
3.1. Characteristic Bearing Frequencies

The characteristic bearing defect frequencies for a bearing are its Fundamental Train Frequency (FTF), Ball Passing Frequency of Outer race (BPFO), Ball Passing Frequency of Inner race (BPFI), Ball Spin Frequency (BSF), and rolling element defect frequency. These frequencies are dependent on the dimensions of the bearing and can be evaluated using theoretical equations.

-**Fundamental Train Frequency (FTF)**, often called cage frequency, is the rotation rate of the cage supporting the rollers in the bearing and is given by the below formula.

\[
FTF = \frac{N}{2} \left[ 1 - \left( \frac{d}{D} \right) \cos \varphi \right]
\]

- The rollers/balls passing the outer race will generate the ball pass frequency on the outer race and is given by the below formula.

\[
BPFO = \frac{n}{2} \cdot \frac{N}{60} \left[ 1 - \left( \frac{d}{D} \right) \cos \varphi \right]
\]

where \(N\) – Speed of the shaft (rpm)

Since the present investigation is concerned with healthy and two outer race faults, only FTF and BPFO have been evaluated, which are listed in Table 2. In order to perform fault diagnosis using Conventional method, these frequencies were compared with the frequency domain plots of the de-noised experimental signals. Presence of these frequencies in the spectrum helps to identify, characterize and label the faults. The calculated frequencies are listed in Table 2.

<table>
<thead>
<tr>
<th>Characteristic Frequency</th>
<th>Case – 1 (800 rpm)</th>
<th>Case – 2 (1000 rpm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FTF</td>
<td>4.965 Hz</td>
<td>6.207 Hz</td>
</tr>
<tr>
<td>BPFO</td>
<td>44.685 Hz</td>
<td>55.858 Hz</td>
</tr>
</tbody>
</table>

3.2. Selection of mother wavelet

In the conventional method, bearing characteristic frequencies are used for identifying the faults. Whereas in the data-driven approach, the initial step is to select the mother wavelet. The selection of mother wavelet was carried out using Minimum Shannon Entropy Criterion (MSEC), Maximum Energy Criterion (MEC) and Maximum Energy to Shannon Entropy Ratio (R(s)), as discussed in the earlier section. Shannon entropy of wavelet coefficients is given by:

\[
S_{\text{entropy}}(n) = - \sum_{i=1}^{m} p_i \log p_i
\]

where \(p_i\) is the energy probability distribution of the wavelet coefficients, defined as:

\[
p_i = \frac{|C_{ni}|^2}{E(n)}
\]

where \(\sum p_i = 1\), and in the case \(p_i = 0\), \(p_i \log p_i\) is taken as zero. \(C_{ni}\) is the \(i^{th}\) detail coefficient and \(E(n)\) corresponds to the total energy associated with all the wavelet coefficients [9]. Maximum Shannon Entropy value signifies that the signal is random and has a uniform distribution. In general, faulty signals are least uniform and least random, thereby having Minimum Shannon Entropy value.

Since the energy content of a signal is a measure that uniquely characterizes a signal, it can be used for wavelet selection. In general, healthy signals have random distributions, whereas the faulty signals have transient impulses, which enhance the energy content compared to the healthy signal.

The amount of energy contained in a signal can be calculated from its wavelet coefficients and is given by:

\[
E_{\text{energy}} = \iint |w_t(s,T)|^2 dsdT
\]

Energy to Shannon entropy ratio is defined as:

\[
R(s) = \frac{E_{\text{energy}}}{S_{\text{entropy}}}
\]
Wavelets from various families, namely, daubechies (db), coiflets (coif), symlets (sym), discrete meyer wavelet (dmey) and bi-orthogonal (bior) wavelets were chosen in the present study. The values of these three parameters, i.e. Shannon entropy, energy and energy to Shannon entropy ratio, have been evaluated for these wavelets at various levels of decomposition – 3, 5 and 7 levels [1,9,18,20,26]. The trends of all three parameters are shown in the Figures 4, 5, 6, 7, 8, 9, 10, 11 and 12.

- From the Minimum Shannon Entropy Criterion plots, i.e., from Figures 4, 5 and 6, it can be observed that as levels of decomposition increased, the Shannon entropy values decreased for each wavelet. Also, after 7 levels of decomposition, the Symlet 8 (sym8) wavelet has the minimum Shannon entropy value among the other wavelets such as coif1, sym3 etc. Hence, according to MSEC, it is best suited for the data set.

- From the Maximum Energy Criterion plots, i.e., from Figures 7, 8and 9, it can be observed that increasing levels of decomposition are accompanied with an increase in the energy value. Employing the maximum energy criteria results in the selection of the daubechies 02 (db02) and symlet 2 (sym2) wavelets after 7 levels of decomposition.

- From the Maximum Energy to Shannon Entropy Ratio plots, i.e., from Figures 10, 11and 12, it can be observed that after 7 levels of decomposition the energy to Shannon entropy ratio values clearly indicates that the daubechies 02 (db02) and symlet 2 (sym2) wavelets have the maximum ratio criteria. This confirms that either wavelet is best suited for the analysis.
Comparison of Conventional Method of Fault Determination with Data-Driven Approach for Ball Bearings in a Wind Turbine Gearbox

Figure 7. Energy after 3 Levels of decomposition

Figure 8. Energy after 5 Levels of decomposition

Figure 9. Energy after 7 Levels of decomposition

Figure 10: Energy to Shannon entropy ratio after 3 Levels of decomposition
The time-domain acceleration data was pre-processed using wavelet feature extraction (and also dimension reduction tools) for classification using SVM. This was done using all the three wavelets, namely, symlet 2 (sym2), daubechies 02 (db02) and symlet 8 (sym8). It was observed that, even though daubechies 02 and symlet 2 wavelets produce good classification accuracy values, daubechies 02 (db02) wavelet performs better in comparison. Therefore, db02 wavelet has been selected for de-noising the experimental time-domain vibration signals after 7 levels of decomposition.

3.3. De-noising and Conventional Approach to Diagnosis

The experimental signal was de-noised using db02 wavelet. The plots used for this identification of the bearing frequencies from the frequency spectrum corresponding to the three cases (one healthy and two outer race faults) are shown in Figures 13, 14, 15, 16, 17 and 18.

- For a healthy bearing operating at 800 rpm, a sharp peak at a frequency of 6.049 Hz that has an acceleration value of 169.5 m/s$^2$ could be observed, see Figure 13. The observed frequency is close to the theoretical frequency (4.965 Hz), see Table 2. For a bearing having 0.25 mm outer race fault operating at 800 rpm, another peak is observed at a frequency of 36.9 Hz, which is closer to the calculated theoretical frequency, see Figure 14. The acceleration value corresponding to this peak is 82.4 m/s$^2$. A similar kind of peak at a frequency of 37.61 Hz that has an acceleration value of 139.5 m/s$^2$ could be noticed in the case of 0.5 mm outer race fault, see Figure 15.

- For a healthy bearing operating at 1000 rpm, a peak at a frequency of 5.34 Hz was observed, which is close to the theoretical FFT (6.207 Hz). The acceleration value at this frequency is found to be 3815 m/s$^2$, see Figure 16. However, for a bearing that has a 0.25 mm outer race fault operating at 1000 rpm, another peak at a frequency of 51.62 Hz with an acceleration value of 101.5 m/s$^2$ was observed. Additionally, two sub-harmonics at frequencies 103.2 Hz and 154.9 Hz that have accelerations of 32.57 m/s$^2$ and 25.55 m/s$^2$ were observed, see Figure 17. For a bearing with 0.5 mm outer race fault, a peak was noticed at a frequency of 51.62 Hz, and the acceleration value corresponding to that peak is 210.4 m/s$^2$, which is almost double the corresponding value in a 0.25 mm outer race fault, see Figure 18. In addition, two sub-harmonics at frequencies 103.2 Hz and 154.9 Hz have accelerations of 101.1 m/s$^2$ and 44.62 m/s$^2$.

- From the above results, it can be interpreted that the peak in the FFT spectra of a bearing operating at 1000 rpm has weakened considerably from the healthy case as the outer race fault progresses. This weakening in the acceleration values can be attributed to the distribution of the energy across a number of frequencies, i.e. sub-harmonics in the fault case.
Another observation that is worth noting is that the acceleration values of the fault frequencies have increased as the fault severity increases, i.e. the characteristic peaks of 0.5 mm outer race fault are stronger than the 0.25 mm outer race fault. It is worth noting that Biswal et al. have estimated the crack thickness values using the acceleration data [2]. Table 3 elaborates the dependency between the acceleration values and the crack thickness observed in the present investigation [2]. Another observation that can be noted from the following table is the fact that the acceleration values corresponding to BPFO at 1000 rpm show higher values than the BPFO at 800 rpm for both values of crack thickness. However, the sub-harmonics show smaller acceleration values, which is as expected. The difference in actual and theoretical values may also be attributed to the slipping or skidding in the rolling element bearings [24].

Table 3. Variation of acceleration with crack thickness

<table>
<thead>
<tr>
<th>Crack thickness (mm)</th>
<th>BPFO at 800 rpm (m/s²)</th>
<th>BPFO at 1000 rpm (m/s²)</th>
<th>1st harmonic at 1000 rpm (m/s²)</th>
<th>2nd harmonic at 1000 rpm (m/s²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>82.40</td>
<td>101.50</td>
<td>32.57</td>
<td>25.56</td>
</tr>
<tr>
<td>0.50</td>
<td>139.50</td>
<td>210.00</td>
<td>101.10</td>
<td>44.62</td>
</tr>
</tbody>
</table>

Figure 13. FFT plot for Healthy Bearing, 800rpm

Figure 14. FFT plot for 0.25 mm outer race fault, 800rpm

Figure 15. FFT plot for 0.50 mm outer race fault, 800rpm
4. Data driven approach to fault diagnosis

Recently, data driven approaches combined with machine learning and artificial intelligence have been used in the field of fault diagnosis. Support Vector Machine (SVM) is a supervised learning model with associated learning algorithms, which is used to analyze data for classification and regression problems. It is used in the present study to develop an efficient fault diagnosis system to classify the various faults from a healthy system.

4.1. Dimensionality reduction

Dimensionality reduction is the process through which the number of random variables under consideration can be reduced by obtaining a set of uncorrelated principal variables (between variables like sum and mean of a data set or the range, minimum and maximum value of a data set). This is important because the incremental contribution to the classification accuracy by addition of each feature is dependent on the correlation between the already existing features and the new feature. Application
of a dimensionality reduction technique removes this correlation while simultaneously ensuring that similar information is stored in new features, which are some combinations of the existing features. Furthermore, this also evaluates the relative contribution of each new feature to the classification accuracy. The working of dimensionality reduction can be explained as: if \( X \) is the original matrix \( D \times N \) containing the initial training set data (with initial features), and \( W \) is the \( N \times N \) matrix whose columns are the vectors of \( X^T X \), then the full feature-decomposition of \( X \) is given by

\[
T = WX
\]

The solution for \( W \) must satisfy

\[
WTW = \langle XX^T \rangle^{-1}
\]

In the present investigation, two Dimensionality reduction techniques were used, namely, Principal Component Analysis (PCA) and Zero-phase Component Analysis (ZCA).

4.1.1. Principal Component Analysis (PCA)

PCA is a statistical technique that uses an orthogonal transformation to convert a set of possibly correlated features into a set of linearly uncorrelated values called Principal Components (PCs). As a result of this transformation, the first principal component has the largest possible variance, i.e. it accounts for as much of the variability in the data as possible. Each succeeding principal component has the highest possible variance under the constraint that it is orthogonal to and is uncorrelated with the preceding components. This also ensures that the variance in the data not captured by the preceding principal components is captured by the subsequent principal component. Also, the principal components are orthogonal because they are the eigen vectors of the covariance matrix, which is symmetric. The solution obtained through PCA must satisfy Equation 8. Since PCA evaluates the orthogonal (global) solution \([WW^T] = I\), the PCA solution \(W_P\) is given by

\[
W_P = D^{-0.5}E^T
\]

Where \( D \) is the diagonal matrix of eigen values and \( E \) is the matrix whose columns are the eigen vectors. In the present investigation, PCA is carried out using a Singular Value Decomposition (SVD) algorithm. The method is completely elucidated in Lee et al. [24].

4.1.2. Zero-phase Component Analysis (ZCA)

The aim of using Zero-phase component analysis is to whiten/sphering the input signal, i.e. apply linear transformation to decorrelate the signal features. In mathematical terms, assume that a signal is represented by a matrix \( X \) belonging to the set of real numbers with dimensions \( D \times N \), where \( N \) is the number of input features. Then, the objective of ZCA is to find a whitening letter \( W \) belonging to real numbers with dimensions \( N \times N \), such that the resulting matrix \( WX \) is an identity sample co-variance matrix. Also, the dimensions of resulting matrix \( WX \) is \( D \times N \). Also, an optional regularization parameter is input, which is the weight given to the identity while computing the co-variance matrix. A value of \( 1 \times 10^{-06} \) has been used for the regularization parameter in this study. Also, the present investigation assumes that the data have a sample mean of zero.

The solution obtained through ZCA must satisfy Equation (8). Since ZCA evaluates the Symmetrical (local) solution \([WW^T] = W^2\), the ZCA solution \(W_Z\) is given by

\[
W_Z = \langle XX^T \rangle^{-0.5}
\]

ZCA is the polar opposite of PCA. It produces local whitening letters, which are ordered according to the phase spectrum of the signal [12].

4.2. Classes and Features

Classification using SVMs has been carried out in both 800 rpm and 1000 rpm cases. Various time-domain and frequency-domain features are extracted from the selected wavelet (db02). The following list shows the various time-domain and frequency-domain parameters used for classification. \( TFi \) where \( i \in [1, 14] \) indicate the time-domain functions and \( FFi \) where \( i \in [1, 4] \) indicate the frequency-domain functions.
The statistical features listed in Table 4 and Table 5 have been implemented in the SVM model developed for classification. All SVM classifiers have been used in the present investigation. Furthermore, in the 800 rpm case, wavelet feature extraction using the three optimal wavelets, namely, db02, sym2 and sym8 as well as dimension reduction techniques such as PCA and ZCA have been implemented to boost the classification accuracy. However, in the case of 1000 rpm, classification accuracies using a normal SVM yielded prediction accuracies very close to 100%; therefore, wavelet feature extraction and dimension reduction techniques were not used in this case. Tables 4 and 5 show the classification accuracies obtained in the case of 800 rpm and 1000 rpm using different combinations of classification features.

From Tables 4 and 5, it can be inferred that in both 800 rpm and 1000 rpm cases, use of both time-domain and frequency-domain features yields close to 100% classification accuracy. However, use of only time-domain features yields mixed results. In the case of 800 rpm, use of time-domain features alone yields a classification accuracy of 93.33%, whereas in the 1000 rpm case, it yields an accuracy close to 100%. Further, use of wavelet feature extraction and dimension reduction tools for boosting classification accuracy in cases corresponding to 800 rpm show that feature extraction with sym2 and db02 wavelets yields good results, with db02 performing better than sym2. However, sym8 shows poorer classification results in comparison with sym2 and db02 wavelets. This is a direct reflection of the R(s) values, which shows identical values for sym2 and db02, but a comparatively lower value for sym8. The difference in performance of sym2 and db02 wavelets could be attributed to the suitability of these particular wavelets to the spectra obtained in the present investigation.

Another important observation to be noted is that in both cases, namely, pre-processed wavelet based SVM analysis and pre-processed normal SVM analysis, data reduction techniques such as PCA and ZCA were applied and the relative contributions (percentages) of each of the principal components to the overall signal were determined. The criteria were set

\[ T_F = \frac{1}{N} \sum_{n=1}^{N} x(n) \]

\[ T_F = \frac{1}{N} \sum_{n=1}^{N} (x(n) - TF_1) \]

\[ T_F = \text{kurtosis}(x(n)) \]

\[ T_F = \max(x(n)) \]

\[ T_F = \min(x(n)) \]

\[ T_F = \text{range}(x(n)) \]

\[ T_F = \text{skewness}(x(n)) \]

\[ T_F = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} x(n)^2 \]

\[ T_F = \frac{1}{\max(x(n))} \sum_{n=1}^{N} |x(n)| \]

\[ T_F = \frac{1}{\max(x(n))} \sum_{n=1}^{N} |x(n)|^2 \]

\[ T_F = \frac{1}{\max(x(n))} \sum_{n=1}^{N} |x(n)|^3 \]

\[ T_F = \frac{1}{\max(x(n))} \sum_{n=1}^{N} |x(n)|^4 \]

\[ F_F = \frac{1}{N} \sum_{k=1}^{N} |s(k)| \]

\[ F_F = \frac{1}{K-1} \sum_{k=1}^{K} |s(k) - FF_1| \]

\[ F_F = \frac{1}{K(FF_1)} \sum_{k=1}^{K} |s(k) - FF_1|^2 \]

\[ F_F = \frac{1}{K(FF_1)^{1.5}} \sum_{k=1}^{K} |s(k) - FF_1|^3 \]
such that the PCs that had greater than 0.01 % contribution were only considered significant. Based on this, it was found that 12 PCs in the case of pre-processed wavelet based SVM analysis and 8 PCs in the case of pre-processed normal SVM analysis were significant. This means that with the application of wavelet transform to pre-process the training data, more PCs were found to contribute significantly to the overall signal, which implies that wavelet transform pre-processing had a positive effect on the efficiency of the classification.

Table 4. Classification accuracies for 800 rpm case (T-D – Time-Domain & F-D – Frequency-Domain)

<table>
<thead>
<tr>
<th>Features used for classification</th>
<th>Additional Pre-processing tools used</th>
<th>Classification accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T-D + F-D features (Normal SVM)</td>
<td>NA</td>
<td>100.00</td>
</tr>
<tr>
<td>T-D features (Normal SVM)</td>
<td>NA</td>
<td>90.83</td>
</tr>
<tr>
<td>T-D features (Normal SVM)</td>
<td>PCA (all 14 PCs)</td>
<td>82.50</td>
</tr>
<tr>
<td>T-D features (Normal SVM)</td>
<td>PCA (8 PCs)</td>
<td>83.34</td>
</tr>
<tr>
<td>T-D features (Normal SVM)</td>
<td>ZCA</td>
<td>91.11</td>
</tr>
<tr>
<td>T-D features (Wavelet transformed SVM)</td>
<td>sym2 wavelet</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NA</td>
<td>90.00</td>
</tr>
<tr>
<td></td>
<td>PCA (all 14 PCs)</td>
<td>90.00</td>
</tr>
<tr>
<td></td>
<td>PCA (12 PCs)</td>
<td>88.34</td>
</tr>
<tr>
<td></td>
<td>ZCA</td>
<td>91.67</td>
</tr>
<tr>
<td>T-D features (Wavelet transformed SVM)</td>
<td>db02 wavelet</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NA</td>
<td>88.33</td>
</tr>
<tr>
<td></td>
<td>PCA (all 14 PCs)</td>
<td>93.34</td>
</tr>
<tr>
<td></td>
<td>PCA (12 PCs)</td>
<td>94.44</td>
</tr>
<tr>
<td></td>
<td>ZCA</td>
<td>95.00</td>
</tr>
<tr>
<td>T-D features (Wavelet transformed SVM)</td>
<td>sym8 wavelet</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NA</td>
<td>73.81</td>
</tr>
<tr>
<td></td>
<td>PCA (all 14 PCs)</td>
<td>75.56</td>
</tr>
<tr>
<td></td>
<td>PCA (12 PCs)</td>
<td>74.45</td>
</tr>
<tr>
<td></td>
<td>ZCA</td>
<td>78.33</td>
</tr>
</tbody>
</table>

Table 5. Classification accuracies for 1000 rpm case (T-D – Time-Domain & F-D – Frequency-Domain)

<table>
<thead>
<tr>
<th>Features used</th>
<th>Pre-processing tools used</th>
<th>Classification accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T-D + F-D features (Normal SVM)</td>
<td>NA</td>
<td>100.00</td>
</tr>
<tr>
<td>T-D features (Normal SVM)</td>
<td>NA</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Figure 19. Comparison of conventional approach and data driven approach to fault diagnosis
It is also worth noting that dimension reduction techniques such as PCA and ZCA seem to perform better when wavelet feature extraction is used, i.e. the difference between the accuracy obtained without using dimension reduction techniques as compared to the accuracy obtained using dimension reduction techniques is more when wavelet feature extraction is used. This observation can be clearly inferred from the cases corresponding to wavelet feature extraction using sym2 and db02 wavelets, where the classification accuracies in almost all of the cases are greater than 90%. To be specific, dimension reduction on the initial data set without the use of wavelet feature extraction yields poorer results than the normal SVM on time-domain data. Hence, to conclude, feature extraction using db02 wavelet in combination with dimensionality reduction techniques such as PCA and ZCA yielded better results compared to the other cases. Also, ZCA yields comparatively better results than PCA in all cases, both with and without using wavelet feature extraction. To make the comparison prudent, the conventional error is evaluated by obtaining frequencies corresponding to a significant peak closest to the theoretical frequency the peak represents. The comparison is shown in Figure 19. The bar-graph in Figure 20 illustrates the error percentage values when various data driven cases are compared.

Figure 20. Bar graph illustrating the error percentage values in data-driven approach to fault diagnosis

5. Conclusions

In the present investigation, a comprehensive method for fault diagnosis of rolling element bearings in a wind turbine test rig using a comparison between conventional approach and data-driven approach has been presented. As far as conventional approach is concerned, mother wavelet selection based on three different criteria, namely, Minimum Shannon Entropy Criteria (MSEC), Maximum Energy Criteria and Maximum Energy to Shannon Entropy ratio (R(s)) was performed. It was observed that the wavelet db02 at 7 levels of decomposition is best suited for the data set, which can be inferred from both the wavelet’s relative magnitudes of energy and Shannon entropy values and the wavelet-transformed SVM classification results. As far as the dependence of acceleration values corresponding to characteristic defect frequencies on the crack thickness is concerned, it was observed that:

- For a given peak, the corresponding acceleration value increased relatively from 0.25mm crack thickness to 0.50mm crack thickness.
- Higher acceleration values were shown by the peak corresponding to BPFO for 1000 rpm in comparison with the peak corresponding to 800 rpm for both crack thickness values.

In addition to this, an important contribution of the present work has been to discuss the effects of dimension-reduction techniques such as PCA and ZCA along with wavelet-based feature extraction in boosting the performance of the SVM classifier. It was observed that:

- The effect of dimension-reduction techniques in boosting the classification accuracy was much more pronounced when the input time-signal data was pre-processed using wavelet transforms.
- ZCA whitening proved to be better than PCA whitening in boosting the classification accuracy both with and without wavelet transforms.

The conventional and data-driven approaches to fault diagnosis were compared by comparing the error obtained while classifying the fault. The classification accuracy obtained by data-driven approach is as mentioned in Table 4 and Table 5. Through the present investigation, the relative advantage of the data-driven method compared to the conventional method of fault diagnosis can be estimated.
Acknowledgements

This study was funded by ADCRG, BITS Pilani, and Hyderabad Campus. The authors are grateful to the institutions for their support.

References


