Optimal Feature Set for Detection of Inner Race Defect in Rolling Element Bearings

Karthik Kappaganthu [1]**, C. Nataraj** [1]**, Biswanath Samanta** [1]

[1] *Department of Mechanical Engineering, Villanova University, Villanova, PA, 19333, USA*
karthik.kappaganthu@villanova.edu
nataraj@villanova.edu
biswanath.samanta@villanova.edu

ABSTRACT

Rolling element bearings are the key components in many rotating machinery. It is necessary to determine the condition of the bearing with reasonable degree of confidence. Many techniques have been developed for bearing fault detection. Each of these techniques have their own strengths and weaknesses. In this paper various features are compared for detecting inner race defects in rolling element bearings. Mutual information between the feature and defect is used as a quantitative measure of quality and the features are ranked appropriately. Often, a combination of different features is used for bearing fault detection. Hence it is important to understand the interaction of features for classification purposes. This paper addresses this issue and determines the optimal feature set for best detection performance.

1 INTRODUCTION

Bearings are the key load carrying members in a rotating machinery. It is necessary to determine the state of the bearing with reasonable degree of confidence to prevent catastrophic failure. Bearing failure has been studied extensively and numerous methods have been developed for fault detection. Bearing failure can take place because of wear and improper installation. One of the common modes of failure in a rolling element bearing is a point defect on the inner race or the outer race of the bearing. Of these the inner race defect is more difficult to detect.

The techniques developed involve measuring vibration signals and processing them using signal processing techniques to obtain the features. Based on the techniques used the features can be classified into time (Tandon, 1994), frequency (Barkova and Barkov, 1995; Randall and Gao, 1994; Ypma, 2001) and time-frequency (Cade *et al.*, 2005; Mori *et al.*, 1996; Ypma, 2001) domains.

One of the first feature extraction techniques for rolling element bearing fault detection were the time domain techniques. Rolling element bearings with faults showed higher peak to peak vibration compared to a healthy bearing (Tandon, 1994; Barkova and Barkov, 1995). Some of the common time domain features considered are skewness and kurtosis.

Frequency domain methods are among the most used feature extraction techniques for bearing fault detection. When the rolling element enters a defect an impulse acts on the casing. The impulse is exerted at a frequency with which the rolling elements enter the defect. This frequency can be calculated from the geometry of the bearing and rotating speed (Harris, 2002; Nataraj and Pietrusko, 2005; Harsha *et al.*, 2004). The frequency domain techniques use this excitation to detect the defects in the bearing. The frequency component associated with the inner race defect is called the inner race ball pass frequency. The rotation of cage also produces some frequency components. The other frequency components present in a typical bearing signal are the 1X response, its harmonics and sub-harmonics. The presence of harmonics and sub-harmonics indicates nonlinear behavior in general.

Fast Fourier Transform is the most common method to extract the frequency components in a signal. However, in most cases the inner race defect's spectrum does not contain a peak corresponding to the ball pass frequency. This is because the excitation signals need to travel through the rolling element, casing and then be detected at the sensor by which time the signal is masked by other excitations. However, in the spectrum the frequency components spaced at the bearing defect frequency can be found around the bearing resonance frequency. Further, the impulse's excitations are amplitude modulated and can be recognized as side bands. Envelope Spectrum can be the used to obtain this frequency information (Randall and Gao, 1994). Feature extraction in the frequency domain is still a subject of much fascinating research. Novel techniques for optimal filtering, weak signal detection and demodulation are still being developed. (Randall and Sawalhi, 2009; Sawalhi and Randall, 2008; Ho and Randall, 2000; Ypma, 2001) are some interesting publications in this area.

Discrete wavelet transforms (DWT) is a method for obtaining the time-frequency information of the signal. These are useful to extract the transients in the signal

and are hence popular for inner race defect detection. The divide the signal into various levels based on the frequency range in the signal. The energies in these levels are used as a features. More information about the wavelet transforms can be found in (Chan, 1995). Some of the recent work on bearing diagnostics using DWT are (Ocak *et al.*, 2007; Cade *et al.*, 2005; Pan *et al.*, 2009; Djebala *et al.*, 2008; Wu and Liu, 2008; Feng and Schlindwein, 2009; Mori *et al.*, 1996).

Bearing defect detection can be formulated as a classification problem. Classification involves two steps. First extracting features from the measured signal (usually vibration from accelerometers) and then training a classifier such as ANN (Artificial Neural Network) and ANFIS (Adaptive Neuro-Fuzzy Interference System). These trained classifiers can then be used to classify the new data.

The performance of the classifier depends on the quality of data and the quality of features. In this research the quality of features is analyzed. The interaction between the features is also important for classification performance. Too few or too many features would degrade the classifier's performance. Also since some the features are better at lower speeds and some at higher speeds a combination of features is necessary. The optimal feature set must be able to provide superior classification at varying speeds. Some general computational intelligence based algorithms that can be used for feature selection are (Sugumaran *et al.*, 2007; Peng *et al.*, 2005; Guo *et al.*, 2004; Malhi and Gao, 2004; Raymer *et al.*, 2000).

Although there are many studies on optimal feature selection there are only a few that analyze bearing defects quantitatively in particular (Sugumaran *et al.*, 2007). The aim of this study is to develop an optimal feature set based on strong mathematical foundations that can be sufficiently generalized with less human expertise. Some general computational intelligence based algorithms that can be used for feature selection are (Guo *et al.*, 2004; Malhi and Gao, 2004; Raymer *et al.*, 2000).

In this research, information theoretic approach is used to quantify the quality of the features. These techniques measure the quality of features as the mutual-information content between features and the state of the bearing (faulty or healthy). Mutual information is a statistical measure that correlates different random variables (Duda *et al.*, 2001). It can be calculated from the probability distribution between the random variables (Cover and Thomas, 1991).

The advantage to using information theoretic approach is that it is independent of the classifier used. Also, among the various features used, some of them might have similar information among them. Hence using such features together increases the uncertainty and degrades the performance. Information theoretic approach addresses this important issue of interaction of features with each other for classification purposes.

Using mutual information, this paper address three important issues. First, it illustrates a quantitative statistical method to compare features for detection of faults in bearings, second it provides an optimal feature set comprising of features from different domain that together provide good classification accuracy and third it provides a guideline for the features that need
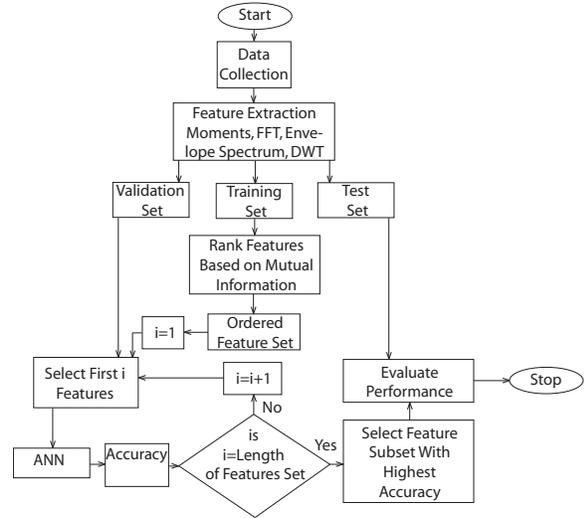

Figure 1: Algorithm for feature selection

to be considered for fault monitoring purposes. In this paper time (skewness, kurtosis), frequency (FFT, envelope) and time-frequency (discrete wavelet transform) domain features are compared with each other.

The methodology is explained in the next section. Section three explains the details of feature ranking using mutual information and a simple algorithm for feature subset selection. The details of the experimental setup and data collection are explained in the fourth section. In the fifth section the typical features obtained from a bearing and the results obtained form the algorithm are shown. The final section deals with conclusions and future work to be done.

2 METHODOLOGY

The flowchart of the process is provided in the Fig. 1.

Data collection is the first step in the proposed method. Vibration data is collected from a system with a faulty bearing and a defect-free bearing over a span of rotating speeds and used for training, validation and testing of the algorithm. The faulty bearing has a localized inner race defect.

Moments, Fast Fourier Transform, Envelope Transform and Discrete Wavelet Transforms are used to obtain the relevant features. Skewness, Fast Fourier Transform (FFT), Envelope magnitudes at Ball Pass Frequency (BPF), Cage Frequency (CF), 1/2X, 1X, 2X, and Discrete Wavelet Transform (DWT) energies and skewness up to level six are used as features. The data is divided into a training set, a validation set and a test set. Care is taken that data in each set is distributed evenly over the entire operating range.

Next, a greedy search algorithm is used to rank the features based on the mutual information content. Greedy search algorithm is a popular sequential search technique used in statistical research (Duda *et al.*, 2001).

Now the validation set is used to extract an optimal feature subset for classification using an ANN as the classifier. The feature subset selection is performed incrementally using the ordered feature set obtained in

the previous stage. The subset with the best ANN classification performance is the optimal solution. This optimal feature subset is then used to test the performance using the test set data.

3 FEATURE RANKING AND SELECTION

3.1 Feature Ranking

As explained earlier the feature ranking is based on mutual information. Let x_i be the random variable with pdf $p(x_i)$ corresponding to the ith feature. Let C be any classifier that maps the features into N_C classes and c_k the corresponding random variable with pdf $p(c_k)$, $k = 1,\ ,2,\ ...,\ N_C$. Note that c_k is a discrete random variable. The entropy and mutual information are the defined as in Eqs. 1 and 2.

$$H(x_i) = -\int p(x_i) \log p(x_i) dx \qquad (1)$$

$$I(x_i; c_k) = -\int p(x_i, c_k) \log \frac{p(x_i, c_k)}{p(x_i)p(c_k)} dx \qquad (2)$$

Further, the entropy and mutual information are related by Eq. 3.

$$I(x_i; c_k) = H(c_k) - H(c_k|x_k) \qquad (3)$$

In order to calculate the mutual information we need to find $p(c_k)$ and $p(c_k|x_i)$ from the data. It is easy to find $p(c_k)$ as it is a discrete random variable. By Bayesian rule we have

$$p(c_k|x_i) = \frac{p(x_i|c_k)p(c_k)}{p(x_i)} \qquad (4)$$

The pdf of a continuous random variable x can be calculated from a given data using a Parzen's Window.

$$p(x) = \frac{1}{N} \sum_{i=1}^{N} \phi(x - x_i, h) \qquad (5)$$

where, N is the number of samples, h is a parameter that defines the size of the window, x_i are the data points and ϕ is a finite valued non-negative density function called the window function. In this work a Gaussian function is used for ϕ (as is done typically).

Using Eqs. 5 and 6, $p(x_i|c_k)$ can be calculated.

$$p(x_i|c_k) = \frac{1}{N_k} \sum_{i=1}^{N_k} \phi(x - x_{k_i}, h) \qquad (6)$$

where, N_k are the number of data points in the kth class and x_{k_i} are the data points belonging to k^{th} class. Using Eqs. 4, 5 and 6 mutual information between a feature and a class can be calculated using Eq. 7.

$$I(x_i, c) = \sum_{k=1}^{N_C} p(c_k) \log p(c_k) - \int \sum_{k=1}^{N_C} p(x_i|c_k)p(c_k) \log p(x_i|c_k) dx \qquad (7)$$

However, in order to calculate the mutual information between a set of features, $x = [x_1\ x_2\ ,...,\ x_n]$ and a class, we would need to calculate the joint pdf $p(x)$ of the feature set and the conditional joint pdf $p(x|c)$.

Although it is possible to do this, it is cumbersome and often inaccurate. A simpler procedure is to use Eq. 8.

$$I(x; c) = \frac{1}{|S|} \sum_{x_i \epsilon S} I(x_i; c) - \frac{1}{|S - 1|^2} \sum_{x_i, x_j \epsilon S} I(x_i; x_j)$$

$$x = \{x : x \epsilon S \subset X\} \qquad (8)$$

The first part of the right hand side of Eq. 8 is the mean of the mutual information of each of the features and class; it is a measure of relevance of the set S. The second part consists of the information between the features themselves; it is a measure of redundancy of the set S. Using this method it is necessary to only calculate the joint pdf of two features at a time. This method, when used in a sequential search, has similar performance to the actual value (Peng *et al.*, 2005).

The feature selection process using these measures is an optimization problem and can be formally defined as in Eq. 9.

$$max\ I(x; c),\ x = \{x : x \epsilon S \subset X\} \qquad (9)$$

where, X is the set containing the features and S is some subset of it. If the size of S is equal to size of X then the solution to Eq. 9 will be an ordered set of features.

Equation 9 can be solved using the greedy search technique. In the first step of this technique, set S is initialized to an empty set and a feature pool set defined as F is initialized to X. Next, S is populated iteratively with a feature from the feature pool such that it maximizes $I(x; c)$ at each stage. The selected feature is then removed from the feature pool. This process is continued till the feature pool is empty.

The algorithm for ranking can be summarized as follows.

1. From the data find $p(c_k)$ and $H(c_k)$, $k = 1, 2, 3, ..., N_C$.

2. Set $S = \{\}, F = X$.

3. While F is not an empty set, DO

    (a) Set $i = 1$, Start Loop 1

    (b) Append the ith element of F to S, i.e. $S_i = \{S,\ F_i\}$.

    (c) Set $j = 1$, Start Loop 2

    (d) Using Eq. 7 find $I(x_j, c)$.

    (e) Using Eq. 5 find $I(x_i, x_j)$.

    (f) If reached the end of S_i End Loop 2, else increment $j \rightarrow j + 1$ and go to Step d.

    (g) Estimate mutual information of set S_i, $I(S_i, c)$ using Eq.8.

    (h) If reached the end of F End Loop 1, else increment $i \rightarrow i + 1$ and go to Step b.

    (i) Find the element x_i^* corresponding to Maximum $I(S_i, c)$.

    (j) Append x_i^* to S and remove it from F.

4. END WHILE

5. The final set S is the ordered feature set.

3.2 Feature Selection

The aim of this stage is to extract an optimal subset S_{opt} from the ordered feature S set obtained in the previous stage. The criterion for optimization is to achieve the least classification error using as few features as possible. The validation data is used to train an ANN and the classification accuracy is the measure of the classification. The algorithm for this is as follows.

1. Initialize $i = 1$ and $S_i = S(1)$.

2. Start Loop

3. Train an ANN using S_i and evaluate classification accuracy a_i.

4. If $S_i = S$ Stop Loop and proceed to step 6, else continue.

5. Increment $i \rightarrow i + 1$ and $S_i \rightarrow \{S_i, S(i+1)\}$ and proceed to step 3.

6. From a_i find i^* corresponding to maximum accuracy and optimal set size.

7. Obtain S_{opt} as S_{i^*}.

4 EXPERIMENTAL SETUP

All the experimental data was collected on a 'Machine Fault Simulator (MFS)' (Spectra Quest, 2009) Fig. 2. It is a test rig with a rotating shaft on a two ball bearings. The shaft and the motor are connected using a flexible coupling to minimized misalignment effects. The shaft is loaded using a bearing loader and balancing disks. The different parts of the system can be conveniently assembled and disassembled. The bearings are placed in the bearing casing and can easily be replaced. The bearing parameters for the system used are given in Table 1. The bearing defect was three mils deep and four degrees wide. The system was loaded with a 5 kg mass. The signals from the MFS were collected using accelerometers placed on the bearing casing; once with a defect-free bearing and once with a bearing with an inner race defect. The signals were captured at a sampling rate of 25 kHz.

The rotating speed was varied between 120 rpm and 3360 rpm with increments of 120 rpm. At each rotating speed, 10 sets of data were collected. Five of these were used in training set, two in validation set, and three in test set. There were 280 samples in all; 140 of these were used for training, 56 for validation and 84 for testing. In each of these sets half of the samples were from a defect-free system and the other half from a system with an inner race defect.

Parameter	Value
Number of Rolling Elements (N_b)	8
Pitch Diameter (D_m)	1.319 in
Rolling Element Diameter (D_b)	.3125 in
Ball Pass Frequency (ω_{bpfi})	4.93 Ω

Table 1: Bearing Parameters

5 RESULTS

5.1 Features

The data collected from the experimental setup agree with the typical observations in similar systems. The
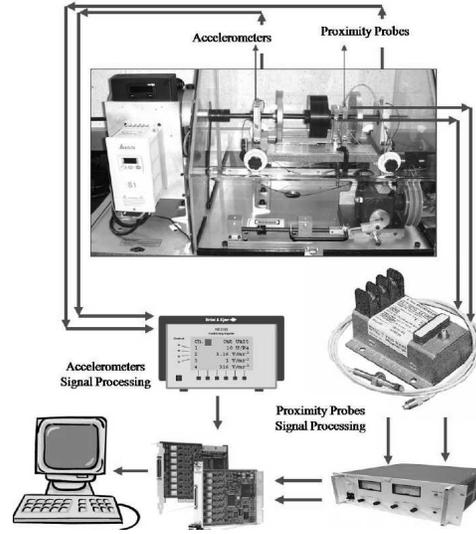

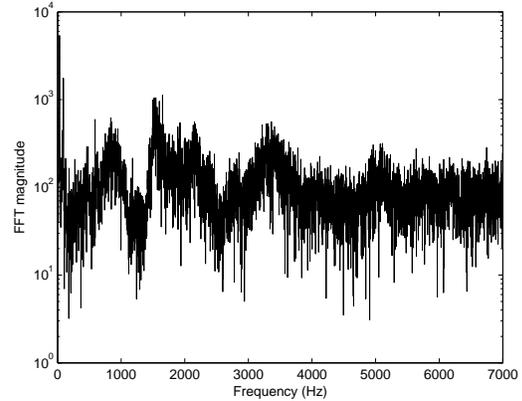Figure 2: Experimental setup

Figure 3: FFT of the measured signal

FFT of the signal for a bearing with defect is shown in Fig. 3. The bearing casing resonance is observed at 3.5 kHz. To evaluate the envelope spectrum the signal is bandpass filtered about this frequency.

Typical spectra of a defect-free and a faulty bearing at lower frequency range are shown in Figs. 4 and 5 respectively. The rotating speed was 30 Hz and the ball pass frequency which can be calculated using Eq. 10 was about 148 Hz. In Eq. 10 N_b is the number of rolling elements, D_b is the diameter of the rolling element, D_m is the mean diameter of bearing and Ω is the rotating speed. It can be seen that there is no peak in the spectrum at the inner race ball pass frequency. However, in the envelope spectrum clear peaks are visible at ball pass frequency and its harmonics.

$$\Omega_{bpfi} = N_b(1 + D_b/D_m)\Omega/2 \qquad (10)$$

The reconstructed DWT detail signal from a defect-free bearing and a bearing with inner race defect are shown in Figs. 6, and 7 respectively. For brevity signals only between level 3 and level 5 are shown, it can
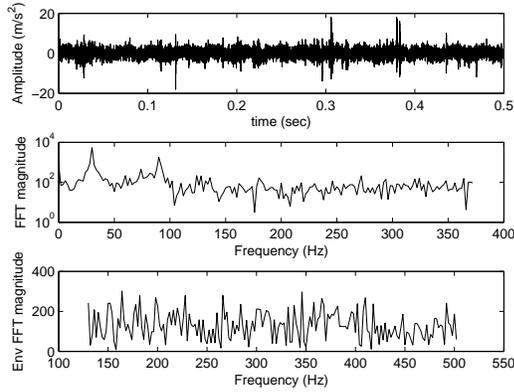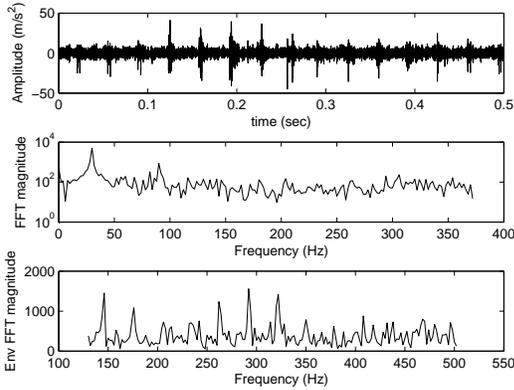
Figure 4: Feature extraction for a defect-free bearing

Figure 5: Feature extraction for a bearing with an inner race defect

Figure 6: DWT of a defect-free bearing signal

be seen that the signal from a bearing with inner race defect has more energy at lower orders. It is important to note that the features are strong nonlinear functions of rotating speeds. The rotating speed contains useful information and therefore needs to be a part of the feature set.

5.2 Mutual Information Based Feature Selection

Twenty eight features were input to the mutual information based ranking algorithm. The final ordered feature set obtained using the feature ranking algorithm explained in section 3 is given in Table 2. The initial order of the feature set input to the algorithm is arbitrary. It can be seen that DWT based features ended up higher in the table. Also, the rotating speed is second in the list and clearly contains important information. DWT energies at levels four and three have more information than other features. The envelope magnitude at ball pass frequency is next feature in the list. This is an important observation, it suggests that DWT based features are better than envelope spectrum based features in this data. Hence DWT based features should be used for studies like fault monitoring and prognostics.
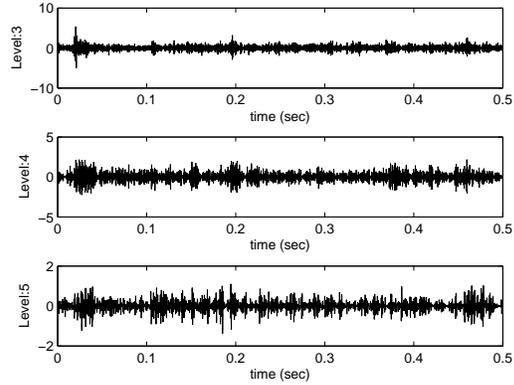

Figure 7: DWT of a bearing signal with an inner race defect
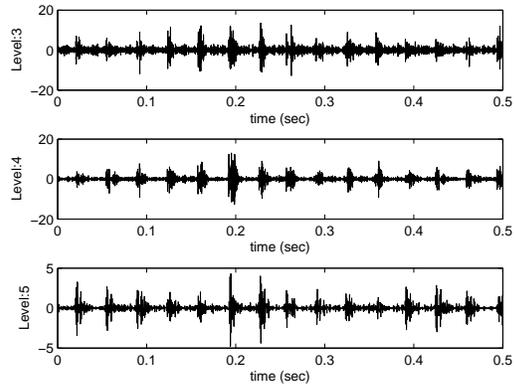
Figure 8: Mutual information

	Ordered Features
1	DWT Detail Energy at Level 4
2	Rotating Frequency (Ω)
3	DWT Detail Energy at Level 3
4	Env Mag at ω_{bpfi}
5	Env Mag at $\Omega/2$
6	FFT Mag at $\Omega/2$
7	Env Mag at ω_{cage}
8	Env Mag at 2Ω
9	FFT Mag at ω_{cage}
10	FFT Mag at 2Ω
11	FFT Mag at ω_{bpfi}
12	DWT Detail Energy at Level 1
13	DWT Skewness at Level 4
14	DWT Detail Energy at Level 2
15	DWT Skewness at Level 3
16	DWT Approx Energy at Level 6
17	DWT Approx Energy at Level 2
18	DWT Skewness at Level 5
19	DWT Approx Energy at Level 5
20	DWT Approx Energy at Level 3
21	DWT Detail Energy at Level 5
22	DWT Approx Energy at Level 4
23	DWT Approx Energy at Level 1
24	Skewness
25	DWT Skewness at Level 6
26	DWT Skewness at Level 2
27	DWT Detail Energy at Level 6
28	DWT Skewness at Level 1

Table 2: Features

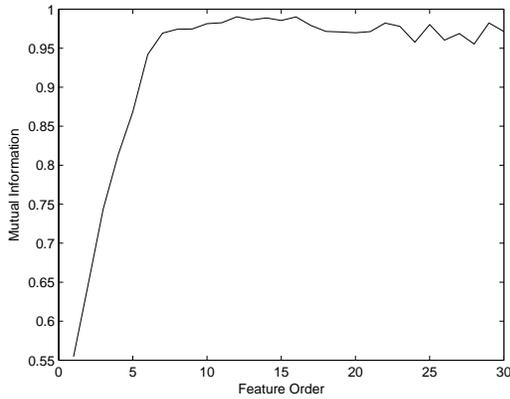|	DF	IRD	
DF	97.8%	2.2%	Validation set
	97.2 %	2.8%	Test set
IRD	3.6%	96.4%	Validation set
	4.2%	95.8%	Test set

Table 3: Confidence Matrix

Another observation is that envelope and FFT magnitude at 1/2X , 2X are also important features. It is a measure of the nonlinearity in the system. The general observation in many studies is that these frequency components are dominant in a bearing with a defect. This has been verified in our experiments and their importance has been validated by this procedure. As had been suggested by many previous studies (Randall, 1987; Cade *et al.*, 2005; Pan *et al.*, 2009) time domain, FFT, DWT approximation signal based features are not effective for ball bearing fault detection. This has been corroborated by this procedure, they rank low in the table. The mutual information calculated is shown in Fig. 8. The mutual information estimate has a steep increase till the first six features then peaks at about the twelfth feature which is DWT detail energy at level 1. Features up to eleven or thirteen would probably give similar performance but for the present data we use the first twelve features as a reasonable mean value. The FFT magnitude at ω_{bpfi} ended up at eleventh place and has added very little relevant information to feature set. It should be noted that even though DWT energy at third and fourth level have higher information content they alone are not sufficient for accurate classification.

Using the ordered feature set and the algorithm explained earlier for feature selection, an ANN was used for classification of the validation data. The mean accuracy of ten ANN runs on the validation set was the criterion for feature selection.

The ANN had one hidden layer with five neurons. Tan-Sigmoid activation function was used and a termination criterion was the error being less than 10^{-10} units or thousand iterations. The mean performance of the ANN per feature is shown in Fig. 9. The accuracy of classification shows a trend similar to mutual information.

The optimum feature set consists of the first twelve features. When this feature set was used with the validation data, the algorithm correctly classified a defect-free system with an accuracy of 97.8 %, a system with a faulty bearing with an accuracy of 96.4 % and overall accuracy of the algorithm was 97.1 %. The corresponding performance on the test set was 97.2%, 95.8% and 96.5% respectively. The variance was 2.56% for the defect-free case and 4.32% for the bearing with an inner race defect. The confidence table for validation set and test set is shown in Table 3. These indicate excellent performance.

6 CONCLUSION

Mutual information was used to rank and compare features for detecting inner race defects in rolling element bearings. Time, frequency and time-frequency domain features were extracted and ranked. DWT detail energy at level four was ranked the highest. However, this feature alone was not capable of good classification performance. As expected the envelope magnitude at the ball pass frequency ranked higher than the FFT magnitude at the same frequency. The DWT approximation based energies and skewness features degraded the information content in the feature set. An ANN was used to determine the optimal feature subset from the ranked feature set. The ANN classification accuracy showed similar trend to the mutual information content. Another interesting observation is the presence of FFT and envelope magnitudes at $\Omega/2$
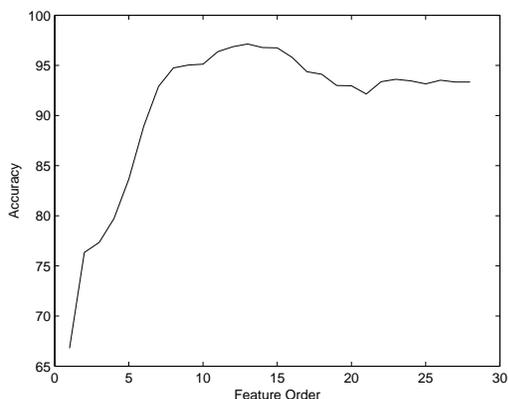
Figure 9: Classification performance on validation data

and 2Ω. These features are a measure of the nonlinearity in the system, which means that the bearing with an inner race defect excites higher nonlinear behavior of the system. It would be interesting to perform nonlinear analysis on a bearing model with a defect; this will be pursued in future work.

Using a validation data set an optimal feature subset was extracted. The optimality condition was mean minimum classification error using ten ANN simulations. The optimal feature set in this case had twelve features. The method showed excellent performance on a test data set. The other advantage of this method is that this feature set guarantees high performance over a range of rotating speeds. Further work is in progress to compare the outcomes of the proposed technique with some other benchmark techniques and to incorporate model based features and evaluate their advantages and disadvantages for the purpose of diagnostics and prognostics. A study on the relationship between defect magnitude and various features using mutual information is also being pursued.

ACKNOWLEDGEMENTS

REFERENCES

(Barkova and Barkov, 1995) A. Barkova and N. Barkov. Condition assesment and life prediction of rolling element bearings. *Sound and Vibration*, 1995.

(Cade *et al.*, 2005) Iain S. Cade, Patrick S. Keogh, and M. Necip Sahinkaya. Fault identification in rotor/ magnetic bearing systems using discrete time wavelet coefficients. *IEEE/ ASME Transctions on Mechatronics*, 10(6):648–657, December 2005 2005.

(Chan, 1995) Y. T. Chan. *Wavelet Basics*. Kluwer Academic Publihsers, Boston, 1995.

(Cover and Thomas, 1991) T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley and Sons, New York, 1991.

(Djebala *et al.*, 2008) Abderrazek Djebala, Douredine Ouelaa, and Nace Hamzaoui. Detection of rolling bearing defects using discrete wavelet analysis. *Mechanica*, 43:339–348, 2008.

(Duda *et al.*, 2001) Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification*. Wiley-Interscience, New York, 2001.

(Feng and Schlindwein, 2009) Yanhui Feng and Fernando S. Schlindwein. Normalized wavelet packets quantifiers for condition monitoring. *Mechanical Systems and Signal Processing*, 23:712–723, 2009.

(Guo *et al.*, 2004) H. Guo, L. B. Jack, and A. K. Nandi. Automatic feature extraction for bearing fault detection using genetic programming. *Eighth International Conference on Vibrations in Rotating Machinery - IMechE Conference Transactions*, 2:363–372, 2004.

(Harris, 2002) Tedric A. Harris. *Rolling Bearing Analysis*. Wiley-Interscience, 2002.

(Harsha *et al.*, 2004) S. P. Harsha, K. Sandeep, and R. Prakash. Nonlinear behaviors of rolling element bearings due to surface waviness. *Journal of Sound and Vibration*, 272:557–580, 2004.

(Ho and Randall, 2000) D. Ho and R. B. Randall. Optimisation of bearing diagnostic techniques using simulated and actual bearing fault signals. *Mechanical Systems and Signal Processing*, 14:763–788, 2000.

(Malhi and Gao, 2004) Arnaz Malhi and Robert X. Gao. PCA-based feature selection scheme for machine defect classification. *IEEE Transcations on Instrumentation and Measurement*, 53:1517–1525, 2004.

(Mori *et al.*, 1996) K. Mori, N. Kasashmi, T. Yoshioka, and Y. Ueno. Prediction of spalling on ball bearings by applying discrete wavelet transform to vibration signals. *Wear*, 8:195–162, 1996.

(Nataraj and Pietrusko, 2005) C. Nataraj and Robert Gerad Pietrusko. Dynamic response of rigid rotors supported on rolling element bearings with an outer raceway defect. In ASME, editor, *Proceedings of IDETC/CIE 2005*, Long Beach California, USA, September 2005.

(Ocak *et al.*, 2007) Hasan Ocak, Kenneth A. Loparo, and Fred M. Discenzo. Online tracking of bearing wear using wavelet packet decomposition and probabilistic modeling. *Journal of Sound and Vibration*, 302:951–961, 2007.

(Pan *et al.*, 2009) Yuna Pan, Jin Chen, and Lei Guo. Robust bearing performance degradation assessment method based on improved wavelet packet - support vector data descriptions. *Mechanical Systems and Signal Processing*, 23:669–681, 2009.

(Peng *et al.*, 2005) Hanchuan Peng, Fuhui Long, and Chris Ding. Feature selection based on mutual

information : Criteria of max-dependency , max-relevance and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27:1226–1238, 2005.

(Randall and Gao, 1994) R. B. Randall and Y. Gao. Extraction of modal parameters from the response of power cepstrum. *Journal of Sound and Vibration*, 176:179–193, 1994.

(Randall and Sawalhi, 2009) R.B. Randall and N. Sawalhi. Signal processing tools for tracking the size of a spall in a rolling element bearing. In *IUTAM Symposium on Emerging Trends in Rotor Dynamics*, 2009.

(Randall, 1987) R. B. Randall. *Frequency Analysis*. Bruel & Kjaer, Naerum, Denmark, 1987.

(Raymer *et al.*, 2000) Michael L. Raymer, William F. Punch, Erik D. Goodman, Leslie A. Kuhn, and Anil K. Jain. Dimensionality reduction using genetic algorithms. *IEEE Transactions on Evolutionary Computation*, 4:164–171, 2000.

(Sawalhi and Randall, 2008) N. Sawalhi and R. B. Randall. Simulating gear and bearing interactions in the presence of faults part1. the combined gear bearing model and simulation of localized bearing faults. *Mechanical Systems and Signal Processing*, 22:1924–1951, 2008.

(Spectra Quest, 2009) Inc Spectra Quest. Http: www.spectraquest.com, 2009.

(Sugumaran *et al.*, 2007) V. Sugumaran, V. Muralidharan, and K. I. Ramachandran. Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing. *Mechanical Systems and Signal Processing*, 21:930–942, 2007.

(Tandon, 1994) N. Tandon. A comparision of some vibration parameters for condition monitoring of rolling element bearings. *Measurement*, 12:285–286, 1994.

(Wu and Liu, 2008) Jian-Da Wu and Chiu-Hong Liu. Investigation of engine fault diagnosis using discrete wavelet transform and neural network. *Expert Systems and Applications*, 35:1200–1213, 2008.

(Ypma, 2001) A. Ypma. *Learning Methods of Machine Vibration Analysis and Health Monitoring*. PhD thesis, Delft University, 2001.

Karthik Kappaganthu received his Bachelors degree in Mechanical Engineering from Kakatiya University, India, and his M.S. degree in Mechanical Engineering from Villanova University, in 2004 and 2006, respectively. He is currently pursuing Ph.D in Engineering at Villanova University.

C. Nataraj received his B.S. degree in Mechanical Engineering from Indian Institute of Technology, and his M.S. and Ph.D. degrees from Arizona State University. He is currently a Professor and the Chair of the department of Mechanical Engineering at Villanova University. He is the founder and former Director of the Center for Nonlinear Dynamics & Control, an interdisciplinary research center. His research interests are in modeling, dynamics and control of nonlinear systems.

B. Samanta received B.Tech. (Honors) and Ph.D. in Mechanical Engineering from Indian Institute of Technology (IIT), Kharagpur. He is currently in the Department of Mechanical Engineering at Villanova University, Villanova, PA. Prior to joining Villanova, he held academic positions at Sultan Qaboos University, Oman and IIT, Kharagpur. His major research interests include broad areas of system dynamics and control, advanced signal processing, prognostics and health management, and applications of computational intelligence in engineering and biomedicine. He has over ninety refereed research articles published by professional bodies like ASME, IMechE, AIAA and IEEE. He is a member of ASME, ASNE, a senior member of IEEE, IEEE Computational Intelligence Society (CIS), and IEEE Engineering in Medicine and Biology Society (EMBS).