

Inchoate Fault Detection Framework: Adaptive Selection of Wavelet Nodes and Cumulant Orders

M. F. Yaqub, *Student Member, IEEE*, Iqbal Gondal, *Member, IEEE*, and Joarder Kamruzzaman, *Member, IEEE*

Abstract—Inchoate fault detection for machine health monitoring (MHM) demands high level of fault classification accuracy under poor signal-to-noise ratio (SNR) which persists in most industrial environment. Vibration signals are extensively used in signature matching for abnormality detection and diagnosis. In order to guarantee improved performance under poor SNR, feature extraction based on statistical parameters which are immune to Gaussian noise becomes inevitable. This paper proposes a novel framework for adaptive feature extraction based on higher order cumulants (HOCs) and wavelet transform (WT) (AFHCW) for MHM. Features extracted based on HOCs have the tendency to mitigate the impact of Gaussian noise. WT provides better time and frequency domain analysis for the nonstationary signals such as vibration in which spectral contents vary with respect to time. In AFHCW, stationary WT is used to ensure linear processing on the vibration data prior to feature extraction, and it helps in mitigating the impact of poor SNR. K-nearest neighbor classifier is used to categorize the type of the fault. Simulation studies show that the proposed scheme outperforms the existing techniques in terms of classification accuracy under poor SNR.

Index Terms—Abnormal vibration detection, higher order cumulants (HOCs), machine health monitoring (MHM), stationary wavelet transform (WT) (SWT).

I. INTRODUCTION

MACHINE health monitoring (MHM) is crucial in all industrial processes to achieve high reliability, reduced man power, and scheduled maintenance. MHM specifically deals with abnormality detection and diagnosis. To diagnose the abnormality, it is important to record certain physical parameters which vary according to the variation in the operation of the machine, and “vibration” is one of such parameters. Vibration pattern changes according to the variations in the machine dynamics. In case of rotary machinery, malfunctioning in the operation of the bearing is the most common fault. It has been investigated that 40% of the total machine faults are because of the bearing [1]. These bearing faults change machine dynamics and generate certain vibrations patterns. In this paper, three types of bearing faults are studied, i.e., inner-race, outer-race, and ball faults. Vibrations caused by each fault result in certain characteristic frequency which is dependent upon speed of rotation and geometry of the bearing. The vibration

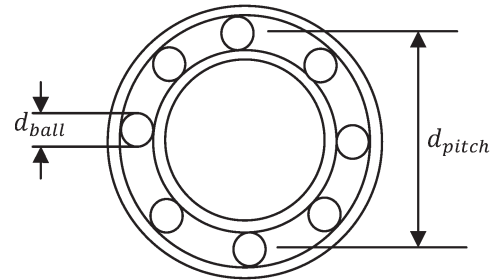


Fig. 1. Rolling element ball bearing geometry.

characteristic frequencies [2] for inner-race (f_{ID}), outer-race (f_{OD}), and ball (f_{BD}) defects can be represented by

$$f_{ID} = \frac{n}{2} f_{rm} \left(1 + \frac{d_{ball}}{d_{pitch}} \cos \phi \right) \quad (1)$$

$$f_{OD} = \frac{n}{2} f_{rm} \left(1 - \frac{d_{ball}}{d_{pitch}} \cos \phi \right) \quad (2)$$

$$f_{BD} = \frac{d_{pitch}}{2d_{ball}} f_{rm} \left(1 - \left(\frac{d_{ball}}{d_{pitch}} \right)^2 \cos^2 \phi \right) \quad (3)$$

where f_{rm} , d_{pitch} , d_{ball} , n , and ϕ represent the frequency of rotation, pitch diameter, ball diameter, number of balls, and the contact angle, respectively, as highlighted in Fig. 1.

In MHM and fault diagnosis, one of the prevalent issues is the residual life estimation and fault severity estimation as contributed by the authors [3]. It facilitates the maintenance staff to schedule the repair by optimizing demand–supply relationships instead of sudden breakdown in the equipment. In order to intimate the maintenance staff about the variation in residual life of the equipment, it is vital to diagnose the type of the fault at the inception stage. In the field of fault diagnosis, a lot of techniques are proposed in the literature to diagnose the type of the fault, also proposed by the authors [4], [5]. The performance of these techniques deteriorates under inchoate and early-stage fault detection due to the following two challenges.

- 1) Reliability under poor signal-to-noise ratio (SNR): Vibration signatures for MHM are weak in the early stage of faults, and the resultant SNR is very poor. An efficient system should diagnose the fault reliably under initial conditions.
- 2) Immunity to noise: In harsh industrial conditions, SNR may vary drastically over the time in industrial environment. It is crucial for the MHM system to be adaptive to the changes in SNR.

Manuscript received October 26, 2010; revised September 18, 2011; accepted September 19, 2011. Date of publication November 22, 2011; date of current version February 8, 2012. The Associate Editor coordinating the review process for this paper was Dr. John Sheppard.

The authors are with Monash University, Gippsland, Vic., Australia (e-mail: farrukh.yaqub@gmail.com).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIM.2011.2172112

In order to categorize the abnormality precisely and reliably, it is crucial to devise signal processing techniques and statistical tools to ensure improved performance under poor SNR. This work focuses on the performance enhancement of MHM system under poor SNR.

Vibration signal is nonstationary in nature, i.e., its spectral contents vary with respect to time. In order to extract detailed and robust information about the vibration signal, it is crucial to investigate the frequency contents of the signal. For nonstationary signals, Fourier transform (FT) and short-time FT (STFT) do not perform well [6]. To combat with nonstationary signal, wavelet transform (WT) is preferably used as compared to conventional FT. In WT, the incoming vibration signal is decomposed into multiple frequency subbands called “nodes” (see Section II). It provides multiresolution analysis in time and frequency domains on contrary to STFT providing only fixed time and frequency resolution both for low and high frequencies [6]–[8].

Wavelet decomposition results in a large number of coefficients which depend upon the number of decomposition levels. First- and second-order statistics are extensively used as feature extraction after wavelet decomposition such as computing energies [9], means and variances [10], standard deviation [11], root-mean-square value [12]–[15], autocorrelation [16], and envelope [17] for the frequency subbands. However, feature extraction schemes based on first- and second-order statistics [9]–[19] suffer from severe performance degradation under poor SNR. The first contribution of AFHCW is to extract the features based on higher order statistics which provide immunity to Gaussian noise [20], [21].

Only a few of the techniques use higher order statistics in order to extract the features after WT [20], [21]. These techniques make use of either discrete WT (DWT) or wavelet packet transform (WPT) to perform time and frequency domain analysis for the vibration signal. Both DWT and WPT involve down-sampling [6] which is a nonlinear process. The down-sampling of the vibration data prior to feature extraction based on higher order statistics may change the data distribution. The second contribution of AFHCW is to recognize this change in the data distribution and provide remedy by using stationary WT (SWT) in which entire processing is essentially linear which guarantees performance enhancement in terms of fault detection accuracy.

The feature extraction schemes in the literature based on higher order statistics [20], [21] extract features using third- and fourth-order statistics, i.e., skewness and kurtosis, respectively. The third contribution of AFHCW is to investigate the fault detection capabilities for the extracted features based on statistical parameters with orders higher than “4” and record their promising characteristics. Moreover, all the existing techniques [9]–[16], [20], [21] lack in developing a generalized criterion with analytical validation to determine the dominance levels for the extracted features for different decomposition nodes after wavelet transformation. The fourth contribution of the AFHCW is to incorporate the adaptive feature selection approach not only in deciding the statistical order but also in deciding the prominent decomposition node using SWT. The adaptiveness is based on fault signature dynamics, i.e., selecting the optimal

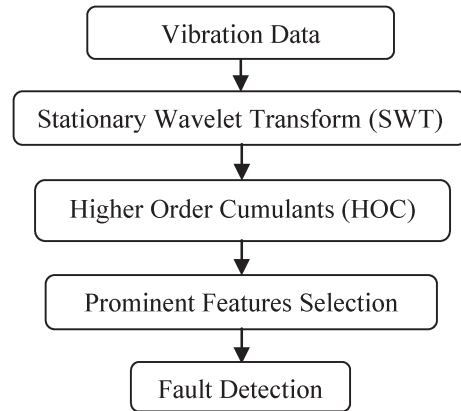


Fig. 2. Framework for MHM.

statistical order and decomposition node based on vibration data itself.

In order to authenticate and validate the performance of the proposed technique, classification performance is quantified with two extensively used publicly available data sets [22], [23], and comparison is provided with “seven” well-established techniques in the literature: those by Yen [9], Goumas [10], Lou [11], Seker [20], Li [21], Samanta [24], and Malhi [25].

This paper is organized as follows. Section II discusses the framework for the feature extracted based on higher order statistics and WT, Section III presents the criterion for adaptive feature selection, Section IV illustrates the fault detection mechanism, Section V compares results with the existing techniques, and Section VI contains the concluding remarks.

II. FEATURE EXTRACTION BASED ON HIGHER ORDER STATISTICS AND WT FRAMEWORK

The overall system can be explained in five steps, data acquisition (vibration data), preliminary processing of the data (SWT), feature extraction [computing higher order cumulants (HOCs)], prominent feature selection, and pattern matching for classification. AFHCW adopts hierarchical paradigm to diagnose machine faults from a given data set as shown in the flow chart in Fig. 2.

Digitized vibration data are first decomposed using SWT in order to ensure linear operation on the vibration data prior to feature extraction. Linearity of the operation to conduct time–frequency analysis of the vibration data helps in mitigating the impact of Gaussian noise and results in better fault detection accuracy (details are provided in Section II-C). SWT helps in investigating the frequency contents of the vibration data in different frequency ranges, i.e., nodes. For the decomposed data, HOCs are computed for each node. Features extracted using different cumulant orders behave differently under varying SNR and type of the faults. Decomposition node and cumulant order are selected adaptively, i.e., dominance level of the features is computed on the basis of the characteristics of the data sets.

A. Time/Frequency Domain Analysis for Vibration Signal

Among the time–frequency domain signal processing techniques, e.g., Discrete Fourier Transform (DFT), STFT, and WT, WT can be used for comprehensive analysis of nonstationary vibration signal to reliably extract time and frequency domain contents [6]. DFT of a nonstationary signal $x[n]$ (4) does not exploit the variation in frequency contents with respect to time. Rather, it averages out the frequency content over the whole signal range [6]–[8]

$$X(k) = \sum_{n=0}^{N-1} x[n]e^{-j(\frac{2\pi}{N})kn}, \quad k = 0, 1, \dots, (N - 1). \quad (4)$$

The shortcomings of DFT can be compensated by STFT (5), but STFT suffers from the problem that it gives the same time and frequency resolution for low and high frequencies. The time and frequency resolution remains the same because window size $w[n]$ remains constant throughout the analysis [6]–[8]

$$X(m, k) = \sum_{n=0}^{N-1} x[n]w[m-n]e^{-j(\frac{2\pi}{N})kn}, \quad k=0, 1, \dots, (N-1). \quad (5)$$

In order to overcome the drawback of fixed time–frequency resolution in STFT, wavelet transformation can be used which has the tendency to perform multiresolution analysis.

In case of DWT, low frequencies have good frequency resolution and poor time resolution while high frequencies have good time resolution but poor frequency resolution. This is the case in most of the physical systems, i.e., for low frequencies, frequency resolution is more important than time, and for high frequencies, time resolution is more important than frequency [2], [9], [10], [20]. Adaptive time–frequency resolution in DWT can be obtained by its built-in tendency to adjust window size according to frequency ranges, i.e., for low frequencies, window size in the time domain is large, ensuring good frequency resolution, and for high frequencies, window size is smaller, guaranteeing good time resolution.

Fig. 3 gives the decomposition tree for DWT. Digitized vibration data are passed through high-pass $h[n]$ and low-pass $g[n]$ quadrature mirror filters (QMFs). QMFs are finite-impulse response (FIR) filters or infinite-impulse response filters. Filter selection is a very crucial part of analysis of the signal using WT. In AFHCW, Daubechies (Db5) filter [9] is used which is a FIR filter. After filtration, data are down sampled (because only half of the frequencies are left). The decomposition at any $i - th$ level results into two components: The low-frequency component is known as approximation coefficients A_i , and the high-frequency component is known as the detail coefficients D_i [26]. The detailed and approximation coefficients at level $i + 1$ are obtained by decomposing approximation coefficients A_i through low-pass and high-pass filters, resulting into A_{i+1} and D_{i+1} . At every decomposition level, data are first passed through QMF filters and then down sampled

$$y_{approx}[n] = x[n] * g[n] \quad (6)$$

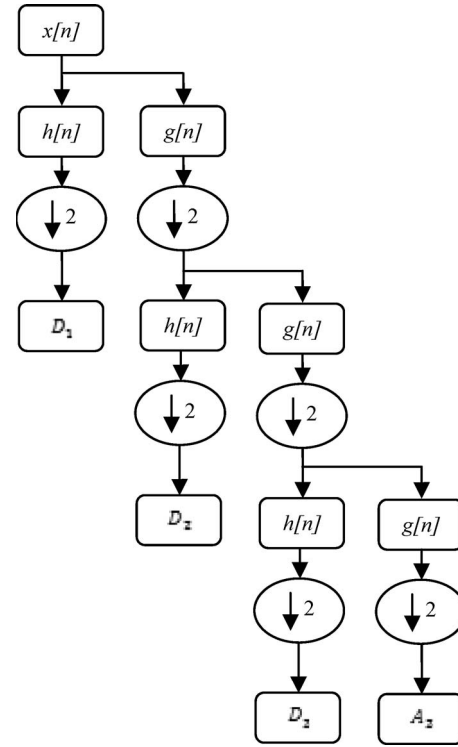


Fig. 3. DWT.

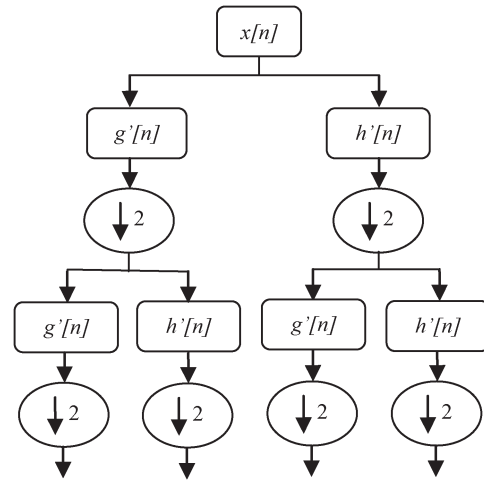


Fig. 4. DWPT.

or

$$y_{approx}[n] = \sum_{k=-\infty}^{k=\infty} [k] \times g[n - k] \quad (7)$$

and

$$y_{detailed}[n] = x[n] * h[n] \quad (8)$$

or

$$y_{detailed}[n] = \sum_{k=-\infty}^{k=\infty} x[k] \times h[n - k]. \quad (9)$$

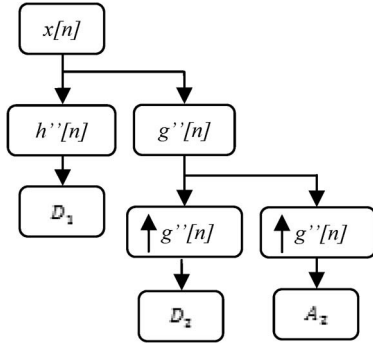


Fig. 5. SWT.

It is clear from the decomposition tree in Fig. 3 that high frequencies (higher decomposition levels) have poor frequency resolution and low frequencies (lower decomposition levels) have good frequency resolution. Certain applications demand good time and frequency resolution both at low and high frequencies. In that case, discrete WPT (DWPT) can be used in which both the approximation and detailed components are decomposed [6], [11], [12], [20], [27] as in Fig. 4, contrary to DWT in which only the approximation component is decomposed. WPT gives more detailed insight into frequency contents for the incoming data at the cost of increased computational complexity because both the detailed and approximation components are decomposed.

In our proposed framework, the type of the distribution for the vibration data is to be preserved before and after the decomposition (details are provided in Section II-C). In case of DWT and WPT, signal is processed into two steps, i.e., first filtration then down-sampling. Filtration of data using FIR filters is a linear process which does not change the type of the Gaussian distribution of the data [28]. However, down-sampling may change the nature of distribution before and after wavelet decomposition. In order to protect the type of distribution, down-sampling is to be avoided.

SWT avoids down-sampling and ensures linear operation on the vibration data. In case of SWT, instead of down-sampling the data at each decomposition level, filters are up-sampled as shown in Fig. 5. The main cost of avoiding down-sampling is the extra computational overhead in processing the same number of coefficients at each decomposition level as compared to DWT in which computational overhead is halved after each decomposition level. The advantage of SWT is the linear operation on the incoming vibration data. Computational cost for SWT and WPT is identical because, in both techniques, the number of coefficients to be processed at each decomposition level remains the same.

In SWT, the decomposition of the data into n_{decomp} levels results in $n_{decomp} + 1$ frequency nodes as shown in Fig. 5. SWT performs multiresolution analysis, i.e., decomposition of the signal into nodes. In AFHCW, data are decomposed into four levels, resulting into five nodes, i.e., four nodes for detailed coefficients $cD_1 \rightarrow cD_4$ and one node for approximation coefficients cA_4 as highlighted in Fig. 6. Fig. 6 shows the coefficients of five nodes in case of inner-race fault. As we move down the tree, $cD_1 \rightarrow cA_4$, frequency contents vary from

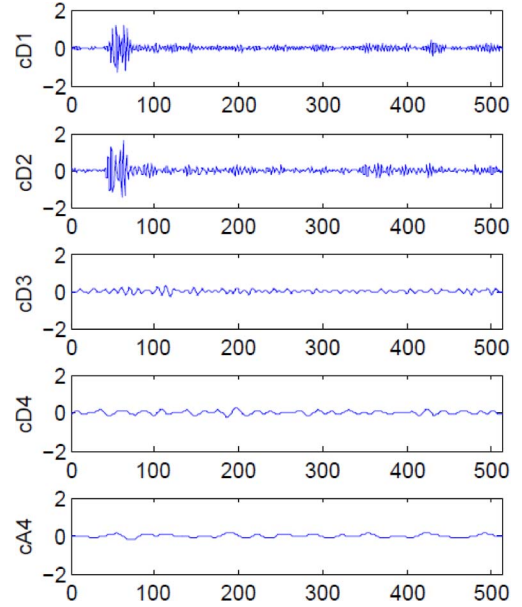


Fig. 6. SWT decomposition for inner-race fault.

high to low frequencies. The decomposition until level 4 is also justified because cA_4 and cD_4 are very smooth signals as most of the transients in the recorded vibration signals are captured by the higher nodes.

B. HOCs for Fault Detection

Cumulant moments are the statistical parameters which are used to characterize any random distribution such as vibration (nonstationary). The cumulant generating function $g(t)$ for a random variable X is defined by (10)[29]

$$g(t) = \log(E(e^{tX})). \quad (10)$$

Then, cumulants of X are given by

$$\begin{aligned} K_1 &= g'(0) \\ K_2 &= g''(0) \\ K_3 &= g'''(0) \\ &\vdots \\ K_n &= g^{(n)}(0) \end{aligned} \quad (11)$$

where e^{tX} is a moment generating function, and \log of moment generating function is a cumulant generating function. The relationship between the cumulants and moments is given by

$$K_r = \mu_r - \sum_{k=1}^{r-1} \binom{r-1}{k-1} K_k \mu_{r-k}. \quad (12)$$

In (12), μ_r is the r -th centralized moment given by

$$\mu_r = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^r \quad (13)$$

where x_i is the i -th sample, N is the total number of samples, and μ is the mean value.

In AFHCW, feature extraction is based on computing HOCs rather than moments because of the following two vital properties.

- 1) From (10), it is clear that the cumulant of the sum of two random variables X and Y is the sum of the individual cumulants, i.e.,

$$K_r(X + Y) = K_r(X) + K_r(Y) \quad (14)$$

where X and Y are statistically independent. This property helps in mitigating the impact of Gaussian noise in AFHCW if vibration data are processed linearly (see Section II-C).

- 2) For Gaussian distribution (15), cumulants generate zero result if the order of cumulant is higher than “2,” i.e.,

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \text{ and} \quad (15)$$

$$K_r(f(x)) = 0, \quad \text{if } r \geq 3 \quad (16)$$

which legitimates the need for higher order statistics based on computing cumulants to guarantee immunity for Gaussian noise.

The standardized third- and fourth-order cumulants [20], [21] represent skewness and kurtosis as in (17) and (18)

$$S_X = \frac{\left[\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^3 \right]}{\sigma^3} \quad (17)$$

$$K_X = \frac{\left[\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^4 \right]}{\sigma^4}. \quad (18)$$

Skewness measures “*asymmetry*” in the data distribution relative to the Gaussian distribution, and kurtosis measures “*peakedness*” of the distribution relative to the Gaussian distribution. In developing AFHCW, we study the fault detection performance of higher order statistics until order 7, i.e., $3 \leq r \leq 7$. The maximum statistical order is limited to “7” as simulation results dictate that further increment in the cumulant order does not give any improvement in the fault detection accuracy (results are provided in Section V).

C. SWT and Cumulants for MHM

In industrial environment, two considerations should be addressed while designing an MHM system.

- 1) Noise level in the vibration signatures could vary drastically over the time. An MHM system should guarantee better performance under varying SNR.
- 2) In order to detect and diagnose faults at the initial stage when actual vibration signatures are very poor, it is vital to test the system’s performance under low SNR.

In the vibration signal, Gaussian noise, either colored or white, is the main contributor toward noise [21]. Further emphasis of this study is to propose the feature extraction technique which mitigates the impact of Gaussian noise. The captured vibration signature may be considered as the original

vibration signal plus Gaussian noise. In (19), $v[n]$ represents the pure vibration signal, and $g[n]$ represents the Gaussian noise

$$s[n] = v[n] + g[n]. \quad (19)$$

In the flowchart in Fig. 2, vibration data are decomposed into different nodes using SWT, so for any arbitrary decomposition level d , the type of the distribution is preserved as given by

$$\begin{aligned} s'[n] &= v'[n] + g'[n] \\ s''[n] &= v''[n] + g''[n] \\ &\vdots \\ s^d[n] &= v^d[n] + g^d[n]. \end{aligned} \quad (20)$$

The cumulants for the coefficients of any decomposition node d [exploiting the property in (14)] can be computed as given in (21) and (22)

$$K_r(s^d[n]) = K_r(v^d[n] + g^d[n])$$

or

$$K_r(s^d[n]) = K_r(v^d[n]) + K_r(g^d[n]) \quad (21)$$

for $r \geq 3$

$$K_r(s^d[n]) = K_r(v^d[n]), \quad \because K_r(g^d[n]) = 0. \quad (22)$$

In (16), it was concluded that, for Gaussian distribution, if the order of the cumulant is greater than “2,” then the result is zero. In (22), the criterion in (16) is applied which justifies that the impact of Gaussian noise is eliminated as long as the cumulant order is higher than “2.” This immunity to the Gaussian noise is achieved by ensuring linear processing on the vibration data prior to feature extraction, which is achieved in SWT and validates the significance of SWT [28]. Computational cost in the case of SWT is higher than that of DWT because, at every decomposition level, the number of coefficients to be processed essentially remains constant as compared to DWT in which the number of coefficients is half that of the previous decomposition level. However, the results in (22) hold true if processing is linear, prior to feature extraction, so that the data distribution remains the same before and after decomposition.

III. ADAPTIVE DECOMPOSITION NODE AND CUMULANT ORDER SELECTION

AFHCW develops a generalized criterion for adaptive feature selection which quantifies the dominancy level for every feature corresponding to a particular decomposition node and the cumulant order. Fig. 7 gives the intuitive criterion for the proposed feature selection. The features are selected based on its tendency of class separability. In Fig. 7, the symbols “+” and “×” represent the data points belonging to two classes (considered for simplicity, but in the paper there are three classes). The test query (point) is represented as “*start*.” In AFHCW, the dominancy level of the features is defined adaptively based upon its tendency to differentiate between the data points of two classes. Fig. 7(a) shows that, with this location of the

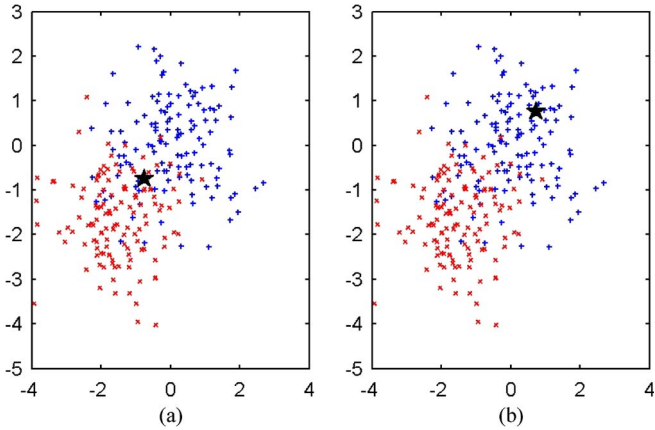


Fig. 7. Adaptive decomposition node and cumulant order selection.

test point, the probability of wrong classification is higher. Fig. 7(b) considers another scenario in which the confidence level is higher to classify the test point. Fig. 7 gives the pictorial representation for only one feature (for simplicity). There are $(n_{decomp} + 1) \times (n_{cum} - 2)$ feature values in total, and the dominance level is computed for each feature value.

Observing the hierarchical paradigm in Fig. 2, vibration data are decomposed using SWT to n_{decomp} levels resulting in $n_{decomp} + 1$ nodes. For each node, cumulants are computed until orders r such that $3 \leq r \leq n_{cum}$. If decomposition nodes are represented in rows and cumulant order varies along the column, the resultant feature matrix contains order $(n_{decomp} + 1) \times (n_{cum} - 2)$ as given by

$$\mathbf{F} = [f_{ij}]_{(n_{decomp}+1) \times (n_{cum}-2)}. \quad (23)$$

According to (23), $(n_{decomp} + 1) \times (n_{cum} - 2)$ features are extracted. The proposed feature selection technique selects the dominant decomposition node and cumulant order, which promises maximum likelihood for accurate classification. Let us assume that there are c_t classes in total. To test the adaptive feature selection, vibration data are split into two portions, one for training and the other for testing. The training and test data are further divided into small chunks containing equal number of samples of the digitized vibration data, i.e., windows. In the training phase, feature matrices are computed for each window of the data for every class. In the testing phase, feature matrix F_{Test} is obtained for the test window of data according to (23). In order to select optimal cumulant order and decomposition node adaptively, the criterion to define dominance for every element in the test feature matrix F_{Test} is given in Proposition 1.

Proposition 1: For each of the $(n_{decomp} + 1) \times (n_{cum} - 2)$ features, the dominance level is quantified on the basis of tendency of the features to ensure class separability (as defined in Fig. 7), maximization of the accumulated distances from $c_t - 1$ classes, and minimization of the distances from one of the classes.

In order to quantify the separability of each feature value, the mean value of the training data points for each class is given by

$$\mathbf{M}_c = [\bar{f}_{ij}]_{(n_{decomp}+1) \times (n_{cum}-2)}. \quad (24)$$

Computations of cumulants involve the power of their orders (13), and to give equal weighting to all the extracted features using different cumulants, it is necessary to normalize the features against their cumulant orders. In AFHCW, feature matrix for the test window F_{Test} and mean matrices for every class M_c are normalized as in (25) and (26), i.e., compute the r -th root of the feature value corresponding to the r -th order cumulant

$$\tilde{\mathbf{F}}_{Test} = \left[r=j+2 \sqrt{\text{abs}(f_{ij})} \right]_{(n_{decomp}+1) \times (n_{cum}-2)} \quad (25)$$

$$\tilde{\mathbf{M}}_c = \left[r=j+2 \sqrt{\text{abs}(\bar{f}_{ij})} \right]_{(n_{decomp}+1) \times (n_{cum}-2)}. \quad (26)$$

To quantify the dominance level of each of the feature value in the test feature vector, the distance matrices are computed for every class as in

$$\mathbf{D}_x = \text{abs}(\tilde{\mathbf{F}}_{Test} - \tilde{\mathbf{M}}_c)_{(n_{decomp}+1) \times (n_{cum}-2)}. \quad (27)$$

Combining the distance matrices for each class, the resultant 3-D array is given in

$$\mathbf{D} = [d_{ijk}]_{(n_{decomp}+1) \times (n_{cum}-2) \times c_t}. \quad (28)$$

The dominance level of every element in the test feature matrix F_{Test} is computed from the 3-D array (28) of difference matrices as in

$$\text{dom}_{ij} = \left(\sum_{k=1}^{c_t} d_{ijk} \right) - 2 \times \min(d_{ijk} | k \in \{1, 2, \dots, c_t\}). \quad (29)$$

The dominance level computed in (29) validates Proposition 1. The indices “ i ” and “ j ” for the element having the maximum value of dom_{ij} correspond to the most dominant decomposition node and the statistical order, respectively. In the above calculations, the dominant feature selection is adaptive, as feature selection is based on the test feature matrix, i.e., the characteristics of the incoming data to be diagnosed are being used to quantify the dominance levels for all the features in the test feature matrix F_{Test} .

IV. FAULT DIAGNOSIS

Fault diagnosis is related to the classification of unknown faults; in this paper, K -th nearest neighbor (KNN) has been used as a classifier. KNN classifier determines the type of unknown fault based on majority rule [30]. The distance of the test point is computed from K nearest neighbors of the training data points. The fault type of unknown test data point is classified as the class which contains the maximum number of neighbors out of K nearest neighbors to the test data point [30]. The dominant feature vector obtained from Section III is used as input for the KNN classifier for fault diagnosis. The classification accuracy of the proposed fault diagnostic model is computed as given in

$$H_r = \frac{\text{Accurate Detection}}{\text{Total Trials}} \times 100\% \quad (30)$$

where H_r represents the “hit ratio” and it measures the percentage of the accurately detected unknown events to the total number of unknown events. Better classification accuracy ensures robustness and reliability of the fault diagnostic model. In Section V, the results are presented for the experiments conducted with three different types of the faults, i.e., inner-race, ball, and outer-race faults. The overall vibration data set is split into two portions, i.e., training data sets (67%) and test data sets (33%). Fault diagnostic model is built using the training data sets, and the test data sets are used to characterize the fault detection accuracy. The fault detection accuracy is measured as the ratio of the correctly detected fault points to the total number of test points as defined in (30).

V. EXPERIMENTAL RESULTS

This section presents the performance evaluation of AFHCW by first explaining the data acquisition model and the simulation setup developed for experimentation. The performance evaluation is presented for features extracted against a particular statistical order with explicit emphasis on the performance of second-order statistics [9]–[19] under poor SNR. The classification performance of AFHCW is presented in Section V-C. In order to validate the performance of the proposed technique, the studies are conducted on publicly available and extensively used actual and simulated vibration data sets. Moreover, Section V-D substantiates the enhanced performance of the proposed scheme in comparison with the “seven” well-established techniques in the literature: those by Yen [9], Goumas [10], Lou [11], Seker [20], Li [21], Samanta [24], and Malhi [25].

A. Data Acquisition and Simulation Setup

1) *Actuation Vibration Data*: Fig. 8 illustrates the experimental setup for recording the actual vibration data sets [22]. Experiments were conducted with three different faulty bearings: inner race, rolling element, i.e., ball, and outer race. Faults in the bearing were created by electrodischarge machining. Faulty bearings are supporting the shaft of the motor, and the load is 2HP with a speed of 1750 r/min. The data have been collected through accelerometers using a 16-channel digital-audio-tape recorder and sampled at the rate of 12 000 samples per second.

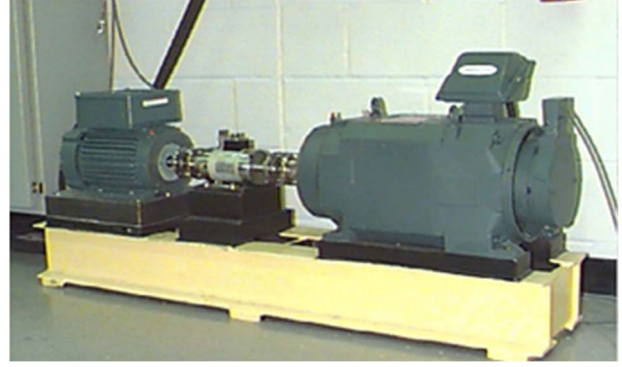


Fig. 8. Experimental setup [22].

2) *Simulated Data*: For simulation studies of AFHCW, vibration data for the bearing faults based on the transfer function with continuous hammering of the rotary machine [23] have been used. The transfer function is given by (31), and the most common input forces for the vibration signal, i.e., mass unbalance, misalignment of the shaft, and bearing defects, are given by (32)–(34), shown at the bottom of the page, respectively. The harmonic amplitudes A , number of harmonics N , phase relationships θ , speed of rotation f_s , and bearing defect frequencies f_b are adjusted [23] for the forces (32)–(34) to simulate the vibration data for inner-race, ball, and outer-race faults with different severity.

3) *Simulation Setup*: Simulations are carried out using MATLAB. The captured vibration data are subdivided into two portions, one part is for training and the other is for testing. KNN classifier is used to measure the performance of the proposed scheme. In the fault diagnosis model, two parameters are optimized: 1) the dominant number of features to be considered for classification as obtained from (29) and 2) the number of nearest neighbors, i.e., K . These parameters are optimized using grid search and fivefold cross validation, and the optimal value of K is “4” [30], [31]. The optimality criterion is based upon maximization in diagnostic accuracy (30). To check the robustness of AFHCW in comparison with the existing techniques, experiments are conducted under varying SNR, as adverse as -10 dB. In order to incorporate the impact of randomness, experimental results are averaged for 1000 experimental runs.

$$H(s) = \frac{2.283s^4 + 642.6s^3 + 4.457 \times 10^8 s^2}{s^4 + 619.3s^3 + 4.42 \times 10^8 s^2 + 4.234 \times 10^{10} s + 5.697 \times 10^{15}} \quad (31)$$

$$F_{mass-unbalance} = A_1 \cos(2\pi f_s t + \theta_1) \quad (32)$$

$$F_{misalignment} = f_s \sum_{k=1}^{N_m} A_k^m \cos(2\pi k f_s t + \theta_m^k) \quad (33)$$

$$F_{bearing} = f_s \sum_{k=1}^{N_b} A_k^b \cos(2\pi k f_b t + \theta_b^k) \quad (34)$$

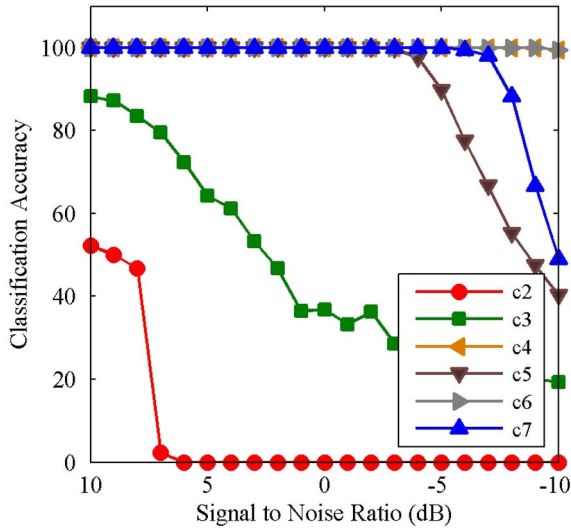


Fig. 9. Fault detection accuracies for ball fault.

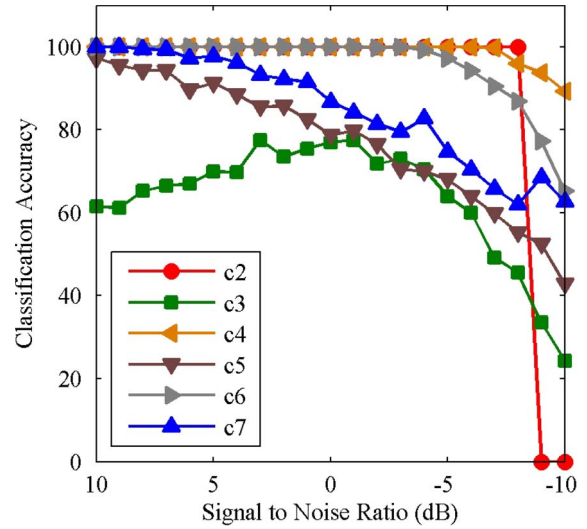


Fig. 10. Fault detection accuracies for inner-race fault.

B. Performance Evaluation for Individual Cumulant Order

In these studies, features are extracted by computing cumulants for the nodes obtained after SWT. The fault detection accuracies for all the cumulant orders c_2 – c_7 are observed explicitly for each type of the faults. Fault detection accuracies are measured using (30). In these experiments, vibration data belong to actuation faults. SNR is varied from 10 to -10 dB. Fault detection at poor SNR values guarantees diagnostic capability of the proposed system at the incipient stage.

Fig. 9 presents the individual classification performance for the extracted features based on cumulant orders c_2 – c_7 in case of ball fault. In Fig. 9, fault detection accuracy for the feature extracted based on second-order cumulant is the worst case scenario because second-order statistics [9]–[19] are the most vulnerable to Gaussian noise. Overall, features extracted based on fourth- and sixth-order cumulants are giving better classification accuracies as compared to all other cumulants. According to (22), the fault detection accuracies for features extracted based on HOCs (with order higher than 2) should not change in the presence of Gaussian noise under ideal scenario. The observed variation is due to the finite number of samples in each window for which features are extracted. However, the overall variation in the fault detection accuracies for the features based on HOCs is significantly lower as compared to second order.

Fig. 10 illustrates the classification performance in case of inner-race fault when features are extracted for different cumulant orders. It shows that the classification performance for the extracted features based on second-order cumulants suddenly drops to zero when the SNR reduces beyond a certain level, i.e., -8 dB. Features extracted based on fourth-order cumulant results in good classification accuracy with poor SNR values. Fig. 10 further shows that the variation in the fault classification accuracy for the features based on higher order statistics is higher than that based on second-order statistics.

Fig. 11 gives the classification performance of individual cumulants in case of outer-race faults. By observation, it seems that second-order cumulant is performing better (100% accu-

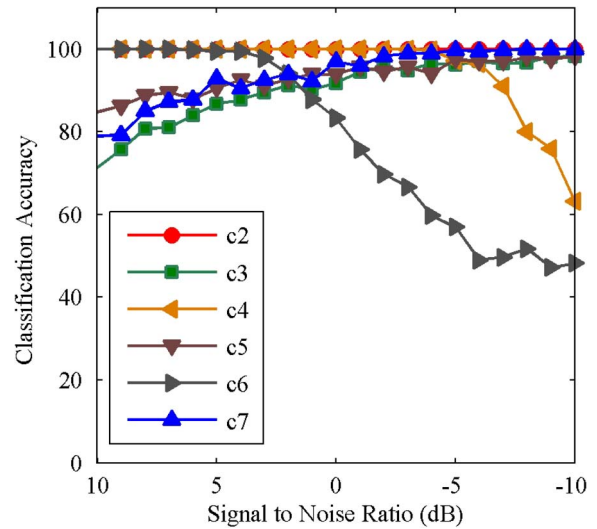


Fig. 11. Fault detection accuracies for outer-race fault.

racy) as compared to other cumulants. Since, for the other two faults, i.e., inner race and ball, second-order cumulant is giving zero fault detection accuracy, i.e., $H_r = 0$, it is always identifying the fault as outer race irrespective of the fault type. Moreover, the fourth-order cumulant, which is giving better performance for the inner-race and ball faults (see Figs. 9 and 10), gives relatively poor performance in the case of the outer-race fault. Overall, features based on third-, fifth-, and seventh-order cumulants are giving better fault detection accuracies.

It is clear, as we decrease the value of SNR, that the performance of the classifier with the extracted features based on second-order cumulant drops suddenly because second-order statistics are the most susceptible to Gaussian noise. The inability of the second-order cumulant to perform well under poor SNR justifies the usage of HOCs.

In Figs. 9–11, the fault detection accuracies vary for the extracted features using different cumulant orders with the variation both in the SNR values and type of the faults. Table I lists the best case cumulant orders for different types of faults

TABLE I
FAULT TYPE VERSUS BEST CASE CUMULANT ORDERS

Fault Type	Cumulant Order
Inner Race	4
Ball	4, 6
Outer Race	3, 5, 7

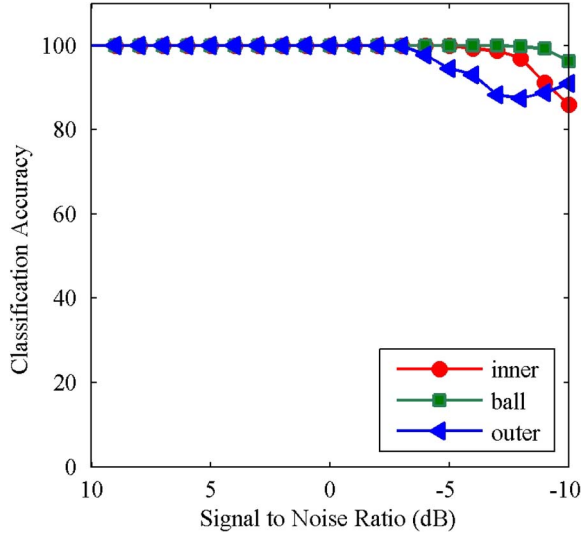


Fig. 12. Adaptive cumulant order selection—SWT.

which are giving overall better performance for different types of faults. However, none of the cumulant order is performing well for all the three types of faults, indicating the need for adaptive selection of the order of cumulants as in the proposed technique presented in Section III.

C. Adaptive Decomposition Node and Cumulant Order Selection

AFHCW proposes a novel technique for adaptive decomposition node and cumulant order selection. Features are extracted by computing HOCs of the nodes obtained after SWT. Out of the whole range of the features as in (23), the dominance levels are computed according to (29). Fig. 12 gives the classification performance of AFHCW for the actual vibration data in case of inner-race, ball, and outer-race faults. Observing the classification accuracies, Fig. 12 shows that AFHCW not only outperforms the feature extraction schemes based on second-order statistics [9]–[19] but also ensures best results as compared to features extracted based on individual cumulant orders (see Figs. 9–12).

Table II gives the whole summary of analysis both for the actual as well as simulated vibration data sets. It lists the worst case classification accuracies for any type of fault when SNR is -10 dB. In case of actual vibration data sets, the diagnostic accuracy is enhanced in AFHCW by almost 23% (86%–63%) as compared to the fourth-order cumulant [21] which is the best case scenario in case of individual performance of the cumulants. The worst case accuracy for fault detection in case of second-order cumulant is 0% because these are the most vulnerable to Gaussian noise.

Table III illustrates the reason for limiting the cumulant order to “7.” It shows that increasing the cumulant order beyond a

TABLE II
WORST CASE CLASSIFICATION ACCURACIES (SNR = -10 dB)

Symbol	Details	Actual Data	Simulated Data
C_2	2nd Order	0.00%	0.00%
C_3	3rd Order	19.25%	50.42%
C_4	4th Order	63.25%	55.57%
C_5	5th Order	40.25%	11.29%
C_6	6th Order	48.25%	46.71%
C_7	7th Order	49.00%	23.28%
C_{2-7}	Adaptive	86.14%	77.00%

TABLE III
CUMULANT ORDER OPTIMIZATION (SNR = -10 dB)
WORST CASE CLASSIFICATION ACCURACIES

Symbol	Details	% Accuracy
C_{2-4}	Adaptive	70.86
C_{2-5}	Adaptive	78.21
C_{2-6}	Adaptive	61.50
C_{2-7}	Adaptive	86.14
C_{2-8}	Adaptive	34.43
C_{2-9}	Adaptive	31.64

certain upper limit results in degradation of the classification accuracy. The decrement in the classification accuracy is due to the fact that higher order statistics contain lesser information regarding data distribution as compared to lower order statistics.

AFHCW uses SWT rather than WPT and DWT, as WPT involves down-sampling [6] and may change the type of the data distribution prior to feature extraction. The change in the nature of data distribution results in invalidity of (22), so as a result, immunity to the Gaussian noise is lost and overall diagnostic capability of the system is degraded. The computational complexity of WPT and SWT is identical because the number of coefficients to be dealt at any decomposition level is the same. In WPT, the number of coefficients per node is halved after every decomposition level, so the behavior of the distribution cannot be precisely determined for a reduced number of coefficients. Fig. 13 gives the classification performance for WPT which is poorer as compared to that for SWT (see Fig. 12). It validates that linear operation on the vibration data prior to feature extraction based on higher cumulants enhances the overall system’s diagnostic capability.

DWT also involves down-sampling [6] on the data prior to feature extraction based on HOCs as in WPT. The nonlinear operation may change the data distribution, and then, (22) does not remain valid, which would cause degradation in the system’s performance. Fig. 14 gives the fault detection accuracies in case of DWT. Comparing the results in Fig. 14 with that in Fig. 12 (SWT), it validates that the linear operation in case of SWT enhances the system’s fault detection accuracy.

Table IV presents the summary of results for different wavelet techniques, i.e., DWT, WPT, and SWT. The fault detection capability for WPT is better as compared to that for DWT [21] because WPT gives more detailed time–frequency analysis by decomposing both the detailed and approximation coefficients (see Fig. 4). Overall, time–frequency domain analysis in AFHCW using SWT enhances the fault detection capability of the system by 16% as compared to WPT and 42% by DWT.

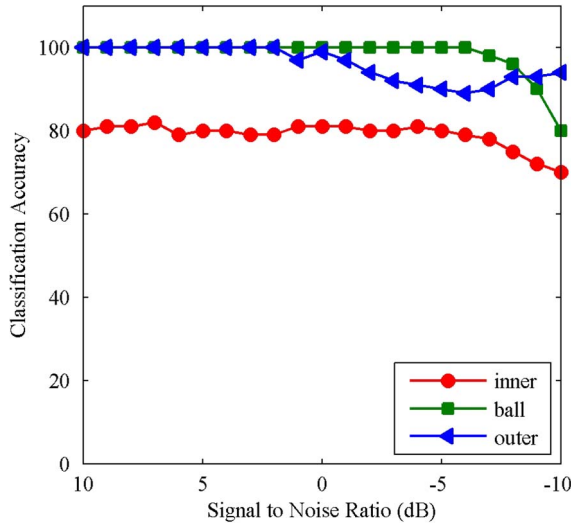


Fig. 13. Adaptive cumulant order selection—WPT.

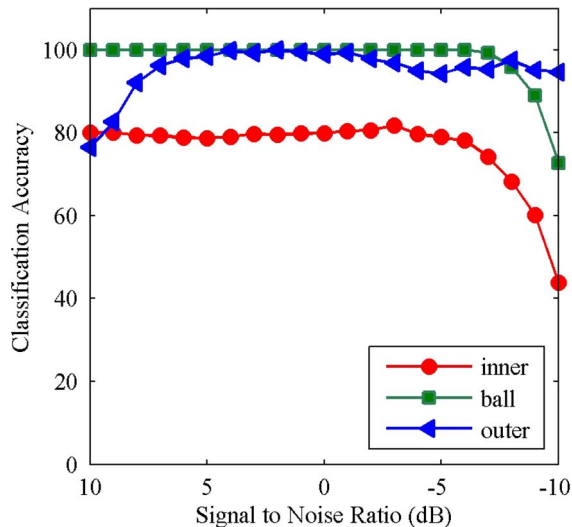


Fig. 14. Adaptive cumulant order selection—DWT.

TABLE IV
WORST CASE CLASSIFICATION ACCURACIES (SNR = -10 dB)

Symbol	Details	% Accuracy
DWT	Discrete Wavelet Transform	43.86
WPT	Discrete Wavelet Packet Transform	70.60
SWT	Stationary Wavelet Transform	86.14

D. Comparative Performance Evaluation of AFHCW

Comparative performance of AFHCW with seven well-established existing schemes, i.e., those by G. G. Yen [9], S. K. Goumas [10], X. Lou [11], S. Seker [20], F. Li [21], B. Samanta [24], and A. Malhi [25], using both publicly available actual and simulated vibration data [22], [23] shows that AFHCW has outperformed all the techniques. Table V presents the comparative performance evaluation of AFHCW with the actual vibration data in terms of average classification accuracies. Experiments are conducted for three different types of faults, i.e., inner-race, outer-race, and ball faults, and the average of the fault

TABLE V
COMPARATIVE PERFORMANCE (ACTUAL VIBRATION DATA) AVERAGE ACCURACY AT -10 dB

Scheme	Accuracy	Scheme	Accuracy
G. G. Yen [9]	33.33%	F. Li [21]	86.00%
S. K. Goumas [10]	33.33%	B. Samanta [24]	33.33%
X. Lou [11]	33.33%	A. Malhi [25]	38.09%
S. Seker [20]	78.25%	AFHCW	91.23%

TABLE VI
COMPARATIVE PERFORMANCE (SIMULATED VIBRATION DATA) AVERAGE ACCURACY AT -10 dB

Scheme	Accuracy	Scheme	Accuracy
G. G. Yen [9]	39.28%	F. Li [21]	79.33%
S. K. Goumas [10]	33.33%	B. Samanta [24]	33.33%
X. Lou [11]	33.33%	A. Malhi [25]	33.33%
S. Seker [20]	71.33%	AFHCW	85.33%

classification accuracies is computed. It shows that AFHCW performs better as compared to the already existing techniques under incipient fault conditions.

Table VI presents the comparative performance evaluation for the proposed scheme for the simulated data sets at -10 dB. These results show that AFHCW yields much superior classification accuracy and robustness against noise as compared to the existing techniques. Classification performance of the existing techniques deteriorates significantly under poor SNR conditions, whereas AFHCW guarantees sustained performance in harsh industrial environment when SNR may vary drastically over the time.

VI. CONCLUSION

In this paper, we present an MMH technique, named AFHCW, that adaptively extracts vibration features based on cumulant order and decomposition node in WT and provides high accuracy in identifying faults, specifically under poor SNR conditions. In order to mitigate the impact of Gaussian noise in industrial environment, the usage of HOCs and SWT is very promising. AFHCW adopts a criterion to test the dominance level of the features for adaptive selection of the cumulant order and decomposition node to capture the dynamics of the vibration data. Simulation studies justify that the adaptive selection of the statistical order enhances the diagnostic performance of the proposed system by 23% for actual vibration data sets and 21% for simulated vibration data sets, as compared to the individual performances of the extracted features based on a particular cumulant order. The comparative performance evaluation with existing schemes validates that the diagnostic capability of AFHCW is enhanced by 5.23% for actual vibration data and 6% for simulated data. In a nutshell, the overall performance of the proposed technique promises precise and early detection of machine faults under poor SNR for MMH.

REFERENCES

[1] J. Morel, "Vibratory monitoring and predictive maintenance," *Techniques de l'Ingénieur, Measurement and Control*, vol. RD, art. R6100, Dec. 2002.
 [2] M. E. H. Benbouzid, "A review of induction motors signature analysis as a medium for faults detection," *IEEE Trans. Ind. Electron.*, vol. 47, no. 5, pp. 984-993, Oct. 2000.

- [3] M. F. Yaqub, I. Gondal, and J. Kamruzzaman, "Machine fault severity estimation based on adaptive wavelet nodes selection and SVM (accepted for publication)," in *Proc. IEEE Int. Conf. Ind. Electron. Appl.*, 2011, pp. 1951–1956.
- [4] M. F. Yaqub, I. Gondal, and J. Kamruzzaman, "Severity invariant feature selection for machine health monitoring," *Int. Rev. Elect. Eng.*, vol. 6, no. 1, pp. 238–248, Feb. 2011.
- [5] M. F. Yaqub, I. Gondal, and J. Kamruzzaman, "Machine health monitoring based on stationary wavelet transform and 4th order cumulants," *Aust. J. Elect. Electron. Eng.*, 2011, to be published.
- [6] Z. K. Peng and F. L. Chu, "Application of the wavelet transform in machine condition monitoring and fault diagnostics: A review with bibliography," *Mech. Syst. Signal Process.*, vol. 18, no. 2, pp. 199–221, Mar. 2004.
- [7] T. W. S. Chow and S. Hai, "Induction machine fault diagnostics analysis with wavelet technique," *IEEE Trans. Ind. Electron.*, vol. 51, no. 3, pp. 558–565, Jun. 2004.
- [8] T. Tarasiuk, "Hybrid wavelet-Fourier spectrum analysis," *IEEE Trans. Power Del.*, vol. 19, no. 3, pp. 957–964, Jul. 2004.
- [9] G. G. Yen and K. Lin, "Wavelet packet feature extraction for vibration monitoring," *IEEE Trans. Ind. Electron.*, vol. 47, no. 3, pp. 650–667, Jun. 2000.
- [10] S. K. Goumas, M. E. Zervakis, and G. S. Stavrakakis, "Classification of washing machines vibration signals using discrete wavelet analysis for feature extraction," *IEEE Trans. Instrum. Meas.*, vol. 51, no. 3, pp. 497–508, Jun. 2002.
- [11] X. Lou and K. A. Loparo, "Bearing fault diagnosis based on wavelet transform and fuzzy inference," *Mech. Syst. Signal Process.*, vol. 18, no. 5, pp. 1077–1095, Sep. 2004.
- [12] L. Eren and M. J. Davaney, "Bearing damage detection via wavelet packet decomposition of the stator current," *IEEE Trans. Instrum. Meas.*, vol. 53, no. 2, pp. 431–436, Apr. 2004.
- [13] K. Teotrakool, M. J. Davaney, and L. Eren, "Adjustable-speed drive bearing-fault detection via wavelet packet decomposition," *IEEE Trans. Instrum. Meas.*, vol. 58, no. 8, pp. 2747–2754, Aug. 2009.
- [14] R. Yan and R. X. Gao, "Energy-based feature extraction for defect diagnosis in rotary machines," *IEEE Trans. Instrum. Meas.*, vol. 58, no. 9, pp. 3130–3139, Sep. 2009.
- [15] E. C. C. Lau and H. W. Ngan, "Detection of motor bearing outer raceway defect by wavelet packet transformed motor current signature analysis," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 10, pp. 2683–2690, Oct. 2010.
- [16] J. Liu, W. Wang, and F. Golnaraghi, "An extended wavelet spectrum for bearing fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 57, no. 12, pp. 2801–2812, Dec. 2008.
- [17] D. F. Shi, W. J. Wang, and L. S. Qu, "Defect detection for bearings using envelope spectra of wavelet transform," *J. Vib. Acoust.*, vol. 126, no. 4, pp. 567–574, Oct. 2004.
- [18] W. J. Staszewski, "Wavelet based compression and feature selection for vibration analysis," *J. Sound Vib.*, vol. 211, no. 5, pp. 735–760, Apr. 1998.
- [19] S. Pittner and S. V. Kmarthi, "Feature extraction from wavelet coefficient for pattern recognition tasks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 1, pp. 83–88, Jan. 1999.
- [20] S. Seker and E. Ayaz, "Feature extraction related to bearing damage in electric motors by wavelet analysis," *J. Franklin Inst.*, vol. 340, no. 2, pp. 125–134, Mar. 2003.
- [21] F. Li, G. Meng, L. Ye, and P. Chen, "Wavelet transform-based higher-order statistics for fault diagnosis in rolling element bearings," *J. Vib. Control*, vol. 14, no. 11, pp. 1691–1709, Nov. 2008.
- [22] *Bearing Data Center*, Jan. 2009. [Online]. Available: http://www.eecs.case.edu/laboratory/bearing/welcome_overview.htm
- [23] H. Ocak and K. A. Loparo, "Estimation of the running speed and bearing defect frequencies of an induction motor from vibration data," *Mech. Syst. Signal Process.*, vol. 18, no. 3, pp. 515–533, May 2004.
- [24] B. Samanta and K. R. Al-Balushi, "Artificial neural network based fault diagnostics of rolling element bearings using time-domain features," *Mech. Syst. Signal Process.*, vol. 17, no. 2, pp. 317–328, Mar. 2003.
- [25] A. Malhi and R. X. Gao, "PCA-based feature selection scheme for machine defect classification," *IEEE Trans. Instrum. Meas.*, vol. 53, no. 6, pp. 1517–1525, Dec. 2004.
- [26] J.S. Walker, *A Primer on Wavelets and their Scientific Applications*. New York: Chapman & Hall, 1999.
- [27] V. R. V. Krishnan and P. B. Anto, "Features of wavelet packet decomposition and discrete wavelet transform for malayalam speech recognition," *Int. J. Recent Trends Eng.*, vol. 1, no. 2, pp. 93–96, May 2009.
- [28] J. S. Bendat and A. G. Piersol, *Random Data: Analysis and Measurement Procedures..* Hoboken, NJ: Wiley, 2010.
- [29] A. Hald, "The early history of the cumulants and the Gram–Charlier series," *Int. Stat. Rev.*, vol. 68, no. 2, pp. 137–153, Aug. 2000.
- [30] K. Umamaheswari, S. Sumathi, S. N. Sivanandam, and K. K. N. Anburajan, "Neuro genetic-nearest neighbor based data mining techniques for fingerprint classification and recognition system," *J. ICGST-GVIP*, vol. 7, no. 3, pp. 39–50, Nov. 2007.
- [31] V. Apnik, V. Levin, E. Le, and Y. Cun, "Measuring the VC-dimension of a learning machine," *Neural Comput.*, vol. 6, no. 5, pp. 851–876, Sep. 1994.



M. F. Yaqub (S'10) received the B.S. and M.S. degrees in electrical engineering from the University of Engineering and Technology (UET), Lahore, Pakistan, in 2007 and 2009, respectively. He is currently working toward the Ph.D. degree from Monash University, Gippsland, Vic., Australia.

Prior to joining Monash, he worked as a Lecturer at the Department of Electrical Engineering, UET. Moreover, he served as a Lecturer (visiting faculty) in the University of Management and Technology, Lahore, and the National University of Computer and

Emerging Sciences, Lahore. His research interests include random pulsewidth modulation in power inverters, advanced digital signal processing, and remote machine health monitoring. He has published over 15 peer-reviewed research papers as first author in renowned international conferences and journals.

Mr. Yaqub is a member of the Pakistan Engineering Council.



Iqbal Gondal (M'91) received the B.Sc. degree in electrical engineering from University of Engineering and Technology Lahore, Pakistan in 1990; and M.Sc. and Ph.D. degrees in electrical engineering from Victoria University of Technology, Australia in 1992 and 1996 respectively.

He is leading intelligent networks research group, and his research interests are wireless networks, sensor networks, gene networks, remote machine health monitoring, social networks, and network mining in Monash University, Gippsland, Vic., Australia. He

has published three book chapters and over 100 refereed conference and journal papers. According to Google Scholar, his publications have been cited over 289 times, with an H-index of 8 and G-index of 14. To date, he has successfully supervised seven and currently supervising six Ph.D. students. Currently, he is working as a Director of Information and Communication Technology Strategy for the faculty in Monash. Prior to joining Monash, he worked in the capacity of a Research Fellow and a Senior Software Systems Engineer for seven years in Singapore and Australia with Delphi (GM), Singapore Manufacturing Technology (SimTech), and other industries. He has experience in network design and development, project management, system design and integration, Supervisory Control and Data Acquisition, intelligent techniques, adaptive systems, and wireless switches for financial services.

Mr. Gondal was also a recipient of a Best Paper award in an international conference. He has served in the capacity of Director of Postgraduate Studies for six years, member faculty board, and member of Monash academic board. He is a member of Communication Society (USA) and Engineers Australia.



Joarder Kamruzzaman (M'00) received the B.Sc. and M.Sc. degrees in electrical engineering from Bangladesh University of Engineering and Technology, Dhaka, Bangladesh, in 1986 and 1989, respectively, and the Ph.D. degree in information system engineering from Muroran Institute of Technology, Muroran, Japan, in 1993.

Currently, he is a Faculty Member in the Faculty of Information Technology, Monash University, Gippsland, Vic., Australia. His research interests include computer networks, sensor networks, and computational intelligence. He has published over 150 peer-reviewed publications

which include 40 journal papers and 6 book chapters and edited two reference books on computational intelligence theory and applications.

Dr. Kamruzzaman was a recipient of the Best Paper award in two IEEE sponsored international conferences. He is currently serving as a Program Committee Member of a number of international conferences and an Editor of an international journal.