IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 23, 2016, accepted August 7, 2016, date of publication September 20, 2016, date of current version February 20, 2017.

Digital Object Identifier 10.1109/ACCESS.2016.2608505

Enhancing Fault Classification Accuracy of Ball Bearing Using Central Tendency Based Time Domain Features

MUHAMMAD MASOOD TAHIR[1], ABDUL QAYYUM KHAN[2], NAEEM IQBAL[2], AYYAZ HUSSAIN[3], AND SAEED BADSHAH[4]

[1]Department of Electrical Engineering, International Islamic University, 44000 Islamabad, Pakistan
[2]Department of Electrical Engineering, Pakistan Institute of Engineering and Applied Sciences, 45650 Islamabad, Pakistan
[3]Department of Computer Science, International Islamic University, Islamabad, Pakistan
[4]Department of Mechanical Engineering, International Islamic University, Islamabad, Pakistan

Corresponding author: M. M. Tahir (masood.phdee38@iiu.edu.pk)

ABSTRACT Time-domain (TD) statistical features are frequently utilized in vibration-based pattern recognition (PR) models to identify faults in rotating machinery. Presence of possible fluctuations or spikes in random vibration signals can considerably affect the statistical values of the extracted features consequently. This paper discusses the sensitivity of TD features against the fluctuations occurred in vibration signals while classifying localized faults in ball bearing. Based on the sensitivity level, the features are statistically processed prior to employing a classifier in PR model. A central tendency-based feature pre-processing technique is proposed that enhances the diagnostic capability of classifiers by providing appropriate values. The feature processing reduces undesired impact of fluctuations on the diagnostic model. Several classifiers are utilized to evaluate the performance of the proposed method, and the results are evident of its effectiveness. The associated advantage of the feature pre-processing over the conventional pre-processing of raw data is its computational efficiency. It is worth mentioning that only few values in feature distributions are required to be processed rather than dealing with big TD vibration data set.

INDEX TERMS Pattern recognition, fault diagnosis, feature processing, central tendency of features.

I. INTRODUCTION

Bearing is a critical component of rotating machinery, where basic dynamic loads and forces are applied. A defective bearing causes malfunctions in the machinery, and may even lead to catastrophic failure [1]. To predict the machinery faults early, vibration analysis has been the most popular and widely used technology [2]. Usually, the identification of ball bearings' localized faults is hard because they produce very low amplitudes in vibration signals as compared to the faults like rotor unbalance or misalignment generated by other joint machinery components. Therefore, the spectra of raw vibration signals contain very little diagnostic information regarding the bearing faults [3], and most of the existing fault diagnosis methods involve certain pre-processing of raw vibration data to facilitate the fault detection process. The pre-processing normally includes noise reduction and extraction of appropriate frequency range before further analysis. In this regard, enveloping [4], [5] and empirical mode & wavelets decompositions [6]–[11] are the most frequently used techniques. Pattern recognition (PR) is another popular domain for automatic diagnosis of the faults [12]. Yet the noise in vibration based PR systems often misleads the statistical classifiers in their training phase [13]. Numerous vibration based machine learning methods have so far been employed to detect the bearing faults utilizing time domain (TD) statistical features [14]–[23]. However, maintaining an optimum classification accuracy using a minimal set of features has been a challenge, in spite of applying certain costly pre-processing methods.

Instead of pre-processing the raw vibration data, this research proposes statistical processing of features prior to employing classifier in PR model. Besides efficiency, the pre-processed features considerably enhance the diagnostic capability of the classifier. The TD features include RMS, mean, variance, skewness, kurtosis, crest factor (CF), impulse factor (IF), shape factor (SF), median and range. Fluctuations or spikes may occur randomly in vibration signals, and can consequently alter the statistical information of these features. The feature processing technique is based on the detection of abnormal values or outliers during data preparation stage of supervised learning [12]. The purpose is to supply only the appropriate features to the classifier for better decision making. The outliers are the values in data pattern that do not adapt an expected behavior, and outlier detection methods have been used in a wide variety of applications such as military surveillance, fraud detection for credit cards, intrusion detection in cyber security, insurance and fault detection in critical systems [24]. It is also a well studied area of data mining, and has been classified mainly into statistical approaches, depth-based approaches, deviation-based approaches, distance-based approaches, density-based approaches and high-dimensional approaches [25]. A number of surveys, review articles and books cover these approaches in machine learning and statistical domains [26]–[31]. Data mining generally utilizes a collection of data instances, i.e. pattern, object, record, point, vector, event, case, sample, observation, entity etc [32]. Each data instance is described using a set of features or attributes, which can be of different types such as binary, categorical, or continuous. The nature of attributes also determines the applicability of the outlier detection methods [32], as the selection of right detection method is vital according to the nature of application and normal behavior of the specific phenomena [24].

The literature survey reveals that a little work has been carried out regarding the accurate extraction of diagnostic features for rotating machinery fault diagnosis. Lee et al. [33] recently examined the sensitivity of diagnostic features for prognostic and health management (PHM) system, with respect to the signal quality and failure modes/ operating conditions of the system like speed, load, or torque. The presented methods utilized several features from time and frequency domains to develop algorithms to identify various faults in bearing, gear and shaft [34]. The frequency domain features included the fault frequencies exhibited by the rotating components. The traditional TD statistical features, i.e. RMS, kurtosis, crest factor etc. were also examined for fault detection and prognosis. The authors emphasized that it is critical to reduce signal noises and eliminate outliers before extracting diagnostic features to obtain accurate results and prevent the system from high false alarm rates. Widely implemented existing outlier detection methods, such as distance based [35], density-based spatial clustering of applications with noise [36], and minimum covariance determinant [37], were employed to identify the data points to be discarded. While focusing on the front end of a PHM system, the authors suggested following steps for reliable outcome from the system; 1) outlier detection to remove abnormal data from raw vibration signal, 2) pre-processing of the vibration signal to reduce unwanted noises, 3) cluster based operational mode detection method to group various operating conditions, and 4) neural network training based feature normalization to mitigate the effects of operating conditions on the features.

Difficulty with the above mentioned method is its complexity and computational cost, especially employing the TD features. Reliability of the TD features may be improved by simpler means rather, particularly for PR-based diagnostic models. The random nature of vibration signals can contain fluctuations, which may be due to even change in dynamic operating conditions. The variations in acquired vibration signals consequently produce outlying values in the extracted features. Physical causes or mechanical phenomena behind the occurrence of signal fluctuations are not discussed in this study. However, an important and valid assumption was made here that the particular phenomena should not be associated with the faults under investigation, i.e. bearing faults in our case. This particular situation allows processing of TD features directly instead of pre-processing the huge set of raw vibration data. Authors developed a feature processing method, which is based on central tendency (CT) of the features distributions. The method deals with the possible outliers adequately while preparing data before incorporating classifier in a PR model. Several classifiers including Support vector machine (SVM), BayesNet, Decision Table, and Decision Tree are used to evaluate the proposed method. All the classifiers are found better decision makers while utilizing processed features. The feature processing method works in two distinct steps; 1) detection of outliers present in the features distributions, and 2) discarding the affected instances or examples before introducing classifier in the model. The features are extracted from the segments of faulty vibration signals to form data set. This study utilizes the median score in any feature distribution as its CT measure, as the median effectively isolated the outliers generated due to the fluctuations in vibration signals. Median-based commonly used outlier detection methods include Box Plot and Median absolute variance (MAD). The Box Plot [38] is employed because of offering functional simplicity and suitability to our application.

A substantial advantage of the proposed pre-processing of TD features over conventional pre-processing of raw data is the computational cost, as only few values in a feature distribution are required to be processed rather than processing the huge vibration signals. The adequately processed features are found to be robust and provide higher diagnostic information to the models. Moreover, the pre-processing of TD features has not been reported so far for the purpose of bearing fault diagnosis. Major contributions of this research include:

- The CT based feature processing method is developed and employed prior to use classifier in vibration based PR model to classify ball bearing's localized faults.
- The presented method ensures the provision of only appropriate features values to the classifier, and enhances the fault classification accuracy.
- In addition to computationally efficient, the method is immune to possible fluctuations present in steady state vibration signals.

This paper is organized as follows: Section II explains the bearing faults and experimental setup. Major steps

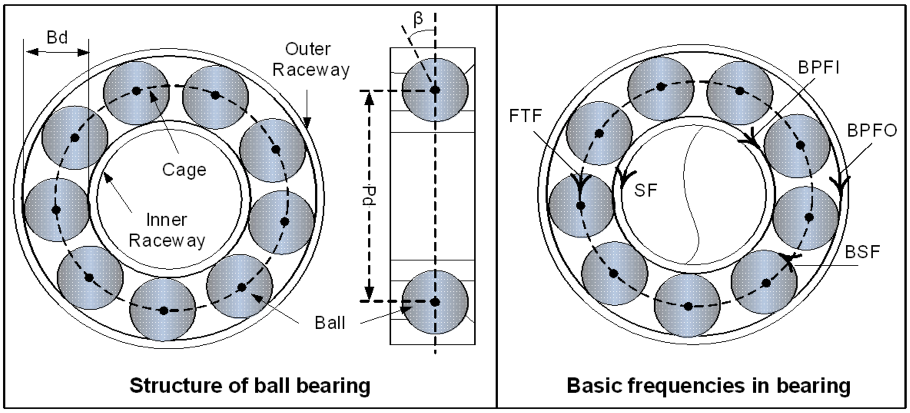

FIGURE 1. Structure and basic frequencies in ball bearing.

involved in the proposed scheme are elaborated in Section III. Section IV discusses the results and findings of the proposed research, whereas the conclusions are drawn in Section V.

II. BEARING FAULTS

There can be several kinds of faults in rolling element bearing, such as surface fatigue damage, bonding and wear. The most common of these faults is the surface fatigue damage, which are further categorized as spalling, crack, or other abnormal conditions [39]. When a localized fault appears on the surface of any element of bearing, cyclical impulsive vibration is originated consequently. For example, an impact is produced when the rolling elements strike a local fault on inner or outer race, or a fault on any rolling element strikes the races. Frequency of the impulsive vibration is known as fault frequency. The value of fault frequency depends on the fault size, rotational speed, and damage location. The main fault frequencies a ball bearing can generate are fundamental train frequency (FTF), ball pass frequency of inner race (BPFI), ball pass frequency of outer race (BPFO), and ball spin frequency (BSF). Figure 1 shows the geometric parameters of ball bearing that involve to generate the characteristic frequencies.

The aforementioned fault frequencies are mathematical described as;

$$FTF = \frac{SF}{2}\left(1 - \frac{B_d}{P_d}Cos\beta\right) \quad (1)$$

$$BPFI = \frac{N_b \times SF}{2}\left(1 + \frac{B_d}{P_d}Cos\beta\right) \quad (2)$$

$$BPFO = \frac{N_b \times SF}{2}\left(1 - \frac{B_d}{P_d}Cos\beta\right) \quad (3)$$

$$BSF = \frac{N_b \times SF}{2}\left(1 - (\frac{B_d}{P_d})^2Cos^2\beta\right) \quad (4)$$

Where SF is the motor driving frequency or rotational frequency of shaft, N_b is the number of bearing balls, B_d is the ball diameter, P_d is the pitch diameter and β is the contact angle.

A. EXPERIMENTAL SETUP

The data set from Curtin University (CU) [40] was used to evaluate the performance of the proposed method. Radial vibration measurements are taken using a machinery fault simulator test rig from SpectraQuest. An accelerometer is mounted on the top of outboard bearing housing to acquire data for inner race (IR) and outer race (OR) bearing faults. Ball bearings model MB ER-16K are used to rotate healthy shaft containing a loader in the middle, as shown in schematic of the setup in Figure 2. The bearing model contains 9 balls ($N_b = 9$) having diameter (B_d) 7.94 mm, whereas the pitch diameter (P_d) is 38.50 mm. Motor speed was 29 Hz measured using tachometer. Vibration signals along with their respective speed signals are captured at the sampling rate of 51200 samples/sec. For more details, the reader is referred to [40].

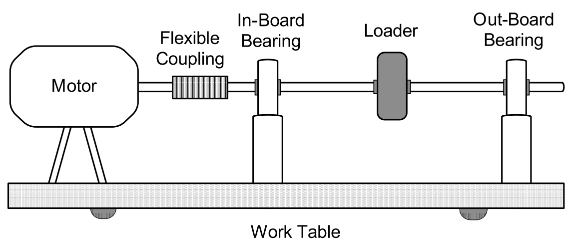

FIGURE 2. Schematic of experimental setup.

B. DATA VALIDATION

The metal to metal impacts produce a *ringing* of the bearing's housing and its support structure, which is modulated by the fault frequency. The low-level impulses having an amplitude-modulating effect on the vibration signal spread over a wide frequency range. Therefore, the conventional frequency analysis are often not be able to show the bearing's fault frequencies, and thus envelope analysis has been used as benchmark method for the purpose over many years [3]. The enveloping is based on demodulation of high frequency resonance associated with fault impacts. It extracts the signal of interest from an overall raw vibration signal while focusing on a narrow band range in the high frequency band. In spite the envelope analysis is a powerful technique, the improper selection of the frequency band can render the analysis ineffective [41]. Therefore, appropriate frequency bands were selected using the fast kurtogram method proposed by Antoni [42] to find the bands in terms of central frequency along with the bandwidth [43].

To validate the data set, enveloping was implemented for IR and OR vibration signals using NI LabVIEW. The function *OAT Envelope Detection (Even Angle Output)* was employed prior to obtain power spectra of the envelope signals. The function utilizes a frequency band and tachometer signal for extracting even-angle envelope signal. The angle-domain output signal maintains constant number of samples per revolution to mitigate the effect of speed variations. Table 1 shows BPFI and BPFO frequencies representing IR and OR faults. The calculated center frequencies and bandwidths of IR and OR faults are also shown in the table respectively, i.e. CF_{IR} & BW_{IR}, and CF_{OR} & BW_{OR}.

TABLE 1. Bearing fault frequencies along with the central frequencies and bandwidths (Hz).

SF	FTF	BPFI	BPFO	CF_{IR}	BW_{IR}	CF_{OR}	BW_{OR}
29	11.5	157.4	103.6	24533	2133	9066	1066

Figure 3(a) shows the enveloped spectrum of IR fault, in which BPFI is present along with the side-bands of shaft speed. Figure 3(b) shows the several harmonics of BPFO to represent OR fault. Hence, the data set contains all the required information.

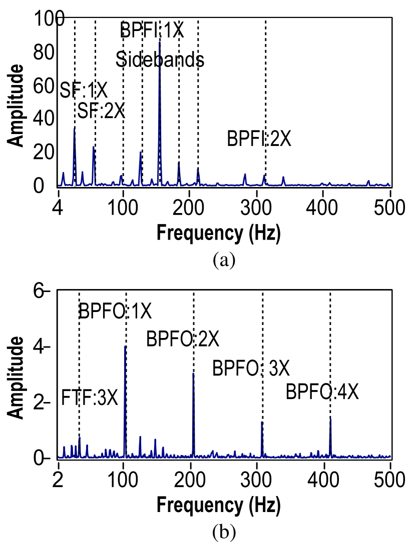

FIGURE 3. Enveloped spectra of the bearing faults. (a) Enveloped spectrum of IR fault. (b) Enveloped spectrum of OR fault.

III. MATERIALS AND METHODOLOGY

The proposed fault diagnostic scheme consists of four steps, which are elaborated in the block diagram in Figure 4.

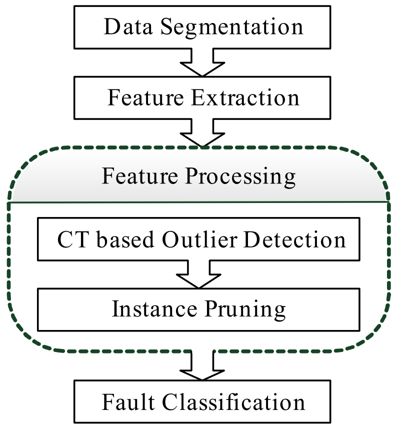

FIGURE 4. Block diagram of bearing's fault diagnostic scheme.

Details of the proposed methodology are in the following subsections.

A. DATA SEGMENTATION AND FEATURE EXTRACTION

At first step, vibration signals were segmented, and features were extracted from the segments to form data set at second step. Every signal of 10 seconds duration was divided into 40 segments. As the motor speed was 29 Hz, each segment holds vibratory history of more than 7 revolutions of the shaft. In this way, the segments contained a valid sample length to compute trustworthy statistical features. The following ten feature were extracted from every segment of each fault to form the data set for the supervised learning and fault classification.

$$RMS = \left(\frac{1}{N}\sum_{i=1}^{N}[X(i)]^2\right)^{\frac{1}{2}} \quad (5)$$

$$Mean(\mu) = \frac{1}{N}\sum_{i=1}^{N}X(i) \quad (6)$$

$$Variance(\sigma^2) = \frac{1}{N}\sum_{i=1}^{N}(X(i)-\mu)^2 \quad (7)$$

$$Skewness = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{(X(i)-\mu)}{\sigma}\right)^3 \quad (8)$$

$$Kurtosis = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{(X(i)-\mu)}{\sigma}\right)^4 \quad (9)$$

$$CrestFactor = \frac{max(|X|)}{RMS} \quad (10)$$

$$ImpulseFactor = IF = \frac{max(|X|)}{\frac{1}{N}\sum_{i=1}^{N}|X(i)|} \quad (11)$$

$$ShapeFactor = SF = \frac{RMS}{\frac{1}{N}\sum_{i=1}^{N}|X(i)|} \quad (12)$$

$$Median = magnitude\left(\frac{N+1}{2}\right) \quad (13)$$

$$Range = max(X) - min(X) \quad (14)$$

In the above relations, X is the sequence of samples obtained after digitizing the time domain signals, $X(i)$ is the amplitude of i^{th} sample and N is the total number of samples in the sequence.

B. FEATURE PROCESSING

The third step implements the feature processing mechanism, which is the central theme of the proposed research. The feature processing ensures the use of most appropriate data by the classifiers. The processing was done in two phases; firstly the Box Plot outlier detection method was utilized to implement the Median-based Outlier Detection (MOD) procedure, and then the instances were pruned based on the outcome of MOD. Details are in the following subsections.

1) CT BASED OUTLIER DETECTION

The CT attempts to describe a set of data with a single value using mean, median and mode [44]. However, each measure can be more advantageous under different conditions. Mean is frequently used and effective measure when data distribution is symmetric. However, it is susceptible to outliers and skewed data because of including every element of data set as part of its calculation. On the other hand, median score occupies the middle position in an ordered data set and less sensitive to the outliers [44]. Usually, more than half elements in a vibration sample belong to normal distribution, and accordingly outliers in the extracted features should lie above the median score when sorted in an ascending order.

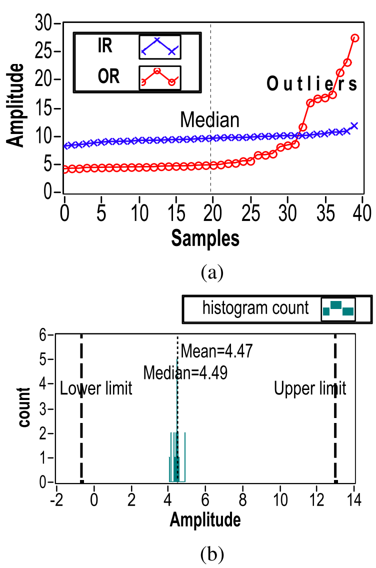

FIGURE 5. Median as central tendency measure. (a) Ascending ordered elements of kurtosis feature extracted from IR and OR vibration data segments. (b) Histogram of lower half distribution of OR kurtosis feature bisected from the median.

Figure 5(a) shows the ascending ordered kurtosis feature, where median values of IR and OR distributions are almost insensitive to the outliers. Figure 5(b) shows histogram of lower half distribution of OR kurtosis feature, i.e. the part of distribution below its median score. The histogram shows that the lower half distribution lies well within the limits, and therefore no outlier is present in this half of distribution.

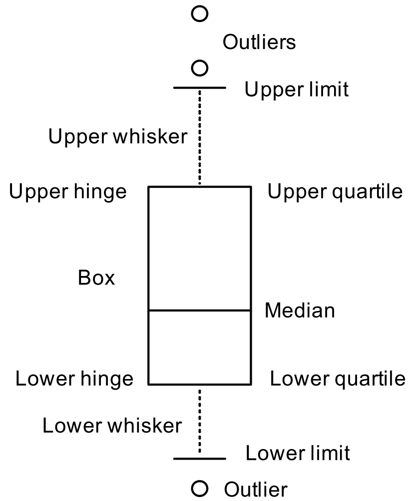

FIGURE 6. Parameters of Box Plot.

Figure 6 shows parameters of Box Plot that include a median, two hinges at lower and upper quartiles (fourths), and two whiskers that connect these hinges to the limits. Box Plot can be constructed using the following rules.

- Arrange the data distribution in ascending order.
- Calculate the first quartile ($Q1$), third quartile ($Q3$) and the inter-quartile range ($IQR = Q3 - Q1$).
- Compute Lower limit $= Q1 - (1.5 \times IQR)$ and Upper Limit $= Q3 + (1.5 \times IQR)$, where the value 1.5 acts as a *scale* to define the limits. Any value in a feature distribution below the Lower Limit or above the Upper Limit was considered as outliers.

The median based Box outlier detection method offers simplicity and suits to our situation.

2) INSTANCE PRUNING

Instance pruning is the process of discarding the unsuitable instances marked by the MOD, which was applied to every feature separately. The instances containing outliers were discarded during the data preparation process. Each element of every features was checked whether that lie within the relevant range, otherwise that element was marked as an outlier. The main stages of the instances pruning are given below;

- Detection of outliers present in the features by employing the MOD.
- Discarding the instances containing one or more outliers.

Table 2 shows the marked outlying value in every feature by the MOD. The algorithm puts a dummy value "99.999" to mark the outliers, so that the affected instances (rows) could easily be discarded from the data set. There is no loss of information as long as the captured vibration data is of appropriate length, from which the instances are generated. The smoother values of features were then fed to the classifiers for training and testing purpose.

TABLE 2. Sample instances containing marked outliers as 99.999.

RMS	Mean	Variance	Skewness	Kurtosis	CF	IF	SF	Median	Range	Class
8.883	0.038	78.915	0.016	7.984	8.864	12.160	1.372	0.014	148.161	OR
19.764	0.037	390.642	-0.089	9.620	8.694	13.019	1.497	0.043	327.657	IR
99.999	0.062	**99.999**	0.224	**99.999**	13.373	19.534	**99.999**	-0.098	**99.999**	OR
19.502	0.044	380.368	-0.111	9.782	7.610	11.540	1.516	0.255	276.110	IR
19.267	0.006	371.236	-0.021	10.034	7.221	11.118	1.540	0.165	273.081	IR
7.930	0.055	62.889	0.050	4.295	4.848	6.353	1.311	-0.039	74.771	OR
7.899	0.064	62.389	0.033	4.344	5.924	7.777	1.313	0.045	82.998	OR
18.973	0.040	359.999	-0.018	9.434	7.793	11.741	1.507	0.103	286.655	IR
8.045	0.044	64.733	0.036	8.288	10.402	13.992	1.345	-0.047	159.318	OR
19.512	0.005	380.729	-0.100	10.766	9.438	14.432	1.529	0.157	**99.999**	IR
8.360	0.050	69.892	-0.073	5.503	6.746	9.048	1.341	0.080	106.921	OR
18.759	0.026	351.926	-0.044	9.685	7.367	11.185	1.518	0.251	267.394	IR
21.021	0.020	441.919	-0.108	9.950	9.528	14.415	1.513	0.099	341.423	IR
99.999	0.069	**99.999**	-0.094	**99.999**	10.973	16.602	**99.999**	**99.999**	243.627	OR
7.823	0.064	61.194	0.076	4.320	5.505	7.208	1.309	0.029	81.223	OR
19.349	0.009	374.399	-0.052	9.578	7.021	10.745	1.530	0.221	266.801	IR
99.999	0.041	**99.999**	0.085	**99.999**	12.640	20.299	**99.999**	0.095	**99.999**	OR
20.143	0.010	405.759	-0.056	8.779	6.885	10.213	1.483	0.231	272.107	IR
7.744	0.045	59.978	0.045	4.353	5.789	7.598	1.312	0.009	81.879	OR

C. FAULT CLASSIFICATION

The supervised learning paradigm was used for bearing's fault classification at final stage. The paradigm requires a dataset with labeled patterns to train a classifier. Once the classifier is trained, it is then employed to test the unknown examples, as elaborated by Figure 7. The SVM, BayesNet, Decision Table and Decision Tree were implemented to evaluate the performance of the proposed method.

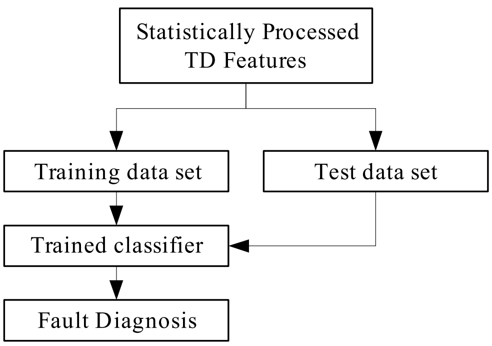

FIGURE 7. Supervised learning and fault classification procedure.

The performance estimation method $K-fold$ cross validation was utilized for training and testing of all the classifiers. The method splits the data D into k equal parts, i.e. D_1, ..., D_k. A single part is retained as the validation data to test the model, and the remaining $k-1$ parts are used as training data. The process is repeated k times, with each of the k parts used exactly once as the validation data. The k results obtained from the folds are then averaged out to produce a global accuracy or single estimation.

$$Accuracy_{global} = \frac{1}{k}\sum_{j=1}^{k} Accuracy_j \tag{15}$$

The advantage of cross validation method is that all examples are involved for both training and validation. This study employed the commonly used 10-fold cross-validation method.

A brief description of the classifiers are presented. The interested reader is referred to [45] for details on SVM, [46] for BayesNet, [47] for Decision Table, and [48] for Decision Tree.

1) SVM

The SVM is an efficiently learning system, which utilizes a hypothesis space of linear functions in a high dimensional feature space. The simple SVM algorithm solves a binary classification problem. The data are separated by a hyperplane defined by support vectors, which are subset of training data as shown in Figure 8. These support vectors can create complex boundaries, and the margin of separation is maximized between each class of data.

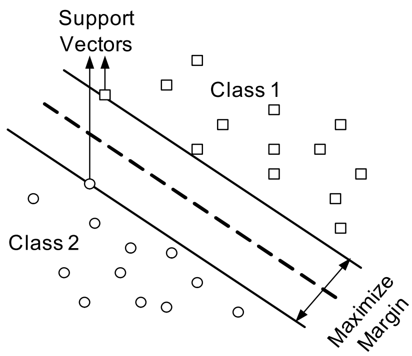

FIGURE 8. Margin and support vectors in SVM.

Suppose N-dimensional input data x_i belong to class 1 or class 2, where $i = 1...N$. The associated labels $y_i = 1$ represent class 1 and $y_i = -1$ correspond to class 2. In case the data are separable linearly, a hyperplane $f(x)$ can be determined to separate the data. The hyperplane follows the rule $f(x_i) \geq 0$ in case x belongs to class 1, whereas $f(x_i) < 0$ if x belongs to class 2.

$$f(x) = w \cdot x + b = \sum_{k=0}^{N} w_k x_k + b \tag{16}$$

where w is the N-dimensional normal vector which defines the hyperplane and b is the learning bias. An optimal hyperplane maximizes the geometrical margin and is obtained by solving the convex quadratic optimization problem $min\ \frac{1}{2}\ \|w\|^2$.

2) BayesNet

Bayesian network is a well established algorithm to represent probabilistic relations among random variables in a set as a directed acyclic graph. The variables are represented by nodes, and are connected via edges depicting causal relations between variables. Conditional probability distribution is given at each variable. In the example show in Figure 9, the edge from node A to node B indicates that A causes B.

Conditional probability distribution (CPD) is specified at each node having parents, whereas the prior probability is specified at node having no parents, i.e the root node. The CPDs of variables B and C, are $P(B|A)$ and $P(C|A)$ respectively, whereas the prior probability of A is $P(A)$. The edges

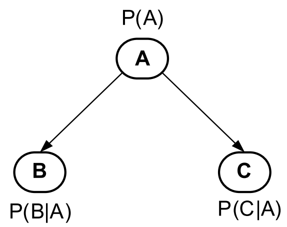

FIGURE 9. A simple Bayesian network.

in the graph represent the joint probability distribution (JPD) of the connected variables, i.e. the JPD for edge (A, B) is $P(B, A)$ representing the probability of joint event $B \cap A$. The fundamental rule of probability calculus shows that

$$P(B, A) = P(B|A) \cdot P(A) \tag{17}$$

Generally the JPD for BayesNet with given nodes $X = X_1, \ldots, X_Z$ is

$$P(X) = \prod_{j=1}^{Z} P(X_j|Parents(X_j)) \tag{18}$$

where $Parents(X_j)$ is the parent set of node X_j. The Equation 18 is known as the chain rule representing the JPD of all variables in the Bayesian network, as the product of probabilities against each variable are its parents values.

Computation of probability for each variable is performed using the known values of other variables. In other words, once some evidence is asserted into the network regarding states of the variables, the effect of evidence is propagated through the network and probabilities of adjacent nodes are updated in every propagation. The inference process can be formalized mathematically as the Bayes theorem:

$$P(X|Y) = \frac{P(Y|X) \cdot P(X)}{P(Y)} \tag{19}$$

The above relation represents the probability of node X given evidence Y. The posterior probability of node X is $P(X|Y)$, which can be computed using known likelihood $P(Y|X))$ and prior probability $P(X)$. The term $P(Y)$ denotes a normalizing factor.

Fault diagnosis, in a qualitative sense, can be seen as the cause-effect or fault-symptom relations. Every fault and symptom is modeled by a random variable in the network with a probability distribution. Taking the observed symptoms or evidences as input to the network, probabilities of every fault are computed accordingly to the Bayes rule.

3) DECISION TABLE

Decision tables are used to model complex rule sets comprising conditions and actions in a compact way. A decision table is formulated to have four quadrants as shown in Table 3. The quadrants on the left describe the conditions as well as the actions being modeled in the table, while the right hand quadrants show the corresponding condition alternatives and action entries. The columns in the right quadrants are called rules. Thus each column has two portions; some of its values are in the condition portion, called inputs, while others are in the action area, termed as outputs. A rule, hence, associates a set of input conditions to a corresponding set of output actions.

TABLE 3. Four quadrants structure.

Conditions	Condition Entries
Actions	Action Entries

Decision tables can be represented in a number of ways according to data being modeled. One way of exploiting decision tables is to model cause-effect relationship by replacing conditions with causes and actions with faults. An example of such an application is machine diagnostics where, on the basis of prior knowledge (rules) connecting observed symptoms to faults, an unknown fault can be classified into known faults.

4) DECISION TREE

The classification algorithm producing decision tree is based on information theory. Construction of the tree is based on the learning data set, that is mentioned below;

- Leaf nodes or answer nodes contains the name of fault class
- Decision nodes or non-leaf nodes specifies some test to be carried out on a single attribute or feature value. A decision node contains one branch and sub-tree against each possible outcome of the test.

Criteria to select the root of tree is based on information gain. The measure is used to select among the candidate features at each step of the tree growth. Information gain (S, A) of a feature A relative to a collection of examples S is defined as;

$$Gain(S, A) = Entropy(S) - \sum_{v \in Value(A)} \frac{|S_v|}{|S|} Entropy(S_v) \tag{20}$$

where $Values(A)$ is the set of all possible values for attribute A, and S_v is the subset of S for which feature A has value v.

The information gain is the expected reduction in entropy, which measures homogeneity in a set of examples. The gain measures how well a given feature separates the training examples according to the target classification.

IV. RESULTS AND DISCUSSION

Vibration data from CU was found appropriate to validate the proposed feature processing based fault diagnosis scheme. Figure 10(a) and Figure 10(b) show the time domain signals of IR and OR bearing faults respectively. The OR signal, containing fluctuations, has been cut into two segments Seg_{AB} and Seg_{BC} as marked in Figure 10(b). The envelope analysis of both the segments was performed using CF_{OR} and BW_{OR} in Table 1. The Seg_{AB} does not contain any fluctuations and the spectrum of its enveloped signal testifies the OR fault frequencies (BPFO) clearly, as shown in Figure 11(a). The enveloped spectrum is quite similar to that of full OR signal

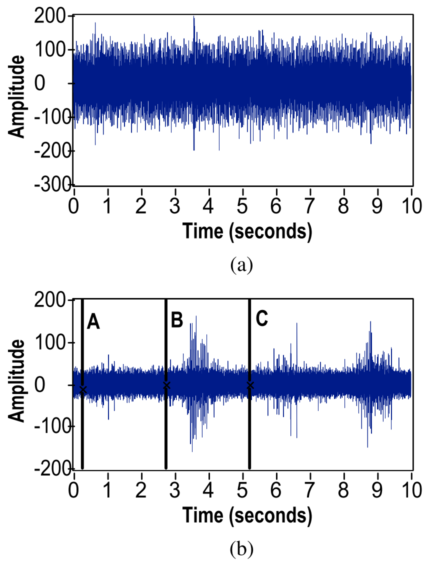

FIGURE 10. **Waveforms of IR and OR faults. (a) Waveform of IR faults. (b) Waveform of OR faults.**

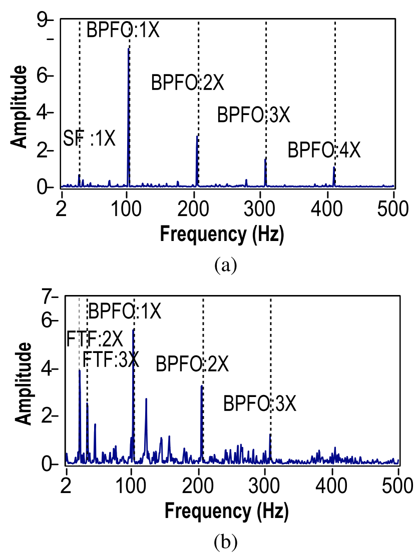

FIGURE 11. **Enveloped Spectra of the segments Seg_{AB} and Seg_{BC} of OR waveform. (a) Enveloped Spectrum of Seg_{AB}. (b) Enveloped Spectrum of Seg_{BC}.**

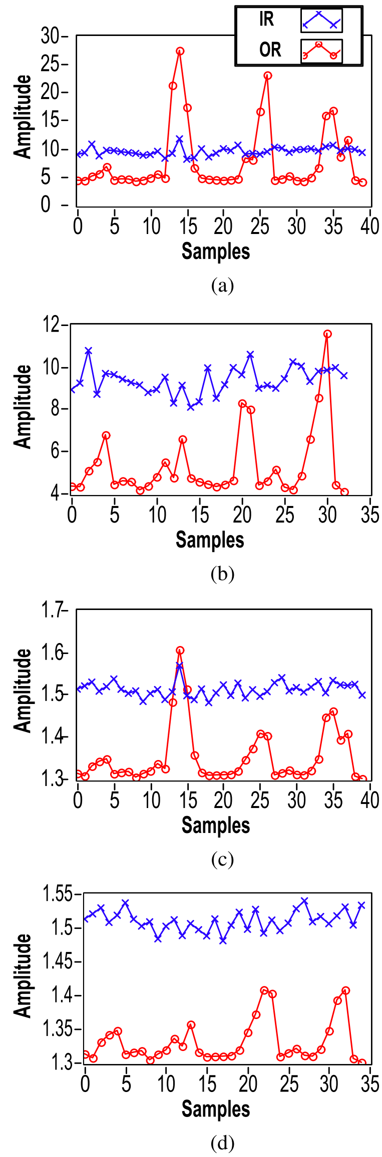

FIGURE 12. **Feature processing via MOD. (a) Disribution of raw values of Kurtosis feature. (b) Kurtosis values after discarding outliers. (c) Disribution of raw values of Shape Factor feature. (d) Shape Factor values after discarding outliers.**

already shown in Figure 3(b). On the other hand, the Seg_{BC} of OR fault contains some fluctuations, and its enveloped spectrum shows some extra frequencies regarding the bearing cage (FTF) in Figure 11(b).

The discussion about the phenomena is out of scope for the present research. It is worth mentioning that the features extracted from OR signal generated outliers due to the fluctuations in Seg_{BC}, which significantly reduced the classification accuracy of the classifier. Nevertheless, the phenomenon exhibited in the Seg_{BC} is undesired to study OR bearing fault.

Outliers in a feature, extracted from different faulty signals, can cause overlapping of elements from those fault classes. This may reduce the diagnostic capability of that particular feature, and is a factor of misleading the classifiers. The MOD adequately handled the issue and ensured the usage of only smoother distributions of the features in diagnostic process.

Figure 12(a) shows the raw elements of kurtosis feature extracted from IR and OR faulty signals, which are overlapping repeatedly with each other mainly due to the outliers in OR feature. These outliers were detected by the MOD, and the respective instances were discarded later by the proposed scheme.

The value of 1.5 for *scale* parameter, suggested by [38], was good enough for the problem as almost all the detected outliers belonged only to the fluctuated parts of OR signal. Figure 12(b) shows the kurtosis elements of IR and OR fault

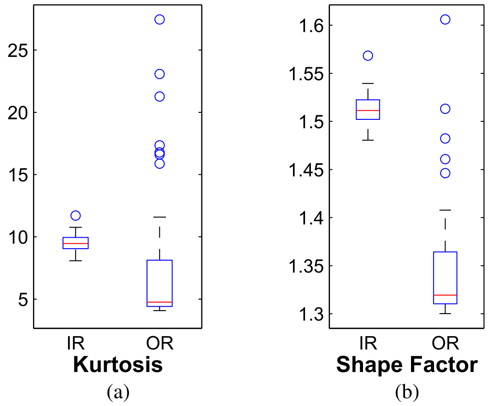

FIGURE 13. Outlier detection via Box Plot. (a) Box Plot applied on kurtosis features from IR and OR faults. (b) Box Plot applied on Shape Factor features from IR and OR faults.

TABLE 4. Individual accuracies of the features before and after processing, using SVM, BayesNet, decision table and decision tree classifiers.

Feature	SVM		BayesNet		D. Table		D. Tree	
	BP	AP	BP	AP	BP	AP	BP	AP
RMS	99.2	100	100	100	100	100	100	100
Mean	64.4	64.4	72.7	72.7	72.2	72.2	73.0	73.0
Variance	100	100	100	100	100	100	100	100
Skewness	83.7	85.2	86.4	87.1	88.3	89.8	94.3	95.1
Kurtosis	93.2	99.2	93.9	100	95.5	100	95.5	100
CF	86.2	86.4	86.1	86.4	88.1	88.3	87.1	87.1
IF	87.1	87.4	88.3	88.3	90.2	90.5	90.5	90.2
SF	93.9	100	95.5	100	95.5	100	95.5	100
Median	79.7	79.9	84.0	84.3	84.0	84.3	84.0	84.0
Range	90.2	100	93.9	100	93.9	100	93.9	100

classes after discarding the outliers. The figure elaborates smoother distributions of the feature against both the faults. Similarly, Figure 12(c) and Figure 12(d) show the SF feature before processing (BP) and after processing (AP) by the MOD. Box Plot outlier detection process is shown in Figure 13(a) and Figure 13(b) for kurtosis and SF features respectively. In this way, every feature was processed separately to mark the outliers in their respective distributions. During the data preparation process, for the training and testing of the classifier, only those instances were selected which are free from outliers.

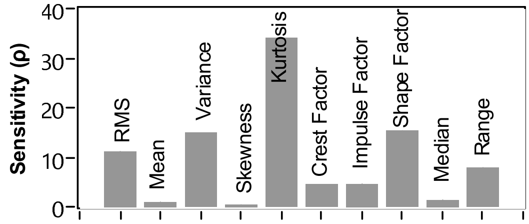

FIGURE 14. Sensitivity of the features against OR fluctuations.

Diagnostic capability of the classifiers was observed against every feature individually. Table 4 elaborates the impact of proposed method on every feature's fault identification ability, using SVM, BayesNet, Decision Table and Decision Tree classifiers implemented in Weka software. Several features improved their performances significantly in terms of enhancing the classification accuracy of the classifiers. Those features were particularly affected whose elements from both fault classes were overlapped due to the fluctuations present in the OR signal. The features include Kurtosis, SF and Range. It is worth noticing that every feature has shown different sensitivity level against these fluctuations.

Figure 14 shows the sensitivity level of every feature (ρ) against the OR signal fluctuations. The following relation was used to calculate their sensitivity levels.

$$\rho = \frac{Range\ of\ Upper\ Half\ Distribution}{Range\ of\ Lower\ Half\ Distribution} \tag{21}$$

The OR ascending ordered feature distribution is bisected into two halves from its median point. The ratio of the upper half to the lower provides an impact of fluctuations through the spread of outliers in the respective halves of the distribution. The Kurtosis feature was affected most by these fluctuations, whereas the skewness showed least sensitivity to the same. Figure 15 shows the skewness feature, which is hardly effected by the fluctuations in OR signal, and consequently the results in Table 4 demonstrate least improvement in the classification accuracy.

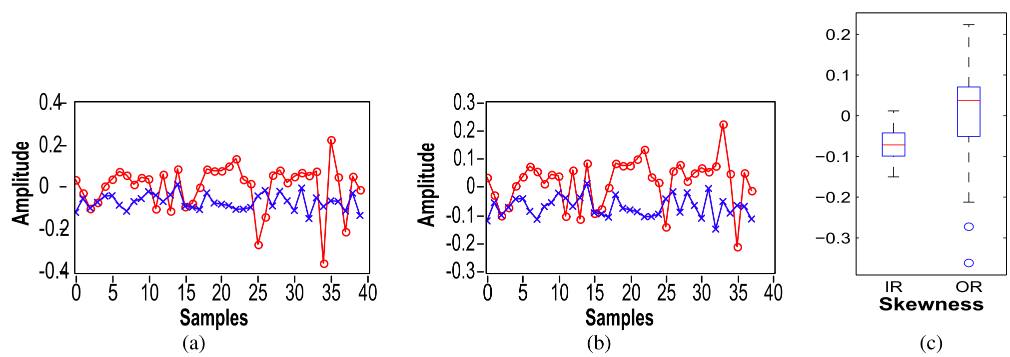

FIGURE 15. Skewness feature processing. (a) Disribution of raw values. (b) After discarding outliers. (c) Box Plot.

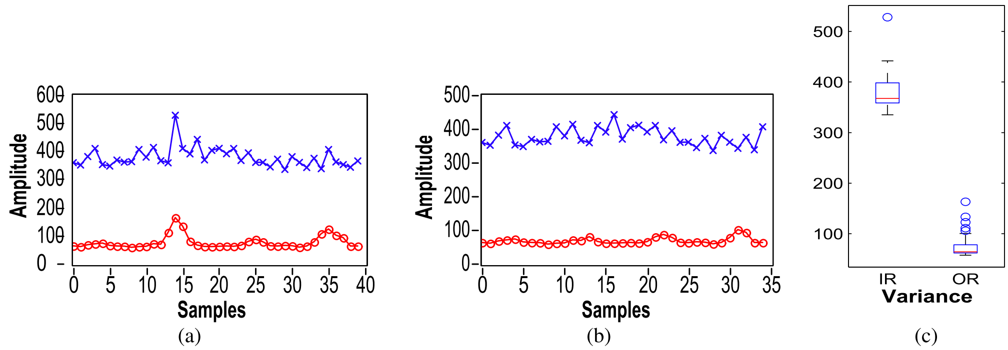

FIGURE 16. Variance feature processing. (a) Disribution of raw values. (b) After discarding outliers. (c) Box Plot.

On the other hand, there may be case where a sensitive feature apparently does not improve its accuracy even after the processing. For instance, Figure 16 illustrate the processing of variance feature. Figure 16(a) exhibits the sensitivity of the feature against the OR signal fluctuations, and consequently Box Plot (Figure 16(c)) detects several outliers in the respective distribution. As a result, the MOD produces quite smoother distribution, as shown in Figure 16(b)). Yet, no enhancement in classification accuracy is noticed for variance feature in Table 4. The reason is that the feature was well separating the faults even before processing, i.e. no overlapping exists between the distributions of two classes. But the unprocessed feature may not demonstrate the same performance if more fault classes are added to the PR-model. Nevertheless, it is worthwhile to disassociate the effect of unrelated fluctuations on the sensitive TD features before further processing.

V. CONCLUSIONS AND FUTURE RESEARCH

In this research, the effect of feature processing on vibration-based PR model was investigated with the intent to diagnose ball bearing's localized faults. It was observed that undesired fluctuations occurring randomly in vibration signals produce outliers in the extracted TD features, which mislead the classifiers in their supervised learning process. It was also noticed that the occurrence of these fluctuations was not associated with the bearing's fault under investigation. On the basis of the above observations, this paper presented a new CT based feature processing method to detect the outliers adequately by implementing the MOD, and the affected instances were pruned at the next stage based on the MOD outcome. In this way, the presented technique assures the employment of only relevant and precise features in the diagnostic models. The SVM, BayesNet, Decision Table, and Decision Tree classifiers were used to evaluate the proposed method, and the classifiers were found better decision makers when processed features were utilized. Several features, i.e. kurtosis, shape factor and range, considerably improved their individual diagnostic capability as per their sensitivity levels to the signal fluctuations and separation ability. A substantial advantage of the proposed feature processing over conventional preprocessing of raw data is its computational efficiency, as only few values in any feature distribution are required to be processed rather than dealing with big TD data.

The future research may investigate the adaptive value of the *scale* parameter to implement the MOD for more accurate results, especially in systems where greater number of fault classes involved, i.e. healthy bearing, ball fault, or multiple localized faults in addition to IR and OR faults.

ACKNOWLEDGMENT

The authors would like to thank Dr. Gareth Forbes from the Department of Mechanical Engineering CU, for providing the vibration data and validating the schematic diagram of his experimental setup used in this research.

REFERENCES

[1] N. Tandon and A. Choudhury, "A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings," *Tribol. Int.*, vol. 32, no. 8, pp. 469–480, 1999.
[2] V. Wowk, *Machinery Vibration: Measurement and Analysis*. New York, NY, USA: McGraw-Hill, 1991.
[3] R. B. Randall, *Vibration-Based Condition Monitoring: Industrial, Aerospace and Automotive Applications*. New York, NY, USA: Wiley, 2011.
[4] N. Sawalhi, R. Randall, and H. Endo, "The enhancement of fault detection and diagnosis in rolling element bearings using minimum entropy deconvolution combined with spectral kurtosis," *Mech. Syst. Signal Process.*, vol. 21, no. 6, pp. 2616–2633, 2007.
[5] W. Wang and H. Lee, "An energy kurtosis demodulation technique for signal denoising and bearing fault detection," *Meas. Sci. Technol.*, vol. 24, no. 2, p. 025601, 2013.
[6] W. Caesarendra, P. Kosasih, A. Tieu, C. Moodie, and B.-K. Choi, "Condition monitoring of naturally damaged slow speed slewing bearing based on ensemble empirical mode decomposition," *J. Mech. Sci. Technol.*, vol. 27, no. 8, pp. 2253–2262, 2013.
[7] 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.
[8] C. Smith, C. M. Akujuobi, P. Hamory, and K. Kloesel, "An approach to vibration analysis using wavelets in an application of aircraft health monitoring," *Mech. Syst. Signal Process.*, vol. 21, no. 3, pp. 1255–1272, 2007.
[9] V. Purushotham, S. Narayanan, and S. A. Prasad, "Multi-fault diagnosis of rolling bearing elements using wavelet analysis and hidden Markov model based fault recognition," *NDT E Int.*, vol. 38, no. 8, pp. 654–664, 2005.
[10] J. Altmann and J. Mathew, "Multiple bandpass autoregressive demodulation for rolling-element bearing fault diagnosis," *Mech. Syst. Signal Process.*, vol. 15, no. 5, pp. 963–977, 2001.

[11] S. Abbasion, A. Rafsanjani, A. Farshidianfar, and N. Irani, "Rolling element bearings multi-fault classification based on the wavelet denoising and support vector machine," *Mech. Syst. Signal Process.*, vol. 21, no. 7, pp. 2933–2945, 2007.

[12] T. W. Rauber, E. M. do Nascimento, E. D. Wandekokem, and F. M. Varejão, *Pattern Recognition based Fault Diagnosis in Industrial Processes: Review and Application*, A. Herout, Ed. Rijeka, Croatia: InTech, 2010.

[13] S. Ericsson, N. Grip, E. Johansson, L.-E. Persson, R. Sjöberg, and J.-O. Strömberg, "Towards automatic detection of local bearing defects in rotating machines," *Mech. Syst. Signal Process.*, vol. 19, no. 3, pp. 509–535, 2005.

[14] B. Samanta, K. Al-Balushi, and S. A. Al-Araimi, "Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection," *Eng. Appl. Artif. Intell.*, vol. 16, nos. 7–8, pp. 657–665, 2003.

[15] L. Jack and A. Nandi, "Fault detection using suppoet vector machines and artificial neural networks, augmnted by genetic algorithms," *Mech. Syst. Signal Process.*, vol. 16, nos. 2–3, pp. 373–390, 2002.

[16] A. Rojas and A. K. Nandi, "Practical scheme for fast detection and classification of rolling-element bearing faults using support vector machines," *Mech. Syst. Signal Process.*, vol. 20, no. 7, pp. 1523–1536, 2006.

[17] B. Samanta and K. 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, 2003.

[18] B. S. Yang, T. Han, and J. L. An, "ART–SKOHONEN neural network for fault diagnosis of rotating machinery," *Mech. Syst. Signal Process.*, vol. 18, no. 3, pp. 645–657, 2004.

[19] L. Zhang, L. B. Jack, and A. K. Nandi, "Fault detection using genetic programming," *Mech. Syst. Signal Process.*, vol. 19, no. 2, pp. 271–289, 2005.

[20] V. Sugumaran and K. Ramachandran, "Automatic rule learning using decision tree for fuzzy classifier in fault diagnosis of roller bearing," *Mech. Syst. Signal Process.*, vol. 21, no. 5, pp. 2237–2247, 2007.

[21] P. Kankar, S. C. Sharma, and S. Harsha, "Fault diagnosis of ball bearings using machine learning methods," *Expert Syst. Appl.*, vol. 38, no. 3, pp. 1876–1886, 2011.

[22] V. Sugumaran and K. I. Ramachandran, "Effect of number of features on classification of roller bearing faults using SVM and PSVM," *Expert Syst. Appl.*, vol. 38, no. 4, pp. 4088–4096, 2011.

[23] M. Saimurugan, K. I. Ramachandran, V. Sugumaran, and N. R. Sakthivel, "Multi component fault diagnosis of rotational mechanical system based on decision tree and support vector machine," *Expert Syst. Appl.*, vol. 38, no. 4, pp. 3819–3826, 2011.

[24] K. Singh and S. Upadhyaya, "Outlier detection: Applications and techniques," *Int. J. Comput. Sci. Issues*, vol. 9, no. 1, pp. 307–323, 2012.

[25] H.-P. Kriegel, P. P. Kröger, and A. Zimek, "Outlier detection techniques," in *Proc. 16th ACM Int. Conf. Knowl. Discovery Data Mining (SIGKDD)*, Washington, DC, USA, 2010.

[26] M. Agyemang, K. Barker, and R. Alhajj, "A comprehensive survey of numeric and symbolic outlier mining techniques," *Intell. Data Anal.*, vol. 10, no. 6, pp. 521–538, 2006.

[27] V. J. Hodge and J. Austin, "A survey of outlier detection methodologies," *Artif. Intell. Rev.*, vol. 22, no. 2, pp. 85–126, 2004.

[28] M. Markou and S. Singh, "Novelty detection: A review—Part 1: Statistical approaches," *Signal Process.*, vol. 83, no. 12, pp. 2481–2497, 2003.

[29] M. Markou and S. Singh, "Novelty detection: A review—Part 2: Neural network based approaches," *Signal Process.*, vol. 83, no. 12, pp. 2499–2521, 2003.

[30] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, no. 3, pp. 1–58, 2009.

[31] J. Zhang, "Advancements of outlier detection: A survey," *ICST Trans. Scalable Inf. Syst.*, vol. 13, no. 1, pp. 1–26, 2013.

[32] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. Reading, MA, USA: Addison-Wesley, 2005.

[33] H. Lee, C. Byington, and M. Watson, "PHM system enhancement through noise reduction and feature normalization," in *Proc. IEEE Aerosp. Conf.*, Mar. 2010, pp. 1–10.

[34] M. Watson, M. Begin, S. Amin, J. Sheldon, H. Lee, and C. Byington, "A comprehensive high frequency vibration monitoring system for incipient fault detection and isolation of gears, bearings and shafts/couplings in turbine engines and accessories," in *Proc. ASME Turbo Expo*, 2007, pp. 885–894.

[35] E. M. Knorr and R. T. Ng, "Algorithms for mining distance-based outliers in large datasets," in *Proc. Int. Conf. Very Large Data Bases*, 1998, pp. 392–403.

[36] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. KDD*, vol. 96. 1996, pp. 226–231.

[37] D. M. Rocke and D. L. Woodruff, "Identification of outliers in multivariate data," *J. Amer. Statistical Assoc.*, vol. 91, no. 435, pp. 1047–1061, 1996.

[38] J. W. Tukey, *Exploratory Data Analysis*. Reading, MA, USA: Addison-Wesley, 1977.

[39] Q. Liu, F. Chen, Z. Zhou, and Q. Wei, "Fault diagnosis of rolling bearing based on wavelet package transform and ensemble empirical mode decomposition," *Adv. Mech. Eng.*, vol. 5, p. 792584, Jul. 2013.

[40] G. Forbes. (2012). *Inner Race and Outer Race Faults Vibration Data Using Machine Fault Simulator*, accessed on 30 Jun. 2014. [Online]. Available: http://data-acoustics.com/measurements/bearing-faults/bearing-1/

[41] E. Bechhoefer and P. Menon, "Bearing envelope analysis window selection," in *Proc. Annu. Conf. Prognostics Health Manage. Soc.*, 2009, pp. 1–7.

[42] J. Antoni, "Fast computation of the kurtogram for the detection of transient faults," *Mech. Syst. Signal Process.*, vol. 21, no. 1, pp. 108–124, 2007.

[43] J. Antoni. (Jul. 2015). *MATLAB Code to Compute Signal's Fast Kurtogram*, accessed on Jul. 25, 2015. [Online]. Available: https://www.mathworks.com/matlabcentral/fileexchange/48912-fast-kurtogram/content/Fast_kurtogram.m

[44] J. H. Watt, *Research Methods for Communication Science*, 1st ed. Boston, MA, USA: Allyn & Bacon, 1995.

[45] V. N. Vapnik, *Statistical Learning Theory*, vol. 1. New York, NY, USA: Wiley, 1998.

[46] H. Kirsch and K. Kroschel, "Applying Bayesian networks to fault diagnosis," in *Proc. 3rd IEEE Conf. Control Appl.*, vol. 2. Aug. 1994, pp. 895–900.

[47] R. Kohavi, "The power of decision tables," in *Proc. Eur. Conf. Mach. Learn.*, 1995, pp. 174–189.

[48] J. R. Quinlan, "Induction of decision trees," *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, 1986.

MUHAMMAD MASOOD TAHIR received the master's degree in mechatronics from the Aachen University of Applied Sciences, Germany in 2004. He is currently pursuing the Ph.D. degree with the Department of Electrical Engineering, International Islamic University Islamabad, Pakistan. He has been involved in research and development of condition monitoring systems for rotating machinery in commercial sector. His research interests include machine condition monitoring, fault diagnosis, machine learning, vibration analysis, and signal processing.

ABDUL QAYYUM KHAN received the master's degree in systems engineering from the Pakistan Institute of Engineering and Applied Sciences (PIEAS), Islamabad, in 2004, and the Ph.D. degree in electrical engineering from the University of Duisburg-Essen, Germany, in 2010. From 2007 to 2010, he was a Guest Researcher with the Institute of Automatic Control and Complex Systems, University of Duisburg-Essen. He is currently an Associate Professor with the Department of Electrical Engineering, PIEAS. His research interests include fault diagnosis in technical processes, linear and nonlinear observer design, robust control of nonlinear systems, and LMI added optimal design.

NAEEM IQBAL received the Ph.D. degree in stability of control systems from Supelec, University of Rennes-I, France, in 1997. He is currently with the Pakistan Institute of Engineering and Applied Sciences, Islamabad, since 1991. He has supervised several Ph.D. scholars in control engineering. His research interests are robust control, non-linear control, and cyber physical systems.

AYYAZ HUSSAIN received the master's degree in computer science from Quaid-i-Azam University, Islamabad, Pakistan in 2000, the Ph.D. degree in computer science from the National University of Computer and Emerging Sciences, NU-FAST, Islamabad, in 2009, and the Ph.D. degree from the Department of Mechatronics, Gwangju Institute of Science and Technology, South Korea, in 2013. He is currently an Assistant Professor with the Department of Computer Science, International Islamic University, Islamabad. His research interests include image processing, machine learning, and computational intelligence.

SAEED BADSHAH received the master's degree in mechanical engineering from NWFP University of Engineering and Technology Peshawar, Pakistan, in 2006, and the Ph.D. degree from the Institute of Mechanics and Mechatronics, Vienna University of Technology, Austria, in 2011. He is currently an Associate Professor with the Department of Mechanical Engineering, International Islamic University, Islamabad. He is a Life Member of Pakistan Engineering Council. He is also member of American Society of Mechanical Engineers, The Institute of Engineers Pakistan and Society of Mechanical Engineers of Pakistan. His research interests include finite-element modeling, experimental modal testing and analysis, condition monitoring of rotating machines, cellular metals, aluminum foam, inverse problems, identification techniques, structural dynamics, structural optimization, and renewable energy.