A NEW WAVELET FEATURE FOR FAULT DIAGNOSIS OF
ROLLER BEARINGS USING DECISION TREE

1 Manju B R, 2 A.R. Rajan, 3 V. Sugumaran

1 Department of Mathematics, Sree Buddha College of Engineering, Kerala, India, manjubrajendran@yahoo.co.in
2 Department of Mathematics, university of Kerala, Thiruvandapuram, Kerala, India
3 SMBS, VIT University, Chennai campus, Chennai, Tamil Nadu, India, v_sugu@yahoo.com

ABSTRACT

“Fault diagnosis of the roller bearings” as pattern classification problem has three main steps namely, feature extraction, feature selection, and classification. A number of machine-learning algorithms have been successfully used to solve the problem with the help of vibration signals. Feature extraction is one of the most important activities in the whole process, as the strength of the features determines the classification accuracy of the classifiers. This paper investigates the use of new discrete wavelet features for feature extraction and compares the result with that of energy definition of discrete wavelet features using Haar wavelet. The extracted features are then classified using decision tree as classifier. The study reveals that the new proposed feature performs better than that of existing energy definition features.

Key words: Fault diagnosis; Discrete Wavelet Transform; New Wavelet Features; Machine Learning

1.0 INTRODUCTION

Roller bearing failures can be caused by several factors, say
1. incorrect design or installation,
2. acid corrosion,
3. poor lubrication,
4. plastic deformation.

The most common defects are produced by material fatigue after a certain running time. This phenomenon begins with the appearance of minute cracks below the surface of the bearing elements. During operation, the cracks progress to the surface due to the cyclic loads, giving rise to surface pitting and spalling. Statistically [1], 90% of the total amount of different roller bearing faults is related to either an inner or outer race fault; the remaining malfunctions are mostly due to a rolling element fault but rarely a cage fault. Hence the focus of the present work is on inner race fault, outer race fault and their combinations. Such defects generate a series of impact vibrations every time a running
Roller passes over the surfaces of the defects, and as is well known, for different fault types, the impacts will appear with different frequency.

Roller bearings, as important components, are widely used in rotary machines; faults occurring in bearings must be detected as early as possible to avoid fatal breakdowns of machines that may lead to loss of production. A visual inspection of certain features of the spectral analysis is prima-facie adequate for fault identification; however, it doesn’t suffice in case of multiple faults or when fault advancement is slow. Vibration signal is seen to be a preferable source in such cases. To detect the faults in bearings many kinds of methods have been developed and without exception, the most important part of the methods is various signal analysis technique. In the early studies, Fourier analysis has been the dominating signal analysis tool for bearing fault detection. But, there are some limitations of the Fourier transform [2]: the signal to be analysed must be strictly periodic or stationary; otherwise, the resulting Fourier spectrum will make little physical sense. Unfortunately, the rolling bearing vibration signals are often non-stationary processes, and their frequency components will change with time. Therefore, the Fourier transform often cannot fulfil the bearing fault diagnosis task well. A technique which accounts for the variation of spectral components with time will be more effective here. Hence, one has to have recourse to time-frequency analysis methods. These generate both time and frequency information of a signal simultaneously through mapping the one-dimensional signal to a two-dimensional time–frequency plane. In fact, in recent studies, the time–frequency analysis methods are widely used to diagnose the faults in bearings. They can determine the time of the impact of occurrence as well as the frequency ranges of the impact location; In effect they identify the existence of faults as well as the causes. Amongst the available frequency analysis, the wavelet transforms stand out due to their unique features; they have been widely used for roller bearing fault detection [3-4].

For roller bearing fault detection, the frequency ranges of the vibration signals that are to be analysed, are often wide; and hence according to the Shannon sampling theorem[18], a high sampling rate is needed. Further the sequence of samples to be collected and processed for meaningful fault detection is quite large in size. Therefore, it is expected that the desired method should have good computing efficiency. Unfortunately, the computing of continuous wavelet transform (CWT) is somewhat time consuming and is not suitable for large size data analysis and on-line fault diagnosis which is of interest here.

The Discrete Wavelet Transform (DWT), which is based on sub-band coding can be used effectively to analyse the signal as long as the time duration of the wavelet is large compared to the sampling period. Note that this requirement is same as the sampling theorem (discussed above) put in a different form. In other words, we are not losing any substantial information in the process of resorting to DWT. The real benefit in the use of DWT is a conspicuous reduction in computation time. It is easy to implement and reduces the computation time and resources required. Hence, it is taken up for study.

J. Lin et. Al [3] reported the use of Morlet wavelet for decomposition and feature extraction from vibration signals from the rolling bearing and the gearbox. [3]. Only Morlet wavelet was used in the study. Goumas et al [4]. used discrete wavelet transform to analyse the transient signals of the vibration velocity in washing machines and fault
features were extracted from the wavelet coefficient [4]. In works reported in literature, researchers mostly have used energy level of each wavelet decomposition coefficient at each level as features. There is a mechanical energy analogy that they quote for justifying the usage. This measure gives an idea about the energy content of the wavelet coefficients. In the literature it is established that the information required for fault diagnosis problem is hidden in the wavelet decomposition coefficients. The challenge is, how one taps that information by using a measure?’. This is the fundamental motivation to look forward to a new measure which will perform better. As a result many measures were tried and the one which performed better alone is presented in this paper. This forms the primary contribution of the paper. The morlet wavelet was used because of its fault diagnosis capability [19-21].

2. EXPERIMENTAL SET UP

The fault simulator setup used for conducting the experiment is shown in Fig. 1. It can simulate faults in bearings, gears, belts etc. It is also designed to study misalignments and eccentricity.

A variable speed DC motor (0.5hp) with speed up to 3000 rpm is the drive. A short shaft of 30mm diameter is attached to the shaft of the motor through a flexible coupling; this is to minimize the effects of misalignment and transmission of vibration from motor. Shaft is supported at its ends through two taper roller bearings. The bearing that is closer to the motor is a good bearing. The bearing at the farther end is the bearing under test; provision is made to change it easily. Sensor is mounted on top of the bearing housing. The selected area is made flat and smooth to ensure effective coupling.

A piezoelectric accelerometer (Dytron) is mounted on the flat surface using direct adhesive mounting technique. The voltage output of accelerometers is proportional to acceleration. Accelerometers are preferred as transducers in machine condition monitoring due to the following reasons:

i) Extreme ruggedness,
ii) Large frequency response,
iii) Large dynamic range (accelerometers can detect very small vibrations without being damaged by large vibrations)
iv) Output is proportional to forces which are the cause of internal damage.
v) High-frequency sensitivity for detecting bearing faults.

The accelerometer is connected to the signal-conditioning unit, DACTRON FFT analyzer, where the signal goes through the charge amplifier and an Analogue-to-Digital converter (ADC). The vibration signal in digital form is input to the computer through an USB port. It is stored directly in the computer’s secondary memory. The signal is then read from the memory and processed to extract different features.
3. EXPERIMENTAL PROCEDURE

In the present study, four SKF30206 roller bearings were used. One was a brand new bearing and was taken to be free of defects. In the other three roller bearings, defects were created. Electron Discharge Machining (EDM) was used to ensure to get precisely defined defect. The size of inner race defect is 0.525 mm wide and 0.827 mm deep and that of outer race defect is 0.652 mm wide and 0.981 mm deep. The sizes of the defects are a bit larger than that one encountered in any practical situations; however, it is in-line with work reported in literature [13].

The vibration signal from the piezoelectric pickup mounted on the test bearing was taken, after allowing for initial running of the bearing. The sampling frequency was 12000 Hz and sample length was 8192 for all speeds and all conditions. The sample length is chosen long enough to ensure data consistency; In addition to data consistency, the following points were considered. Statistical measures are more meaningful, when the number of samples is sufficiently large. On the other hand, as the number of samples increases the computation time increases. To strike a balance, sample length of around 10000 was chosen. In techniques like wavelet based feature extraction, the number of samples is preferably $2^n$. The nearest $2^n$ to 10000 is 8192 and hence, it was taken as sample length. Extensive trials were taken at the set speed (700 rpm) and vibration signal was stored in the data file.
In line with the justification earlier (in section 1), four cases were considered –
i) Normal bearing (without any fault),
ii) Bearing with inner race defect,
iii) Bearing with outer race defect
iv) Bearing with both inner race and outer race defects. Time domain plots of these cases
are shown in Fig. 2.

4. WAVELET ANALYSIS AND FEATURE EXTRACTION
4.1 Brief review of wavelet theory
Wavelet Transform is defined as the integral of the signal \( s(t) \) multiplied by
scaled, shifted versions of a basic wavelet function \( \psi(t) \) —a real-valued function whose
Fourier Transform satisfies the admissibility criteria \[14\]:
\[
C(a,b) = \int_{\mathbb{R}} s(t) \frac{1}{\sqrt{a}} \psi \left( \frac{t-b}{a} \right) dt
\]
\[a \in \mathbb{R}^* - \{0\}, b \in \mathbb{R}.\]

where,
- \( C \) - Continuous wavelet coefficients
- \( t \) - time
- \( a \) - scaling parameter,
- \( b \) - time localisation parameter.
Both ‘\( a \)’ and ‘\( b \)’ can be continuous or discrete variables. Multiplication of each
coefficient by an appropriately scaled and shifted wavelet yields the constituent wavelets
of the original signal. For signals of finite energy, continuous wavelet synthesis provides
the reconstruction formula:
\[
s(t) = \frac{1}{K_{\psi}} \int_{\mathbb{R}} \int_{\mathbb{R}} C(a,b) \frac{1}{\sqrt{a}} \psi \left( \frac{t-b}{a} \right) da db
\]
\[\text{(2)}\]
Use of Discrete Wavelet Transform (DWT) for analysis retains effectiveness without sacrificing accuracy. In this scheme, ‘a’ and ‘b’ are given by:

\[(j,k) \in Z^2 : a = 2^j, b = k2^j, Z = \{0, \pm 1, \pm 2\}\]

Defining

\[(j,k) \in Z^2 : \psi_{j,k}(t) = 2^{-j/2} \psi(2^{-j}t - k)\]
\[\phi_{j,k}(t) = 2^{-j/2} \phi(2^{-j}t - k)\]

A wavelet filter with impulse \(g\), plays the role of the wavelet \(\psi\) and a scaling filter with impulse response \(h\), plays the role of scaling function \(\phi\). \(g\) and \(h\) are defined on a regular grid \(\Delta Z\), where \(\Delta\) is the sampling period (here, without loss of generality, set \(\Delta = 1\)). Discrete wavelet analysis yields

\[C(a,b) = c(j,k) = \sum_{n \in Z} s(n)g_{j,k}(n)\]  
\[a = 2^j, b = k2^j, j \in N, k \in N\]

Corresponding synthesis relation is:

\[s(t) = \sum_{j \in N} \sum_{k \in Z} c(j,k)\psi_{j,k}(t)\]

The detail at level \(j\) is defined as:

\[D_j(t) = \sum_{k \in Z} c(j,k)\psi_{j,k}(t)\]

And the approximation to \(s(t)\) at any level is:

\[A_{J-1} = \sum_{j > J} D_j\]

The following equations hold:

\[A_{J-1} = A_j + D_j\]  
\[s = A_j + \sum_{j > J} D_j\]

In practice, the decomposition can be determined iteratively, with successive approximations being computed in turn, so that a signal is decomposed into many lower-resolution components. This is known as the wavelet decomposition tree.

4.2 Wavelet analysis and feature definition

Figs. 3 to 6 show a combination of signals and their decomposition details at different levels designated ‘d1’ to ‘d5’. Actually, for analysis, thirteen levels are considered (from ‘d1’ to ‘d13’). A careful perusal of the signal details under different conditions bring out that there are considerable changes in the average energy level of some details with respect to its conditions. Feature extraction constitutes computation of specific measures which characterise the signal. The discrete wavelet transform (DWT) provides an effective method for generating features. There are many ways in which researchers defined the measures for DWT. However, amongst them the most common and popular definition is taken as bench mark for comparison. The bench mark ‘energy definition’ of a feature vector is given by
A component $v_{dwt}^{i}$ in the feature vector is related to the individual resolutions by the following equation

$$v_{dwt}^{i} = \frac{1}{n_i} \sum_{j=1}^{n_i} w_{i,j}^2, \quad i = 1, 2, \ldots, 12$$

where, $n_i = 2^{12}, n_2 = 2^{11}, \ldots, n_{12} = 2^0$.

$v_{dwt}^{i}$ is the $i^{th}$ feature element in a DWT feature vector. $n_i$ is the number of samples in an individual sub-band, $w_{i,j}^2$ is the $j^{th}$ coefficient of the $i^{th}$ sub-band.

The proposed feature vector is defined as follows:

$$v_{dwt} = \sum_{i=1}^{n} d_i^2$$

where, $i$ is an index, $n$ is number of data points in each wavelet decomposition coefficients at a given level and $d_i$ is the detail coefficient of $n^{th}$ decomposition. In the new definition, the co-efficients are squared and summed (sum squared definition). The other feature definitions which were considered for the study include mean of the detail coefficients, root mean square of the detail coefficients, standard error of the detail coefficients etc.

![Fig.3. DWT decomposition plot of details using DB2 up to 5 levels for signal of good bearing](image-url)
Fig. 4 DWT decomposition plot of details using DB2 up to 5 levels for signal of inner race fault bearing.

Fig. 5 DWT decomposition plot of details using DB2 up to 5 levels for signal of outer race fault bearing.
5. CLASSIFIER (C4.5 ALGORITHM)

Fault diagnosis can be viewed as a data mining problem where one extracts information from the acquired data through a classification process. A predictive model for classification invokes the idea of branches and trees identified through a logical process. The classification is done through a decision tree with its leaves representing the different conditions of the bearings. The sequential branching process ending up with the leaves here is based on conditional probabilities associated with individual features. Any good classifier should have the following properties [16]:

(1) It should have good predictive accuracy; It is the ability of the model to correctly predict the class label of new or previously unseen data.
(2) It should have good speed.
(3) The computational cost involved in generating and using the model should be as low as possible.
(4) It should be robust; Robustness is the ability of the model to make correct predictions given the noisy data or data with missing values.
(5) The level of understanding and insight that is provided by classification model should be high enough.

It is reported that C4.5 model introduced by J.R. Quinlan [17] satisfies with the above criteria and hence the same is used in the present study. Decision tree algorithm (C4.5) has two phases: building and pruning. The building phase is also called as ‘growing phase’. Both these are briefly discussed here.
5.1 Building Phase

Training sample set with discrete-valued attributes is recursively partitioned until all the records in a partition have the same class. This forms the building phase. The tree has a single root node for the entire training set. For every partition, a new node is added to the decision tree. For a set of samples in a partition $S$, a test attribute $X$ is selected for further partitioning the set into $S_1, S_2, ..., S_L$. New nodes for $S_1, S_2, ..., S_L$ are created and these are added to the decision tree as children of the node for $S$. Further, the node for $S$ is labeled with test $X$, and partitions $S_1, S_2, ..., S_L$ are recursively partitioned. A partition in which all the records have identical class label is not partitioned further, and the leaf corresponding to it is labeled with the corresponding class. The construction of decision tree depends very much on how a test attribute $X$ is selected. C4.5 uses information entropy evaluation function as the selection criteria [17]. The entropy evaluation function is arrived at through the following steps.

Step 1: Calculate $Info(S)$ to identify the class in the training set $S$.

$$Info(S) = -\sum_{i=1}^{K} \left[ \frac{freq(C_i, S / |S|)}{|S|} \right] \log_2 \left[ \frac{freq(C_i, S / |S|)}{|S|} \right]$$

(12)

where, $|S|$ is the number of cases in the training set. $Ci$ is a class, $i = 1, 2, ..., K$. $K$ is the number of classes and $freq(Ci, S)$ is the number of cases included in $Ci$.

Step 2: Calculate the expected information value, $Info_x(S)$ for test $X$ to partition $S$.

$$Info_x(S) = -\sum_{i=1}^{K} \left[ \frac{|S_i|/|S|} Info(S_i) \right]$$

(13)

where $L$ is the number of outputs for test $X$, $S_i$ is a subset of $S$ corresponding to $i^{th}$ output and is the number of cases of subset $S_i$.

Step 3: Calculate the information gain after partition according to test $X$.

$$Gain(X) = Info(S) - Info_x(S)$$

(14)

Step 4: Calculate the partition information value $SplitInfo(X)$ acquiring for $S$ partitioned into $L$ subsets.

$$SplitInfo(X) = -\frac{1}{2} \sum_{i=1}^{L} \left[ \frac{|S_i|}{|S|} \log_2 \left( \frac{|S_i|}{|S|} \right) + \left( 1 - \frac{|S_i|}{|S|} \right) \log_2 \left( 1 - \frac{|S_i|}{|S|} \right) \right]$$

(15)

Step 5: Calculate the gain ratio of $Gain(X)$ over $SplitInfo(X)$.

$$GainRatio(X) = Gain(X) - SplitInfo(X)$$

(16)

The $GainRatio(X)$ compensates for the weak point of $Gain(X)$ which represents the quantity of information provided by $X$ in the training set. Therefore, an attribute with the highest $GainRatio(X)$ is taken as the root of the decision tree.
5.1 Pruning phase

Usually a training set in the sample space leads to a decision tree which may be too large to be an accurate model; this is due to over-training or over-fitting. Such a fully grown decision tree needs to be pruned by removing the less reliable branches to obtain better classification performance over the whole instance space even though it may have a higher error over the training set.

The C4.5 algorithm uses an error-based post-pruning strategy to deal with over-training problem. For each classification node C4.5 calculates a kind of predicted error rate based on the total aggregate of misclassifications at that particular node. The error-based pruning technique essentially reduces to the replacement of vast sub-trees in the classification structure by singleton nodes or simple branch collections if these actions contribute to a drop in the overall error rate of the root node.

5.3 Application of Decision Tree for the Problem under Study

As is customary the samples are divided into two parts:

i) Training set

ii) Testing set.

Training set is used to train classifier and testing set is used to test the validity of the classifier. Ten-fold cross-validation is employed to evaluate classification accuracy. The training process of C4.5 using the samples with continuous-valued attributes is as follows.

1. The tree starts as a single node representing the training samples.
2. If the samples are all of the same class, then the node becomes a leaf and is labeled with the class.
3. Otherwise, the algorithm discretises every attribute to select the optimal threshold and uses the entropy-based measure called information gain (discussed in Section 5.1) as heuristic for selecting the attribute that will best separate the samples into individual classes.
4. A branch is created for each best discrete interval of the test attribute, and the samples are partitioned accordingly.
5. The algorithm uses the same process recursively to form a decision tree for the samples at each partition.
6. The recursive partitioning stops only when one of the following conditions is true:
   a. All the samples for a given node belong to the same class or
   b. There are no remaining attributes on which the samples may be further partitioned.
   c. There are no samples for the branch test attribute. In this case, a leaf is created with the majority class in samples.
7. A pessimistic error pruning method (discussed in Section 2.2) is used to prune the grown tree to improve its robustness and accuracy.

6. Results and Discussions

The fault diagnosis problem is modeled as machine learning problem in which wavelet decomposition co-efficients are used for feature extraction. In defining the features from wavelet decomposition co-efficients, a new measure (root square
definition) is proposed and the corresponding feature vectors are extracted. Refer eqn.(11) for feature definition; The \( V_1, V_2, \ldots V_n \) represents the feature defined at nth level. The features are classified with C4.5 decision tree algorithm. The decision tree built out of new features is shown in Fig. 7. A similar procedure is followed for already existing benchmark definition (energy definition) and the corresponding decision tree is shown in Fig. 8. In Fig. 7, the number of leaves is 8 and the size of the tree is 15 for new feature. The corresponding values of existing features are 9 and 16 as shown in Fig. 8. One can appreciate that if the size and number of leaves are less, it clearly shows the higher strength of the features. If the strength of the features is more, it will have less entropy and can classify the output data points easily leading to small tree. The vice versa is also true.

Fig. 7 Decision tree with energy measure

Fig. 8 Decision tree with new measure
The detailed accuracy like true positives (TP), False positives (FP), precision, recall, and F-measure for existing features is presented in Table 1 and that of proposed feature is presented in Table 3. From Table 1 and Table 3, one can understand that the proposed features are better by noting the TP rate values, Precision, Recall and F-measure, which are supposed to be high for a better feature set. These parameters are higher compared to that of existing features. Also the result is supplemented by the FP rate values which supposed to be less for a better feature and it confirms the above discussion.

Table 1 Detailed accuracy by class for existing feature

<table>
<thead>
<tr>
<th>Class</th>
<th>TP Rate</th>
<th>FP Rate</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>0.98</td>
<td>0.003</td>
<td>0.99</td>
<td>0.98</td>
<td>0.985</td>
</tr>
<tr>
<td>Irf</td>
<td>0.87</td>
<td>0.053</td>
<td>0.845</td>
<td>0.87</td>
<td>0.857</td>
</tr>
<tr>
<td>Orf</td>
<td>0.85</td>
<td>0.04</td>
<td>0.876</td>
<td>0.85</td>
<td>0.863</td>
</tr>
<tr>
<td>Iorf</td>
<td>0.99</td>
<td>0.007</td>
<td>0.98</td>
<td>0.99</td>
<td>0.985</td>
</tr>
</tbody>
</table>

Table 2 Confusion matrix for existing features

<table>
<thead>
<tr>
<th></th>
<th>Good</th>
<th>Irf</th>
<th>Orf</th>
<th>Iorf</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>98</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Irf</td>
<td>1</td>
<td>87</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>Orf</td>
<td>0</td>
<td>13</td>
<td>85</td>
<td>2</td>
</tr>
<tr>
<td>Iorf</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>99</td>
</tr>
</tbody>
</table>

Table 3 Detailed accuracy by class for proposed features

<table>
<thead>
<tr>
<th>Class</th>
<th>TP Rate</th>
<th>FP Rate</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>0.98</td>
<td>0.003</td>
<td>0.99</td>
<td>0.98</td>
<td>0.985</td>
</tr>
<tr>
<td>Irf</td>
<td>0.90</td>
<td>0.047</td>
<td>0.865</td>
<td>0.90</td>
<td>0.882</td>
</tr>
<tr>
<td>Orf</td>
<td>0.87</td>
<td>0.03</td>
<td>0.906</td>
<td>0.87</td>
<td>0.888</td>
</tr>
<tr>
<td>Iorf</td>
<td>0.99</td>
<td>0.007</td>
<td>0.98</td>
<td>0.99</td>
<td>0.985</td>
</tr>
</tbody>
</table>

Table 4 Confusion matrix for proposed features

<table>
<thead>
<tr>
<th></th>
<th>Good</th>
<th>Irf</th>
<th>Orf</th>
<th>Iorf</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>98</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Irf</td>
<td>1</td>
<td>90</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>Orf</td>
<td>0</td>
<td>11</td>
<td>87</td>
<td>2</td>
</tr>
<tr>
<td>Iorf</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>99</td>
</tr>
</tbody>
</table>

The misclassification details are presented in the form of confusion matrix in Table 2 and Table 4. Before going to the discussion of the confusion matrix the way of interpretation of confusion matrix is presented here. Referring to Table 2, the first row of the confusion matrix represents the good condition. The first element in first row, i.e.
location (1, 1), 98 represents the number of data points that belong to the good condition and have been classified correctly as ‘good’. The second element in the first row i.e. location (1, 2), ‘2’ depicts as to how many of the good condition have been misclassified as ‘Irf’. The third element represents the number of data points that has been misclassified as ‘Orf’. The fourth element in the first row i.e. location (1, 4), ‘0’ depicts as to how many of the good condition have been misclassified as ‘Iorf’. Similarly the second row represents the ‘Irf’ condition. The second element in second row represents the correctly classified instances for ‘Irf’ condition and rest of them are misclassified details as explained earlier. Similar interpretation can be given for other elements as well.

To summarize, the diagonal elements shown in the confusion matrix represents the correctly classified points and non-diagonal elements are misclassified ones. Comparing element by element wise in the confusion matrix shown in Table 2 and Table 4, one will appreciate that the proposed feature set is performing well or equal to that of existing feature set. It is to be emphasised that in none of the instances, the proposed features performed inferior to that existing features. The total classification accuracy of existing feature set is 92.5% while that of proposed feature set is 93.5%. The difference is one percentage only; however, it is an increment from 92.5% to 93.5%. In that context, it is a good improvement and can be used for fault diagnosis applications.

7. CONCLUSION

A comparative study of bearing faults and their classification involving a new wavelet feature set and an existing wavelet energy feature set has been carried out leading to some interesting conclusions regarding classification and fault identification. Feature definition based on the root square and average energy level of signals is central to the study here. The results show that the proposed feature set performs better than the existing feature set. This proposed feature set is much suited for practical fault diagnosis of roller bearings compared to energy definition.

8. REFERENCE


