International Journal of Manufacturing and Material Processing
Vol. 1: Issue 2
www.journalspub.com

A Comprehensive Study of Fault Diagnostics of Roller Bearings Using Continuous Wavelet Transform

V. Sugumaran*[1], K. I. Ramachandran[2], Aditya Vasudev Rao[1]
[1]SMBS, VIT University, Chennai Campus, Vandalur-Kelambakkam Road, Chennai, India
[2]Amrita School of Engineering, Ettimadai, Coimbatore, India

Abstract

One of the prominent causes of breakdown of rotating machinery, bearing failure, proves to be costly to the industry. Unexpected machine stoppage disrupts production schedule and maintenance strategies. Therefore, condition monitoring has become a very important research field in recent years. It can be used to avoid unexpected failures of critical systems. Bearing fault diagnostics using vibration signals has proven to be very effective. This paper deals with the fault detection and classification of roller bearings using Continuous Wavelet Transform. This paper is the result of an exhaustive study of different families wavelets and the determination of the best wavelet for the purpose. A decision tree is constructed using the C4.5 algorithm to determine the best wavelet based on classification accuracy. Once the wavelet was selected, those features were used in two classifiers viz. Support Vector Machines (SVM) and Proximal Support Vector Machines (PSVM) to determine classification accuracy. The test results showed that 'rbio1.5' gives maximum classification accuracy among wavelets and PSVM gave 100% classification accuracy for all of the bearing faults thus proving to be most suitable for practical applications.

Keywords: Roller bearings, continuous wavelet transform, statistical features, histogram features, decision tree, support vector machines, proximal support vector machines

Corresponding Author
Email ID: v_sugu@yahoo.com

INTRODUCTION

Bearings are the most common components found in rotating machinery. Rolling element bearings are widely used not only in heavy duty industrial machines but also in the most elementary machines. This wide range of applications makes bearings one of the most important components.

The dynamic nature of operation of roller bearings causes wear and tear of the integral parts and eventually leads to failure. Bearing failure is one of the most frequent causes of industrial machinery breakdown. Such unpredicted breakdown can cause unnecessary delays and thus, can adversely affect production. To avoid this problematic scenario, it is crucial that the condition of the bearing is continuously monitored and faults (if any) should be detected prior to failure. Timely diagnosis can greatly help maintenance strategies and significantly reduce the risk of sudden failure of machinery.

Many techniques are available for bearing fault diagnostics. Amongst them,Vibration and Acoustic Emission signals are widely used because they provide important information of internal faults.

Vibration signals need to be measured at several carefully selected points and the

signals need to be analyzed to study the basic components which makeup the complex raw waveform. The relevant features are extracted from the time domain signal and the various conditions of the bearings are classified based on these extracted features. Such classification cannot be carried out by a direct observation of the visual waveform.

Transformation into a different domain using a suitable transform can be of helpful here. Wavelet transform is a time-frequency signal analysis method which has become very popular in recent years; it is widely used for this purpose. It has the local characteristic of time-domain as well as frequency domain and its time-frequency window size can be varied. In the processing of non-stationary signals, it presents superior performance. In the present study, wavelet analysis is performed to extract features from the time domain signals. There are many types of wavelet transforms which can be used for extracting the features. For example, Continuous Wavelet Transform, Discrete Wavelet Transform (DWT), Stationary Wavelet Transform (SWT), *etc*. Wavelets have been used for condition monitoring in several studies.

S Prabhakar *et al.* showed that DWT can be effectively used for bearing fault detection[1]. Xinsheng Lou et.al. used DWT to generate feature vectors from accelerometer signals of ball bearings. They further used adaptive neural-fuzzy system to perform classification. [2]

V. Sugumaran *et al.* selected features using Decision Tree and classified them through Proximal Support Vector Machine for fault diagnostics of roller bearing. [3] S. Abbasion *et al.* performed denoising of roller bearing vibration signal using Meyer Wavelet and used Support Vector Machine for classification and achieved 100% accuracy. [4] Gang Yu *et al*. used cluster-based wavelet feature extraction. The wavelet coefficients are grouped into clusters using a set of representative signals that pertain to a given diagnostic application. Based on the fact that multi-layer perceptron (MLP) neural networks have slow training speed and limited potential in real world applications, they used a Probabilistic Neural Network (PNN). PNN is robust and has a very fast training speed. [5] Youzhong Li developed an incipient bearing fault diagnosis technique using hybrid wavelet and neural networks.

Noise in the vibration signal was eliminated using adaptive wavelet de-noising, then wavelet packet transform is applied and energy features are extracted. The faults were identified effectively by using a combination of wavelet theory and Artificial Neural Network (ANN) to obtain a specialized neural network. [6] Dong Wang *et al.* proposed an improved combination of Hilbert and Wavelet Transforms for extraction of fault signatures of rolling element bearings. Analysis results showed that fault signatures' extraction capability was greatly enhanced by the method. [7] Wensheng Su *et al.* presented a new hybrid method for fault diagnosis of rolling element bearings. The vibration signal was filtered with a band-pass filter determined by Morlet wavelet which is optimized by genetic algorithm. Further, they applied an auto correlation enhancement algorithm to the filtered signal to further reduce the residual in-band noise and highlight the periodic impulsive feature[8]. J Chebil *et al.* established that the power and flexibility of DWT can be enhanced by using Discrete Wavelet Packet Transform (DWPT).

They found that choice of mother wavelet Sym6 combined with Retail Management System (RMS) feature and Bayesian classifier produce excellent classification results of up to 100% for some of the

faults in rolling element bearings. [9] P.K. Kankar *et al.* performed rolling element bearing fault diagnostics using Continuous Wavelet Transform. They used Complex Morlet Wavelet and Meyer wavelet and performed classifications using Artificial Neural Network (ANN), Support Vector Machine (SVM) and Self Organizing Maps (SOM). They found SVM was most accurate for both wavelet features.

They also found that features selected using Meyer wavelet gives higher faults classification accuracy with SVM classifier. [10] N. Sarvanann *et al.* performed incipient gear box fault diagnosis using Discrete Wavelet Transform (DWT) for feature extraction. Debauchies wavelet was used as a mother wavelet and features were classified using Artificial Neural Network (ANN). [11] P. Konar *et al.* used Continuous Wavelet Transform (CWT) to decompose vibration signals and Morlet and Debauchies were used as Mother wavelets. The wavelet coefficients thus obtained were analysed by SVM and ANN as a comparative study. [12] Hui Li *et al.* used the Hermitian Wavelet amplitude and phase map to detect and diagnose rolling element bearing fault.

The Fourier spectrum of the Hermitian Wavelet is real and the Hermitian wavelet does not affect the phase of a signal in complex domain.

This gives the ability to detect the singularity characteristic of a given signal accurately. [13] Kung Feng *et al.* constructed a new wavelet filter called Anti-Symmetric Real Laplace Wavelet (ARLW) for fault diagnostics of rolling element bearing and used Differential Evolution algorithm for feature selection. [14] Derek Kanneg *et al.* proposed the Wavelet Spectrum (WS) technique for representative feature extraction and bearing incipient fault detection. The effectiveness of the proposed technique is verified by a series of experiments and the test results prove it to be effective[15]. Houcine Bendjama *et al.* tested CWT and Discrete Wavelet Transform(DWT) on real measurement signals collected from vibration system and proved that Wavelet Transform is an effective technique in indentifying type of fault in rotating machinery. [16] Kang Chen *et al.* used one dimensional DWT to decompose the bearing fault signal into multi-layer. Then the decomposed signal is analyzed by Hilbert envelope and spectrum analysis technique to detect the faults. [17] Peng Li *et al*. extracted multi-scale slope features of gearbox and bearing using DWT and achieved high levels of accuracy. [18]

Changqing Shen *et al.* performed fault diagnosis of rotating machinery by using Wavelet Packet Transform (WPT) and a Support Vector Regressive Classifier. They used Distance Evaluation Technique (DET) to reduce the dimensionality of feature space. [19]

Muralidharan *et al.* performed a comparative study of Naive Bayes classifier and Bayes net classifier for fault diagnosis of a monoblock centrifugal pump using wavelet analysis. In their study they simulated cavitation, impeller fault, bearing fault and Bearing and Impeller fault together. They computed the DWT for different conditions of the pump to extract features. [20] V. Muralidharan *et al*. acquired signals of a mono block centrifugal pump, used Continuous Wavelet transform (CWT) for feature extraction and performed classification using Decision Tree (J48) algorithm. [21]

Kumar H.S. *et al.* performed DWT of the bearing vibration signals and used the statistical features extracted from dominant wavelet coefficients as inputs to ANN classifier to evaluate its performance. [22] Lingjie Meng *et al.*

found a solution to alleviate the pseudo-Gibbs phenomenon, occurring in traditional wavelet methods by using a translation invariant wavelet. They used a morphological filter to eliminate narrowband impulses and random noises to a certain extent. They used ensemble empirical mode decomposition (EEMD) to decompose the signals and analyzed using intrinsic mode functions (IMFs). [23].

Thus, it can be noted that though several techniques have been used in wavelet analysis, a comprehensive study is needed. Most of the research work included extracting wavelet coefficients using different family of wavelets or usage of different classifiers. In some cases, the choice of wavelet was predetermined and in others the selection was based on calculation of classification accuracies of a few wavelets.

However, a more exhaustive approach was required for such a diverse field: consisting of many possibilities. This work includes a total of 54 wavelets, considering all scale variations, to determine exactly the best wavelet which gives maximum classification accuracy in fault diagnostics of rolling element bearings. C4.5 algorithm is used to obtain the wavelet based on classification accuracy.

The features extracted include statistical features and histogram features. C4.5 algorithm is then used to obtain the best set of features for classification. This is followed by a comparison between Support Vector Machines (SVM) and Proximal Support Vector Machines (PSVM) to determine the superior technique for the chosen wavelet. This paper will enable people working in this field to directly go ahead with the exact best wavelet, who would otherwise experiment with different wavelets to achieve most accurate classification.

CONTINUOUS WAVELET TRANSFORM

The CWT uses a set of non-orthogonal wavelet frames to provide highly redundant information which is excellent for detection of various types of faults. Wavelet coefficient at each analysis scale can be obtained, allowing us to characterize the local information content. CWT is gives more readability and allows more ease of interpretation. [9]

Continuous Wavelet Transform is defined as the sum over all time of the signal multiplied by scaled, shifted versions of the wavelet function Ψ. [11]

$$L_{\psi f(s,\tau)} = \int f(t)\Psi^*_{s,\tau}(t)dt \qquad \text{Eq. (1)}$$

There is decomposition of *f(t)* into a set of basis function $\psi_{s,\tau}(t)$, called wavelets generated from a single basic wavelet $\psi(t)$, the mother wavelet , by scaling and translation:

$$\Psi_{(s,\tau)}(t) = \frac{1}{\sqrt{|s|}}\psi\left(\frac{t-\tau}{s}\right) \qquad \text{Eq. (2)}$$

Here, s is a scale factor and τ is the translation factor. The factor $\frac{1}{\sqrt{|s|}}$ is for energy normalisation across the different scales.

FEATURE DEFINITION

Typical plots of Continuous Wavelet Transform (CWT) are shown in Figure 1−4. The colour scale is shown on the right side of the plots.

The plots show the coefficients at different scales. Observing all the four plots, there is a variation in wavelet coefficients for different conditions of the bearings. However, if the signals of a particular condition (say, 'good') were taken at different times, there is a phase shift in the variation of wavelet coefficients in a particular scale (tracing horizontally in the plot). To quantify the variation, a few measures are to be computed that can be used as features for fault diagnosis.

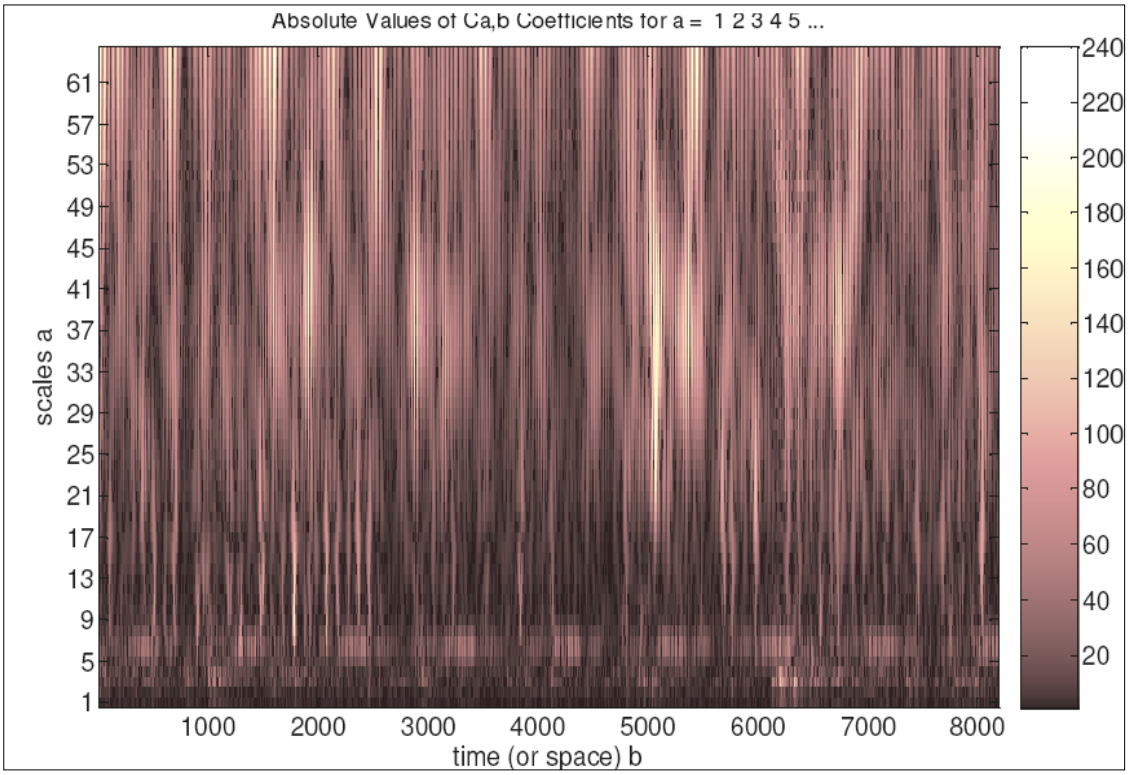

Fig. 1: *CWT Plot of 'Good' Condition with Morlet Wavelet.*

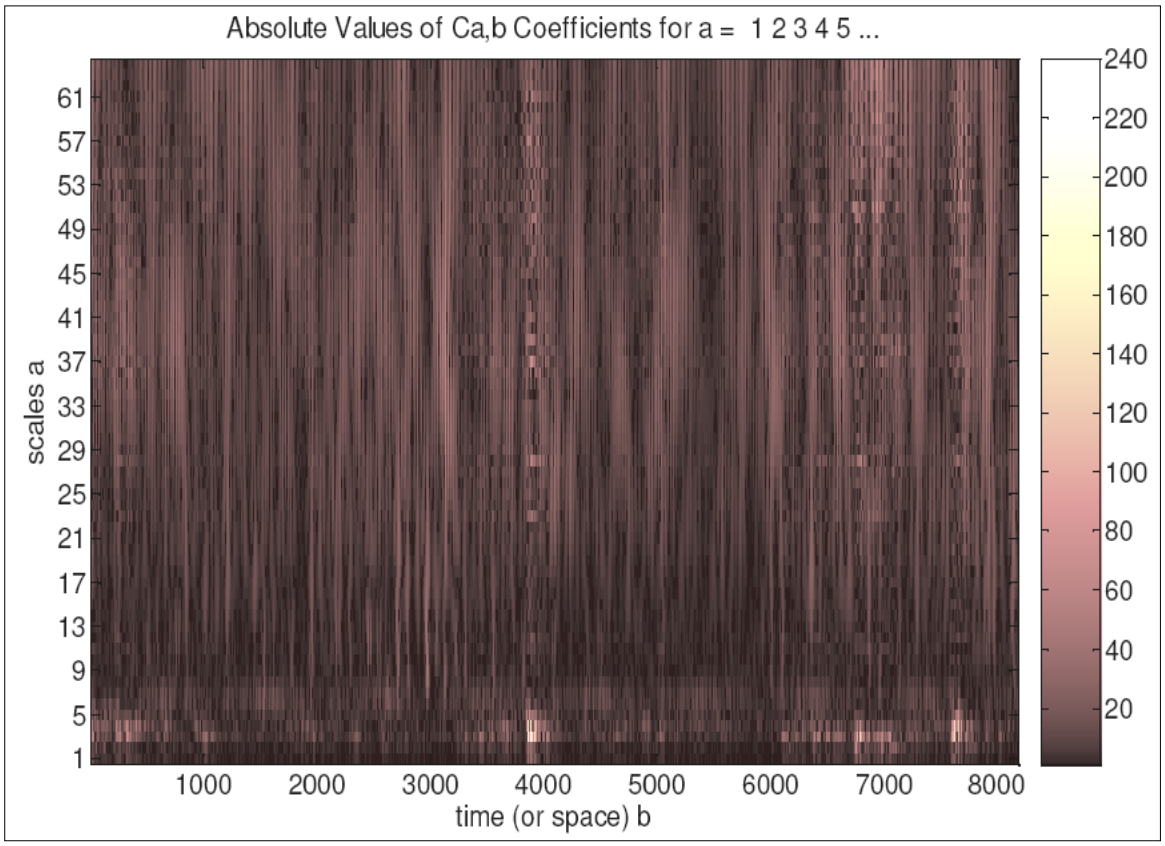

Fig. 2: *CWT Plot of 'IRF' Condition with Morlet Wavelet.*

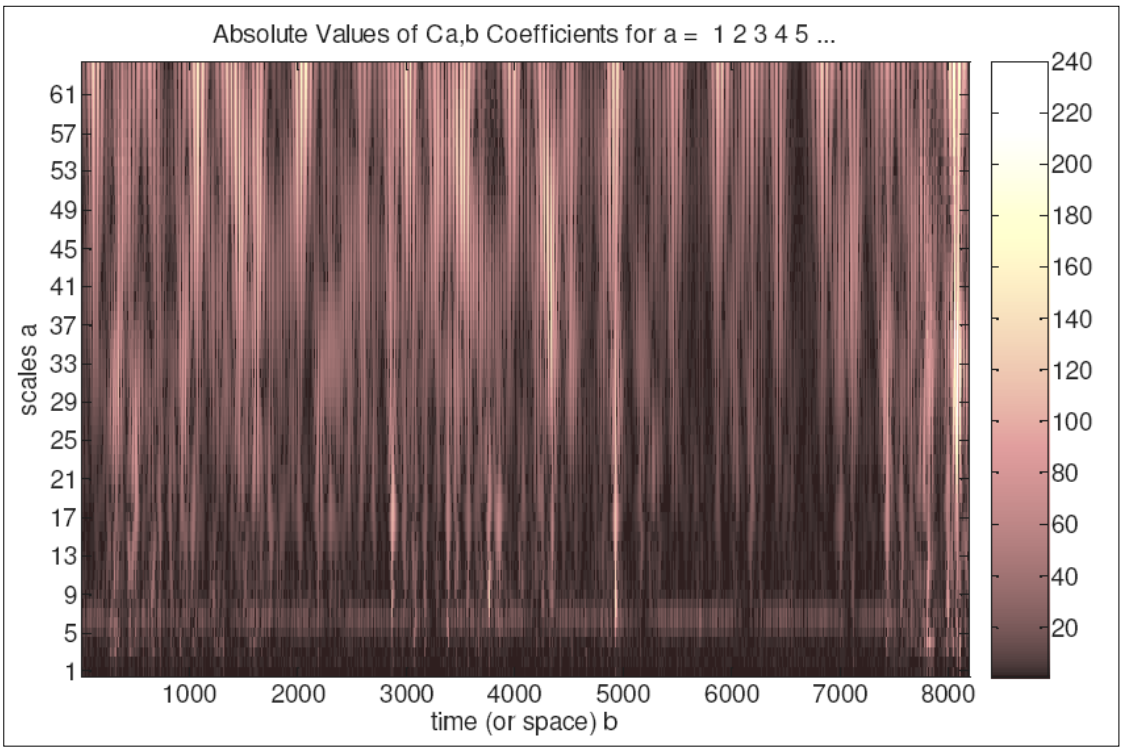

Fig. 3: *CWT Plot of 'ORF' Condition with Morlet Wavelet.*

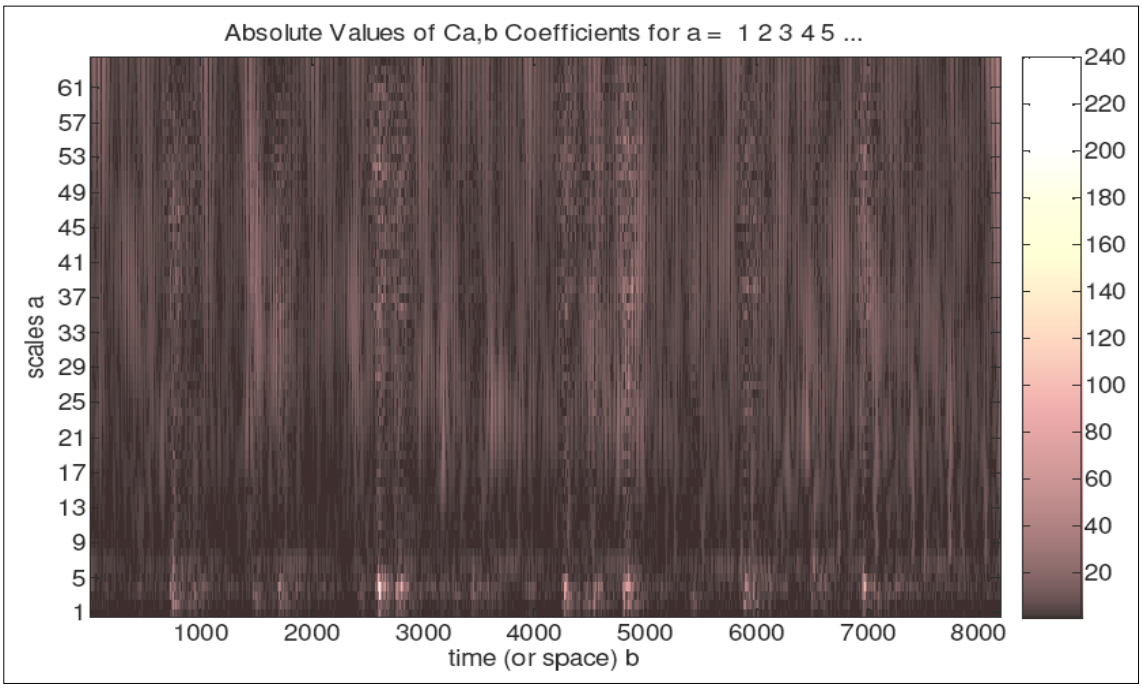

Fig. 4: *CWT Plot of 'IORF' Condition with Morlet Wavelet.*

The experimental setup and data extraction methodology is the one described in reference. [3] In the present study, statistical and histogram measures of continuous wavelet transform coefficients were used. The statistical features are mean, standard error, median, standard deviation, sample variance, kurtosis, skewness, range, minimum, maximum and sum. The total range of continuous wavelet coefficients is divided into twenty sub-ranges and they are used as bin for plotting histogram. The number of data points falling in each sub-ranges

are used as features and they are named as *'h1', 'h2' ... 'h20'*. Initially, the statistical and histogram features were used individually and classification accuracies were found using C4.5 algorithm. It was found that in some cases the statistical features gave higher accuracy while with others histogram features gave higher accuracy. In order to take advantage of both these feature sets, the statistical and histogram features were combined and used in the present study.

DECISION TREE

It is a tree based knowledge representation methodology used to represent classification rules. [3] Main purpose of decision tree is to represent structural information. Decision trees are built recursively, following a top-down approach. [21] A standard tree induced with C4.5 algorithm consists of a number of branches, one root and a number of leaves and nodes. A chain of nodes from root to leaves forms one branch and each node involves one attribute. The occurrence of an attribute in a tree provides information regarding the importance of the associated attribute. At each node, the criterion used to identify the best feature uses the principle of information gain and entropy reduction.

Information gain is the expected reduction in entropy caused by portioning sample according to a particular feature. Information gain *(S, A)* of a feature *A* relative to a collection of samples *S*, is defined as[3],

$$Gain(S,A) = Entropy(S) - \sum_{v \epsilon Value(A)} \frac{|S_v|}{|S|} Entropy(S_v), \quad \text{Eq. (3)}$$

where *Value(A)* is the set of all possible values for attribute *A*, S_v is the subset of *S* for which feature *A* has value v (i.e., S_v={sϵS |A(s) = v}).

Entropy is a measure of homogeneity of a given sample, expressed as,

$$Entropy(S) = \sum_{i=1}^{c} -P_i \log_2 P_i \quad \text{Eq. (4)}$$

Here, *c* is the number of classes, p_i is the proportion of *S* belonging to class *'i'*.

FEATURE SELECTION USING C4.5 ALGORITHM

A typical decision tree using C4.5 algorithm is shown Figure 5. CWT features described in section 3 form the input to the algorithm. In Figure 5, $std\_dev, h1, h3,$ and $h18$ in the ovals are the features. Each rectangle represent a class.

The numbers within the rectangles, but outside parenthesis represent class labels; the numbers 1, 2, 3 and 4 represent good bearings, bearings with inner race faults, bearings with outer race faults, bearings with both inner and outer race faults, respectively. Two additional numbers are present within the parenthesis. The first of these indicates the number of data points that can be classified using that feature set and the second number represents misfits. The absence of the second number indicates the absence of misfits. Further, if the first number is small compared to the total number of samples, the corresponding features can be considered as outliers and hence ignored. The following points are important regarding the feature selection using decision tree:

1. The features appearing in the nodes of the decision tree are in the descending order of importance. Features that have less discriminating capability can be consciously discarded by deciding on the threshold. This concept is used in selecting good features by the algorithm.
2. Features, which have good discriminating capability alone, will eventually appear in the decision tree.

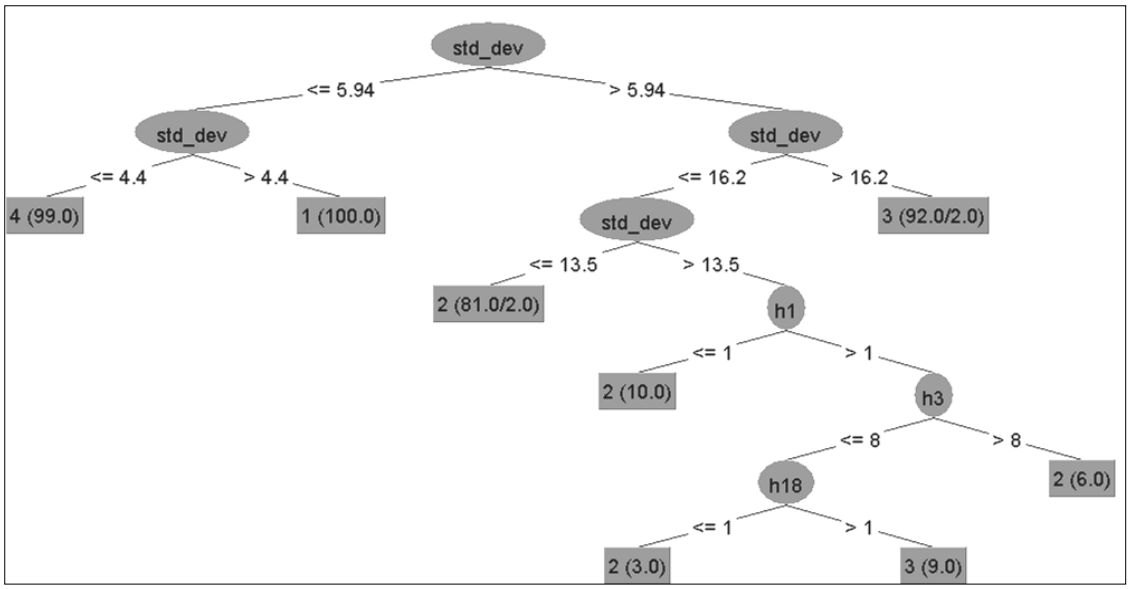

Fig. 5: *Decision Tree with CWT Features.*

Thus in the present case, out of the thirty features started with, only four features - namely $std\_dev, h1, h3,$ and $h18$ have been selected. Figure 5 shows the decision tree built out of these four features. A similar feature selection is carried out for all the wavelets used in the study for all levels.

WAVELET SELECTION USING C4.5 ALGORITHM

The wavelets considered in the present study are Haar wavelet, Coiflet wavelet, Daubechies wavelets, bi-orthogonal wavelet, reverse bi-orthogonal wavelet, Symlet wavelet, Meyer wavelet, Gaussian wavelet, Mexican hat wavelet.. Except Meyer and Haar wavelets, all other wavelets have child wavelets. Taking all these variations into account, 54 wavelets are possible. Now, the task is to find which wavelet amongst them suits well for fault diagnosis of roller bearings. As features are supposed to contain the required information available in the signal and C4.5 algorithm works based on the information gain, it is used for identifying the right choice of wavelet. This can be done by computing classification accuracy of each wavelet; one that has high accuracy is selected as a good wavelet. The classification process is carried out with the selected wavelet and the results are presented in section 7.

The classification accuracies of different wavelets using decision tree (C4.5 algorithm) are shown in Figure 6. In fact, for selecting the right wavelet any classifier can be used. However, only C4.5 algorithm was used here and the results are presented. Each bar in the Figure 6 shows the maximum accuracy in that particular wavelet family. The name of the wavelet is also marked along with its classification accuracy.

The classification accuracy of individual wavelets in a wavelet family using C4.5 algorithm is presented in the form of plots (refer Figures 7, 12, 15, 21, and 23). From the Figure 6, one can note '*rbio1.5*' gives the maximum classification accuracy (97.75%). Having selected the wavelet that gives the highest classification accuracy ('*rbio1.5*'), for the rest of the classifiers, these features are used; results are presented in section 7.

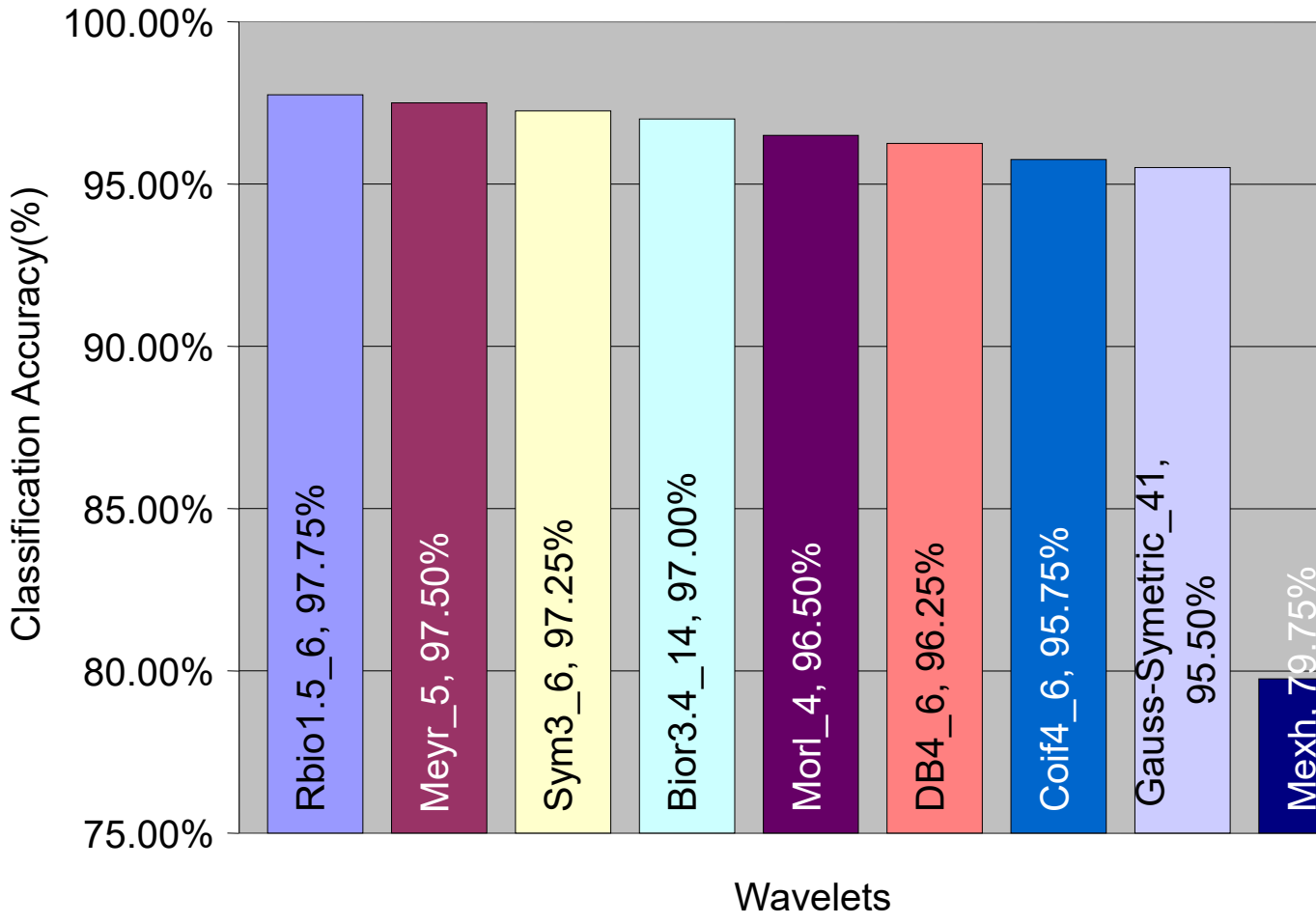

***Fig. 6**: Classification Accuracies of Different Wavelets Using C4.5.*

The classification accuracy comparison of different wavelets in reverse bi-orthogonal wavelet family is presented in Figure 7. The decision tree of CWT features of '*rbio1.5*' and the corresponding confusion matrix are in Figure 8 and 9 respectively.

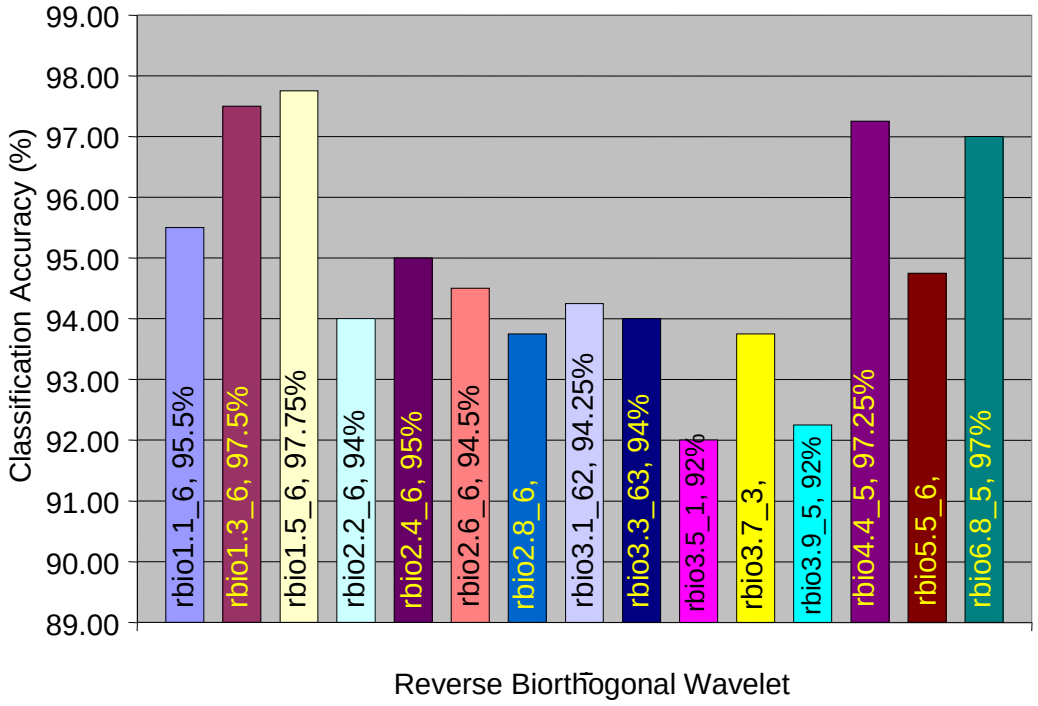

Fig. 7: *Comparison of Reverse Bi-orthogonal Wavelets CWT Features.*

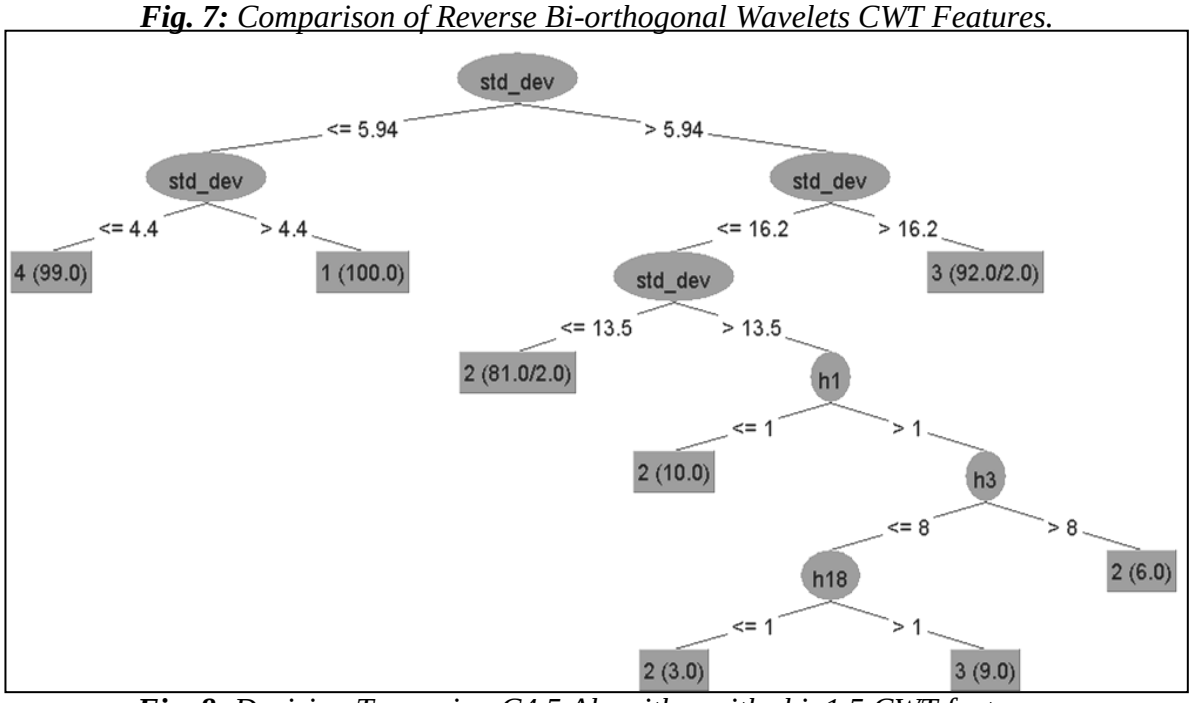

Fig. 8: *Decision Tree using C4.5 Algorithm with rbio1.5 CWT features.*

Observing the decision tree as represented in Figure 8, one can understand that the standard deviation (std_dev) of CWT coefficients plays a important role in classification as it appears in the top most node of the decision tree.

On the contrary, the role of feature '*h18*' is low amongst the four features, as it is used to classify only twelve samples. In other words, if the feature '*h18*'is not used, then the two samples out of twelve samples would have been misclassified.

From the confusion matrix depicted in Figure 9 one can observe that only one instance out of 400 samples is misclassified as 'good' (good bearings) although they were belonging to 'IORF' faulty category (inner and outer race fault) and another one sample is misclassified as 'IRF' while the actual condition of the bearing is 'Good'.

There are totally six misclassifications amongst 'IRF' and 'ORF'.

	Good	IRF	ORF	IORF
Good	99	1	0	0
IRF	0	97	3	0
ORF	0	3	97	0
IORF	1	1	0	98

Fig. 9: *Confusion Matrix of C4.5 Agorithm with rbio1.5 CWT Features.*

The second highest accuracy is given by 'meyr' wavelet at scale 5. The decision tree and the confusion matrix of 'meyr' are shown in Figure 10 and 11 respectively. The decision tree represented in Figure 10 is small and uses only two features. It reveals that those features are more robust features. From the confusion matrix depicted in Figure 11, one can observe that none of the 'Good' conditions is misclassified while only one sample is misclassified as 'Good' whose original condition is 'IORF'. The misclassifications amongst 'IRF' and 'ORF' bring down the overall classification accuracy to 97.5%.

Symlet wavelet with scale 6 gives next highest accuracy (97.25%). The classification accuracies of the symlet family wavelets are presented in Figure 12. It shows that symlet3 gives high accuracy (97.25%).

The decision tree and confusion matrix are shown in Figure 13 and 14 respectively. One can note from decision tree of symlet shown in Figure 13 that again standard deviation (std_dev) performs very well as a features compared to the other three in the decision tree.

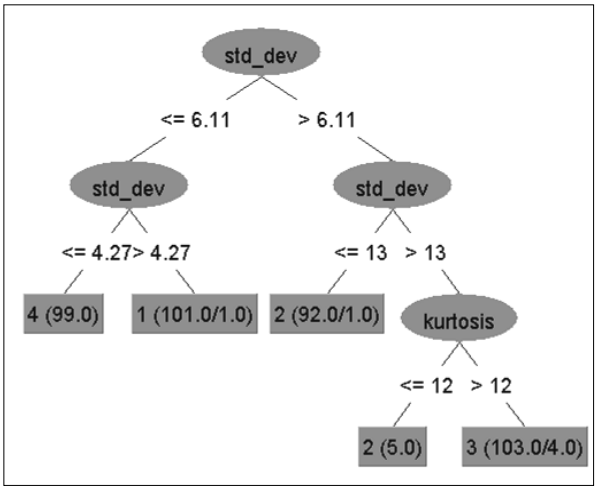

Fig. 10: *Decision Tree using C4.5 Algorithm with Meyr (Scale 5) CWT Features.*

	Good	IRF	ORF	IORF
Good	100	0	0	0
IRF	0	94	6	0
ORF	0	3	97	0
IORF	1	1	0	98

Fig. 11: *Confusion Matrix of C4.5 Algorithm with Meyr (scale 5) CWT Features.*

From the confusion matrix represented in Figure 14, one can observe that only one sample of 'Good' condition is misclassified as 'IRF' and one sample is misclassified as 'Good' whose original condition is 'IORF'.

Again here also, the misclassifications amongst 'IRF' and 'ORF' bring down the

overall classification accuracy to 97.25%. There is no case where the sample is misclassified as 'IORF'. This indicates that the class 'IORF' is defined well and the CWT features are able to identify them consistently from other classes. This is also true for the wavelets already discussed.

The bi-orthogonal wavelets with a classification accuracy of 97% come next. The performance of its family is represented in Figure 15. The highest accuracy (97%) in the family is given by two child wavelets; one '*bior3.4*' wavelet at the scale of 14 and 'bior3.9' at the scale of 5. The decision tree of '*bior3.4*' and '*bior3.9*' are shown in Figure 16 and 18 respectively.

The confusion matrix of '*bior3.4*' and '*bior3.9*' are shown in Figure 17 and 19 respectively. Both are equally good, as the number of branches and number of levels is same. In Figure 17, none of the samples are misclassified as 'IORF' while in Figure 19, there is one sample which is misclassified as 'IORF.

However, this is not very serious compared to misclassifying faulty bearing as 'Good'. Following this thumb rule, '*bior3.9*' is better compared to '*bior3.4*'. The reason for this is '*bior3.4*' has 3 such misclassifications while '*bior3.4*' has only two. The decision tree of 'bior3.9' at scale 5 with C4.5 is shown in Figure 18.

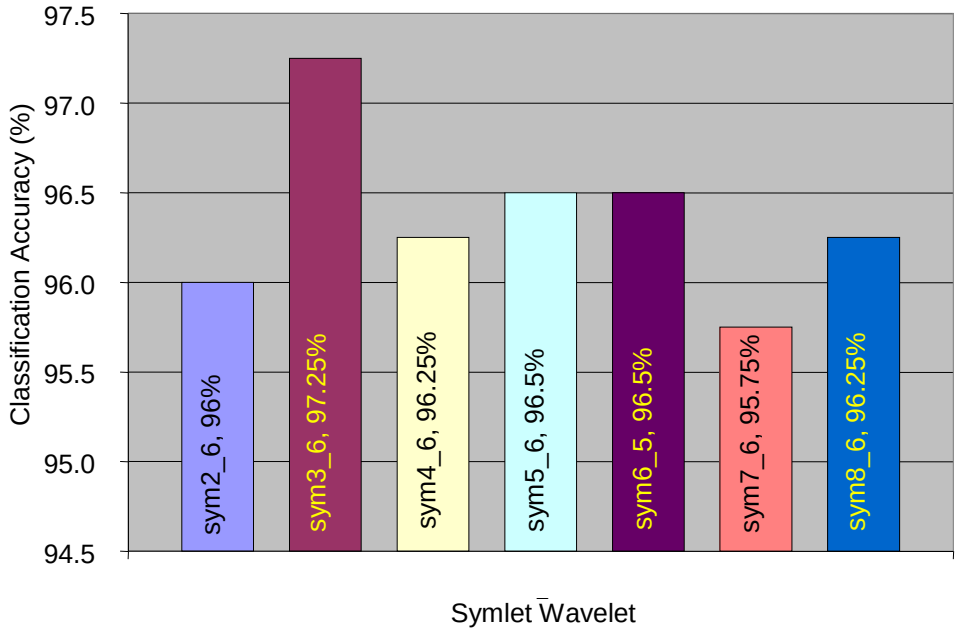

Fig. 12: *Comparison of Symlet Wavelets with CWT Features using C4.5 Algorithm.*

Morlet wavelet at scale 4 gives the classification accuracy of 96.5%. The confusion matrix is represented in Figure 20. Although the overall classification accuracy is a little less than that of the ones discussed earlier, one can note that Morlet does not misclassify any fault conditions (IRF, ORF, IORF) as 'Good'. For this reason alone one can chose Morlet for fault diagnosis of roller bearings. From an application point of view Morlet stands out apart from other wavelets.

	Good	IRF	ORF	IORF
Good	99	1	0	0
IRF	0	97	3	0
ORF	0	5	95	0
IORF	1	1	0	98

Fig. 13: *Confusion Matrix of C4.5 Algorithm with Sym3 (scale 6) CWT Features*

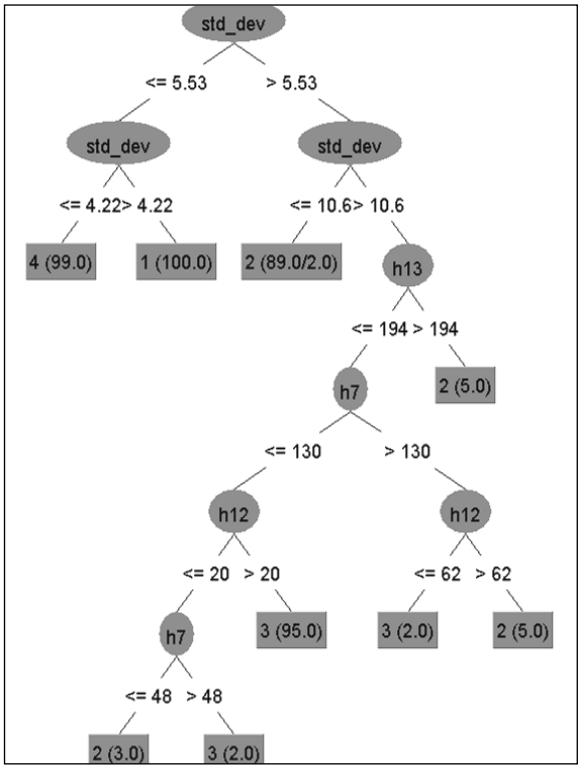

Fig. 14*: Decision Tree using C4.5 Algorithm with Sym3 (Scale 6) CWT Features.*

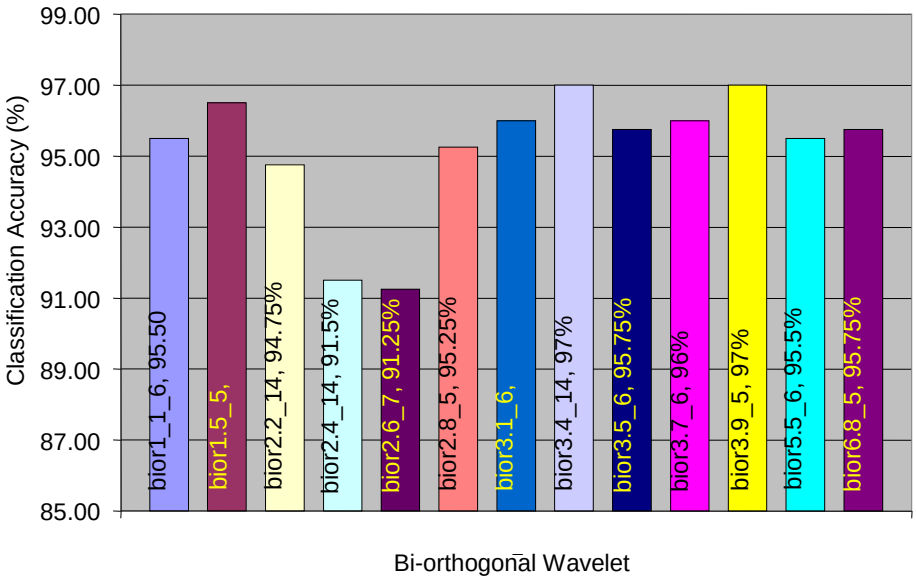

Fig. 15: *Comparison of Bior3.1 DWT Features.*

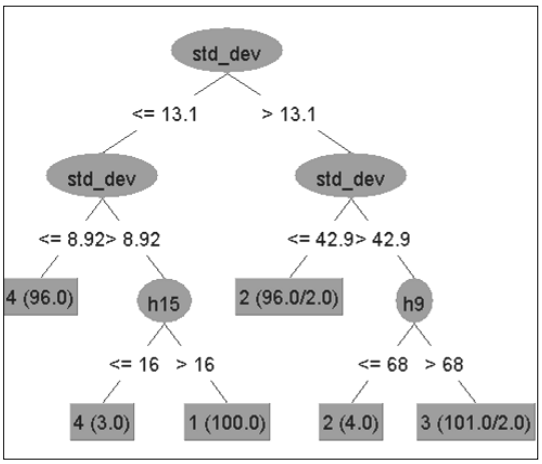

Fig.16: *Decision Tree using C4.5 Algorithm with Bior3.4 (Scale 14) CWT Features.*

	Good	IRF	ORF	IORF
Good	99	1	0	0
IRF	0	97	3	0
ORF	0	4	96	0
IORF	3	1	0	96

Fig. 17: *Confusion Matrix of C4.5 Algorithm with Bior3.4 (scale 14) CWT Features.*

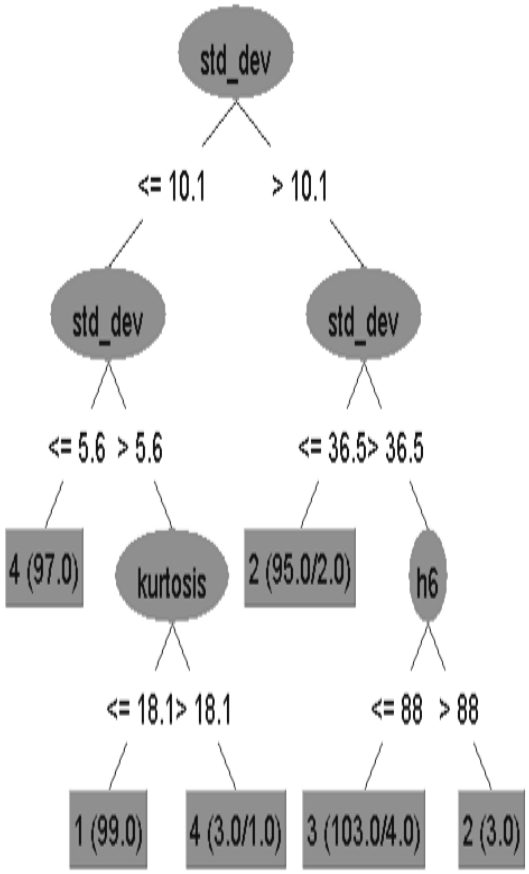

Fig. 18*: Decision Tree using C4.5 Algorithm with Bior3.9 (Scale 5) CWT Features.*

	Good	IRF	ORF	IORF
Good	98	1	0	1
IRF	0	95	5	0
ORF	0	2	98	0
IORF	2	1	0	97

Fig. 19: *Confusion Matrix of C4.5 Algorithm with Bior3.4 (scale 14) CWT Features.*

	Good	IRF	ORF	IORF
Good	97	1	0	2
IRF	0	94	6	0
ORF	0	4	96	0
IORF	0	1	0	99

Fig. 20: *Confusion Matrix of Morlet (scale 4) CWT Features.*

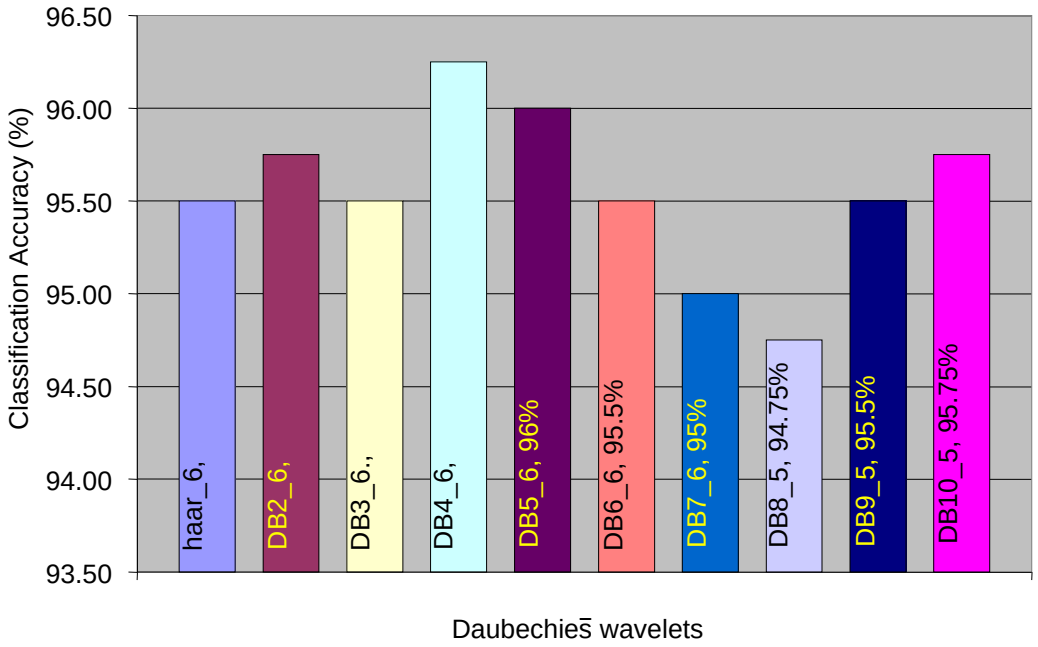

Fig. 21*: Comparison of Daubechies Wavelets Among its Family with CWT Features.*

	Good	IRF	ORF	IORF
Good	99	1	0	0
IRF	0	95	5	0
ORF	0	7	93	0
IORF	1	1	0	98

Fig. 22*: Confusion Matrix of DB4 CWT Features.*

Daubechies wavelet with scale 6 gives next highest accuracy (96.25%). The classification accuracies of the Daubechies family wavelets are presented in Figure 21. It shows that DB4 gives high accuracy (96.25%). The confusion matrix is depicted in Figure 22. From the confusion matrix, one can observe that only one sample of 'Good' condition is misclassified as 'IRF' and one sample is misclassified as 'Good' whose original condition is 'IORF'. Again here also, the misclassifications amongst 'IRF' and 'ORF' bring down the overall classification accuracy to 96.25%. There is no case where the sample is misclassified as 'IORF'.

This indicates that the class 'IORF' is identified well and the CWT features are able to identify them properly from other classes. A similar performance is demonstrated by 'coiflet' ('*coif4*') wavelet with classification accuracy of 95.75%; Figure 23 and 24 depicts its performance.

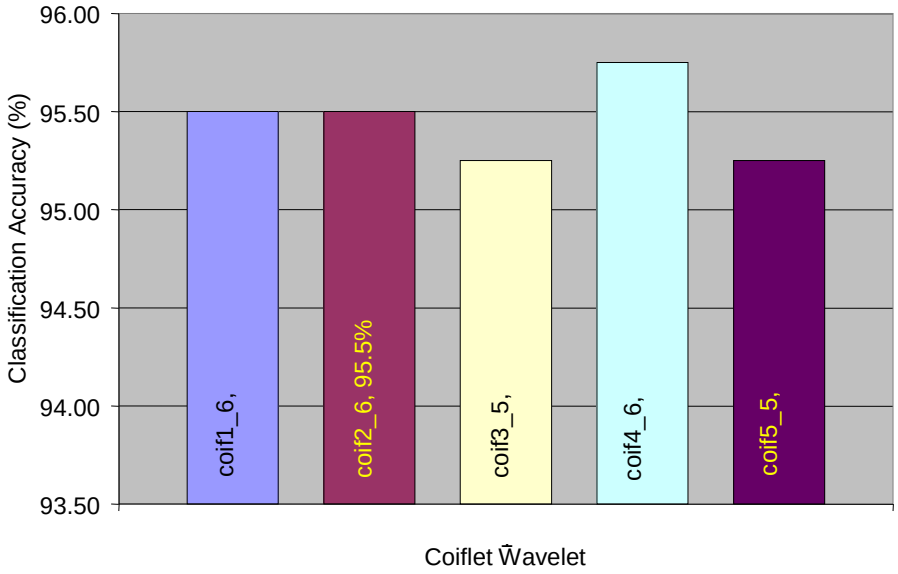

Fig. 23: *Comparison of Coiflet Wavelets among its Family with CWT Features.*

	Good	IRF	ORF	IORF
Good	99	1	0	0
IRF	0	93	7	0
ORF	0	6	94	0
IORF	2	0	1	97

Fig. 24: *Confusion Matrix of Coif4 (scale 6) CWT Features.*

	Good	IRF	ORF	IORF
Good	100	0	0	0
IRF	0	92	8	0
ORF	0	10	90	0
IORF	0	0	0	100

Fig. 25: *Confusion Matrix of Gauss-Symmetric (scale 41) CWT Features.*

Figure 25 shows the confusion matrix of Gauss-symmetric wavelet. At scale 41, it gives the classification accuracy of 95.5%. The confusion matrix reveals that it has very good diagnostic ability between 'Good' and faulty conditions as demanded by many practical applications like Morlet wavelet. Also, the 'IORF' class is well defined and there is no misclassification as 'IORF'.

However, it is not able to distinguish between 'IRF' and 'ORF'. This brings down the classification accuracy to a level where application engineers might consider it unsuitable for practical applications.

CLASSIFICATION

Support Vector Machines

SVM is a supervised machine learning technique which has its basis in statistical learning theory. It is used for analyzing data and recognizing patterns. It can be effectively used in bearing fault diagnostics because it can perform classification and regression in small sample cases. Here 'machine' means an algorithm. The algorithm is given a set of inputs(features) with the associated output values. Each feature can be seen as a dimension of a hyper-plane. In case of two

class problems, the SVM algorithm creates a hyper-plane to separate the hyper-space in two classes. This approach can be extended for multi-class problems. The algorithm woks to achieve maximum separation between the classes. Maximum separation needs a larger margin which helps to minimize 'generalization error'. Generalization error is the error obtained in classification, when a new set of data is fed to the trained classifier.

This process leads to two hyper-planes parallel to the separating plane called 'bounding planes'. It is the distance between these bounding planes, which the algorithm tries to maximize. The nearest data points used to define margin which are beyond the bounding planes are called support vectors. Therefore, points *P1,P2, P3, P4* and *P5*,which belong to *A*– are support vectors (Refer Figure 26).

To mathematically express the above process we use the notations used by V.Sugumaran *et al.*[3] Let 'A' be a m x n matrix with elements in real space. 'D' is a m x 1 matrix representing class label (+1 and –1). '*e*' is a vector containing ones('1's) . 'ν' is a control parameter that defines the weight of error minimisation and separation between bounding planes in the objective function. '*w*' represents the orientation parameter and 'γ ' is the location parameter of separating hyper-plane relative to origin.

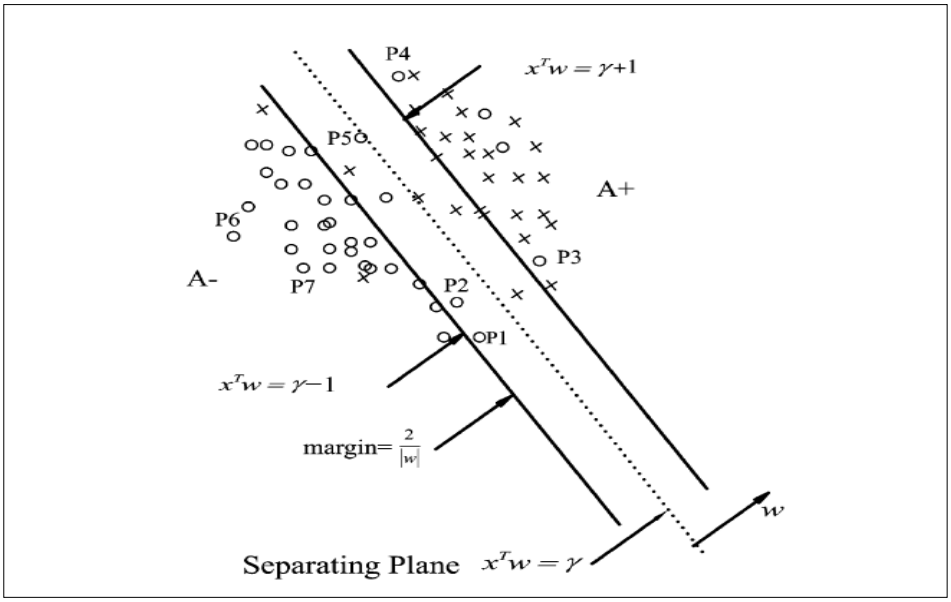

Fig. 26: *SVM Classifier*.

Mathematically,

$$min \quad \nu e'y + \frac{1}{2} w'w, \quad (w, \gamma, y) \epsilon R^{n+1+m}$$

$$such\ that \quad D(Aw - e\gamma) + y \geq e, \quad y \geq 0, \qquad \text{Eq. (5)}$$

$$where$$

$$A \epsilon R^{m \times n}, D \epsilon \{-1, +1\}^{m \times 1}, \quad e = 1^{m \times 1}$$

For classification using SVM, two classes are considered at a time.

There are four CWT features used as input and corresponding to them there are four orientation parameters '*w*', namely, w_1, w_2, w_3, and w_4.

The location parameter γ and control parameter V are tabulated for different pairs of bearing conditions. The training parameters of SVM are presented in Table 1.

Table 1: *SVM Training Parameters with CWT Features.*

Class Pairs	w_1	w_2	w_3	w_4	Location parameter (γ)	Control parameter (ν)
Good *Vs* IRF	0.0444	–0.0887	0.0307	1.8957	–11.7001	0.01
Good *Vs* ORF	–0.004	–0.0104	0.0045	0.3414	–2.9098	0.01
Good *Vs* IORF	–0.0353	–0.0054	–0.0158	–1.7061	7.7737	0.01
IRF *Vs* ORF	0.1267	–0.0546	0.1662	0.6866	–11.1157	0.01
IRF *Vs* IORF	–0.1570	0.0537	–0.0120	–0.9320	5.3756	0.01
ORF *Vs* IORF	–0.0851	–0.0035	–0.0115	–0.3976	3.9084	0.01

The control parameter 'ν' is kept constant as '0.01' in the study. However, it can be changed for different pairs to get good classification accuracy for cases where the accuracy obtained is not acceptable. Here, location parameter (γ) is kept constant for all pairs of bearing conditions ('0.01'). This value is obtained after experimentation with many control parameter 'ν' values by trial and error. The classification results and confusion matrix are presented in Table 2 and Figure 27 respectively. From the confusion matrix shown in Figure 27, one can note that none of the 'Good' bearings is misclassified; but 3% of the faulty conditions of the bearings are misclassified as 'Good'. The overall classification accuracy is only 96.5%.

Table 2: *Classification Results of SVM with CWT Features.*

Test Parameter	Values
Test mode	10-fold cross-validation
Time taken to build model	19.66 seconds
Total Number of Instances	400
Correctly Classified Instances	386 (96.5%)
Incorrectly Classified Instances	14 (3.5%)
Mean absolute error	0.0256
Root mean squared error	0.1211

	Good	IRF	ORF	IORF
Good	100	0	0	0
IRF	1	91	8	0
ORF	0	8	92	0
IORF	2	1	0	97

Fig. 27: *Confusion Matrix of SVM with CWT Features.*

Proximal Support Vector Machines

PSVM is a modified version of SVM. In this classifier, instead of assigning points to one of two disjoint half-spaces, they are assigned to the closest of the two parallel planes which are spaced as apart as possible. This formulation leads to a very simple and fast algorithm for generating a linear or a nonlinear classifier.

Thus, Eq. (5) modifies to

$$\min_{(w,\gamma,y)\epsilon R^{n+1+m}} \quad \nu \frac{1}{2}\|y\|^2 + \frac{1}{2}(w^T w + \gamma^2)$$

$$such\ that, D(Aw - e\gamma) + y = e \quad \text{Eq. (6)}$$

The above formulation can be geometrically interpreted as shown in Figure 28. [3] Referring to Figure 28 y represents deviation (scaled $1/\|w\|$) of a point from the plane from the plane passing through the centroid of the data cluster (A+ or A–) to which the point belongs. Hence, y does not have a non negativity constraint. The margin between the bounding planes is maximised with respect to both orientation w and relative

location γ to the origin. Strong convexity of the objective function has a key role. By replacing the inequality by equality in Eq. (6), it becomes possible to write an explicit solution.

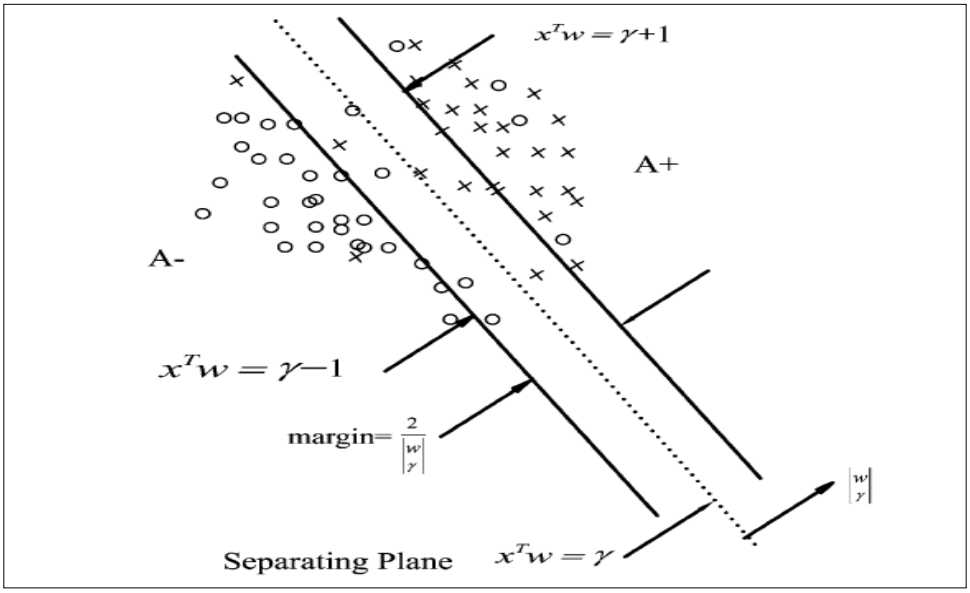

Fig. 28: *PSVM Classifier.*

The geometrical interpretation of the formulation obtained in Eq. (6) can be done as follows: The planes $x^T w - \gamma = \pm 1$ can be thought of as ''proximal'' planes, around which the points of each class are clustered. These planes are pushed as far apart as possible by the term $w^T w + \gamma^2$ in the objective function. This term is the reciprocal of the 2-norm distance squared between the two planes in the (w, γ) space. This interpretation is not based on the idea of maximising the margin. Once training is over, prediction of the class of any new set of features can be made by using the following decision function

$$f(x) = sign(w^T x - \gamma) \qquad \text{Eq. (7)}$$

If the value of $f(x)$ is positive then the new set of features belongs to class A+; otherwise it belongs to class A−.

In PSVM, as stated before, the parameters 'w' and 'γ' define the separating plane. Here, V is weight parameter which is varied by trial and error to get high classification accuracy. While computing classification accuracy, the value of V is varied from '0.0001' to '10000' at different step levels. The confusion matrix for PSVM is given in Figure 29. The classification accuracy is 100% in all the cases. Interestingly, it is the best, one can achieve. Of course, subject to the test conditions. The classification parameters with CWT features are given in Table 3. For many values of V, PSVM gives 100% classification accuracy. However, only one representative value is presented in Table 3.

	Good	IRF	ORF	IORF
Good	100	0	0	0
IRF	0	100	0	0
ORF	0	0	100	0
IORF	0	0	0	100

Fig. 29*: Confusion Matrix of PSVM with CWT Features.*

Table 3: *PSVM Classification Parameters with CWT Features.*

Class pairs	*w1*	*w2*	*w3*	*w4*	Control parameter (ν)	Location parameter(γ)	Accuracy
Good *Vs* IRF	–0.5817	0.0109	–0.0002	0.0029	1	–1.3210	100%
Good *Vs* ORF	–0.5817	0.0109	–0.0002	0.0029	1	–1.3210	100%
Good *Vs* IORF	–0.5817	0.0109	–0.0002	0.0029	1	–0.3210	100%
IRF *Vs* ORF	–0.5817	0.0109	–0.0002	0.0029	1	–1.3210	100%
IRF *Vs* IORF	–0.5817	0.0109	–0.0002	0.0029	1	–1.3210	100%
ORF *Vs* IORF	–0.5817	0.0109	–0.0002	0.0029	1	–1.3210	100%

RESULTS AND DISCUSSION

The CWT features were computed. The wavelet and feature selection was performed using C4.5 algorithm. The process yielded that the wavelet *'rbio1.5'* gives maximum classification accuracy. The decision tree lead to the selection of four features.

Their performance in classification have been presented. The classification accuracies of SVM and PSVM using CWT features are obtained. PSVM proved to have 100% classification accuracy and thus very well suited for practical applications.

REFERENCES:

1. Prabhakar S., Mohanty A.R., Sekhar A.S. Application of discrete wavelet transform for detection of ball bearing race faults. *Tribology International*. 2002; 35: 793–800p.
2. Lou X., Loparo K.A. Bearing fault diagnosis based on wavelet transform and fuzzy inference. *Mechanical Systems and Signal Processing*. 2004; 18:1077–95p.
3. Sugumaran V., Muralidharan V., Ramachandran K.I. Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing. *Mechanical Systems and Signal Processing*. 2001; 21: 930–42p.
4. Abbasion S., Rafsanjani A., Farshidianfar A., *et al.* Rolling element bearings multi-fault classification based on the wavelet denoising and support vector machine. *Mechanical Systems and Signal Processing*. 2007; 21: 2933–45p.
5. Yu G., Li C., Kamarthi S. Machine fault diagnosis using a cluster-based wavelet feature extraction and probabilistic neural networks. *International Journal of Advance Manufacturing Technology*. 2009; 42: 145–51p.
6. Li Y. Study on Incipient Fault Diagnosis for Rolling Bearings Based on Wavelet and Neural Networks. *4th International Conference on Natural Computation.* 2008 Oct 18–20; Jinan. IEEE; 2008. 262–5p.
7. Wang D., Miao Q., Fan X., *et al.* Rolling element bearing fault detection using an improved combination of Hilbert and Wavelet transforms. *Journal of Mechanical Science and Technology.* 2009; 23: 3292-3301p.
8. Su W., Wang F. , Zhu H., Zhixin *et al.* Rolling element bearing faults diagnosis based on optimal Morlet wavelet filter and autocorrelation enhancement. Mechanical *Systems and Signal Processing*. 2010; 24: 1458–72p.
9. Chebil J., Noel G., Mesbah M., *et al.* Wavelet Decomposition for the Detection and Diagnosis of Faults in Rolling Element Bearings. *Jordan Journal of Mechanical and Industrial Engineering.* 2009; 3(4): 260–7p.
10. Kankar P.K., Sharma S.C., Harsha S.P. Fault diagnosis of ball bearings using

continuous wavelet transform. *Applied Soft Computing*. 2011;11: 2300–12p.
11. Saravanan N., Ramachandran K.I. Incipient gear box fault diagnosis using discrete wavelet transform (DWT) for feature extraction and classification using artificial neural network (ANN). *Expert Systems with Applications*. 2010; 37: 4168–81p.
12. Konar P., Chattopadhyay P. Bearing fault detection of induction motor using wavelet and Support Vector Machines (SVM). *Applied Soft Computing*. 2011;11: 4203–11p.
13. Li H., Fu L., Zheng H. Bearing fault diagnosis based on amplitude and phase map of Hermitian wavelet transform. *Journal of Mechanical Science and Technology*. 2011; 25: 2731–40p.
14. Feng K., Jiang Z., He W., *et al*. Rolling element bearing fault detection based on optimal antisymmetric real Laplace wavelet. *Measurement*. 2011; 44: 1582–91p.
15. Kanneg D., Wang W. A Wavelet Spectrum Technique for Machinery Fault Diagnosis. *Journal of Signal and Information Processing* 2011; 2: 322–9p.
16. Bendjama H., Bouhouche S., Boucherit M.S. Application of Wavelet Transform for Fault Diagnosis in Rotating Machinery. *International Journal of Machine Learning and Computing*. 2012; 2(1): 82–7p.
17. Chen K., Li X., Wang F., *et al*. Bearing Fault Diagnosis Using Wavelet Analysis. *International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering. (ICQR2MSE).* 2012 June 15–18; Chengdu. 2012. 699–702p.
18. Li P., Kong F., He Q., *et al*. Multiscale slope feature extraction for rotating machinery fault diagnosis using wavelet analysis. *Measurement.* 2013; 46: 497–505p.
19. Shen C., Wang D., Kong F, *et al.* Fault diagnosis of rotating machinery based on the statistical parameters of wavelet packet paving and a generic support vector regressive classifier. *Measurement*. 2013; 46: 1551–64p.
20. Muralidharan V., Sugumaran V. A comparative study of Naïve Bayes classifier and Bayes net classifier for fault diagnosis of monoblock centrifugal pump using wavelet analysis. *Applied Soft Computing.* 2012; 12: 2023–9p.
21. Muralidharan V., Sugumaran V. Feature extraction using wavelets and classification through decision tree algorithm for fault diagnosis of mono-block centrifugal pump. *Measurement*. 2013; 46: 353–9p.
22. Kumar H.S., Dr. Srinivasa P.P., Sriram N.S., et al. ANN based evaluation of performance of wavelet transform for condition monitoring of rolling element bearing. *International Conference On Design And Manufacturing*, IConDM 2013. Procedia Engineering. 2013; 64: 805–14p.
23. Meng L., Xiang J., Wang Y., *et al*. A hybrid fault diagnosis method using morphological filter–translation invariant wavelet and improved ensemble empirical mode decomposition. *Mechanical Systems and Signal Processing*. 2015; 50–51: 101–15p.