

## Research Article

# Evaluation of Effectiveness of Wavelet Based Denoising Schemes Using ANN and SVM for Bearing Condition Classification

Vijay G. S.,<sup>1</sup> Kumar H. S.,<sup>2</sup> Srinivasa Pai P.,<sup>2</sup> Sriram N. S.,<sup>3</sup> and Raj B. K. N. Rao<sup>4</sup>

<sup>1</sup>Department of Mechanical and Manufacturing Engineering, Manipal Institute of Technology, Manipal University, Karnataka, Manipal 576104, India

<sup>2</sup>Department of Mechanical Engineering, NMAM Institute of Technology, Karnataka, Nitte 574104, India

<sup>3</sup>Department of Mechanical Engineering, Vidya Vikas Institute of Engineering and Technology, Karnataka, Mysore 570028, India

<sup>4</sup>COMADEM International, Birmingham B29 6DA, UK

Correspondence should be addressed to Vijay G. S., vgs.mpl@gmail.com

Received 31 March 2012; Revised 27 September 2012; Accepted 11 October 2012

Academic Editor: Karim Oweiss

Copyright © 2012 Vijay G. S. et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The wavelet based denoising has proven its ability to denoise the bearing vibration signals by improving the signal-to-noise ratio (SNR) and reducing the root-mean-square error (RMSE). In this paper seven wavelet based denoising schemes have been evaluated based on the performance of the Artificial Neural Network (ANN) and the Support Vector Machine (SVM), for the bearing condition classification. The work consists of two parts, the first part in which a synthetic signal simulating the defective bearing vibration signal with Gaussian noise was subjected to these denoising schemes. The best scheme based on the SNR and the RMSE was identified. In the second part, the vibration signals collected from a customized Rolling Element Bearing (REB) test rig for four bearing conditions were subjected to these denoising schemes. Several time and frequency domain features were extracted from the denoised signals, out of which a few sensitive features were selected using the Fisher's Criterion (FC). Extracted features were used to train and test the ANN and the SVM. The best denoising scheme identified, based on the classification performances of the ANN and the SVM, was found to be the same as the one obtained using the synthetic signal.

## 1. Introduction

The detection of fault in the machinery, in its incipient stage itself, has gained prime importance as it avoids machine down time, catastrophic failure of the machinery, threat to human life, high maintenance costs, and so forth. The fault diagnostic techniques based on the vibration signal analysis have become popular in recent times [1, 2]. The problem of the strong noise components masking the weak characteristic signals has always posed challenges to the condition monitoring expert. Several wavelet based signal processing techniques aiming at denoising the measured signal so as to increase the Signal-to-Noise Ratio (SNR) and reduce the Root-Mean-Square Error (RMSE) have been proposed and tried by several researchers [3–7]. The details of the techniques used by some of the researchers have been explained in Section 2.2. The wavelet based denoising technique has gained popularity due to

its effectiveness and ease of application [8]. It overcomes the difficulty of determining the resonant frequency of the system. Therefore, the wavelet technique has been adopted in this work for denoising the bearing vibration signals. The detail coefficients, obtained from the Discrete Wavelet Transform (DWT), generally include a large proportion of the high-frequency noise components along with some of the characteristic information of the machine fault. Suitable compression or suppression of these components would remove the noise. The suppressed detail coefficients can then be used along with the original approximation coefficients in reconstructing the decomposed signal, by using the Inverse Wavelet Transform (IWT), which would now be fairly free of the noise [9, 10].

The Artificial Neural Networks (ANNs) and the Support Vector Machines (SVMs) have been used to a large extent in the fault diagnosis problems with high success rates. The bearing vibration signals are nonstationary signals and

hence a nonlinear mapping from the input space to the output space is required, which is successfully fulfilled by the classifiers like the ANN and the SVM. Several researchers have applied the ANN and the SVM to the bearing fault identification problem. Wang et al. [12] have used the ANN, with difference values of the autoregressive coefficients as inputs, in a rotating machinery fault identification problem. Zarei [13] has proposed to improve the diagnostic abilities of an ANN applied to a four-condition bearing classification problem by using the time domain features alone as the ANN inputs. Kankar et al. [14] have applied the ANN and the SVM to a five-condition ball bearing defect classification problem and obtained high classification accuracies. The SVM is a soft computing tool which performs the tasks executed by an ANN, but with a different approach. The SVM positions a hyper plane between the two classes of data, thus separating the data belonging to the two classes. The ANN's approach is to minimize the error on the training data set which is known as the empirical risk minimization, whereas the SVM's approach is based on the structural risk minimization, in which the upper bound of the generalization error is minimized [15]. Yang et al. [16] have used the energy features extracted from a number of Intrinsic Mode Functions as input vectors to the SVM classifier to diagnose the REB condition. Sugumaran et al. [17] have illustrated the use of a decision tree to identify the best features, extracted from bearing vibration signal, which were given as inputs to the Proximal Support Vector Machine (PSVM) and the SVM. They reported that the PSVM performed better than the SVM. The popularity of the ANN and the SVM classifiers in the REB diagnostics has motivated the authors of this paper to use them in this work.

The objective of any classifier like the ANN or the SVM is to attain a good generalization ability, that is, to exhibit high accuracies on the training and the test data. This calls for the optimal design of the ANN/SVM architecture. One of the requirements of designing an optimal ANN/SVM architecture is to reduce the input dimensionality, that is, to select a few predominantly sensitive features as inputs. This is known as the Dimensionality Reduction Technique (DRT). Researchers have proposed and tried several DRTs. Some of the popular DRTs are Principal Component Analysis (PCA), Fisher's Criterion (FC), Singular Value Decomposition (SVD), Genetic Algorithm (GA), and so forth. Yen and Lin [20] have investigated the effectiveness of the DRTs, namely, Linear Discriminant Analysis and FC for reducing the number of wavelet packet features extracted for analyzing a bearing classification problem. Fuente et al. [21] have used the Fisher's Discriminant Analysis (FDA) for identifying the faults in a real plant in terms of maximizing the scatter between the classes and minimizing the scatter within each class. Chiang et al. [22] and Tang and Li [23] explain the fault diagnosis based on the FDA. Jack and Nandi [24], Samanta et al. [25], and Saxena and Saad [26] have shown in their works that the GA can be effectively used as a DRT along with the optimization of the topology parameters of the ANN/SVM. From the preliminary work carried out by the authors of this paper, it was found that the GA effectively selected the sensitive features, but took a longer time as the

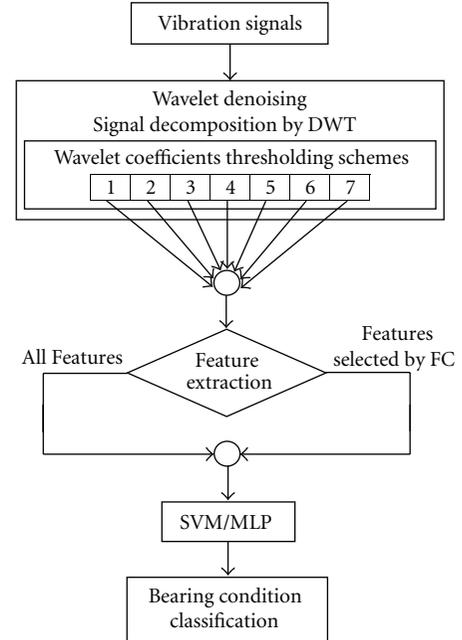


FIGURE 1: Denoising schemes and bearing diagnostic procedure employed in the study.

GA depended on the performance of the ANN or the SVM to compute the fitness value, that is, for every computation of the fitness value, the ANN or SVM had to be run, making the process time consuming. However, the effectiveness of the FC in selecting the sensitive features was found to be comparable with that of the GA based feature selection, and, more importantly, unlike the GA, FC was independent of the performance of the ANN or SVM. Therefore, in this work, FC has been used as a DRT.

In this paper, the effectiveness of seven different wavelet based denoising schemes have been evaluated in terms of the classification accuracies of the ANN and the SVM on the denoised training and the test data, extracted from the REB vibration signals. Firstly, a synthetic signal (representing the vibration signal of a defective bearing) has been corrupted by a Gaussian white noise and subjected to the seven denoising schemes. Secondly, the real-time bearing vibration signals, measured from a customized bearing test rig under one load and two speed conditions, for four conditions of the bearings, have been subjected to the same denoising schemes. The denoising scheme which provided high SNR and low RMSE in the first part of the work provided high classification accuracies (on the training and the test data) in the second part of the work. The focus of this work was to evaluate the best wavelet based denoising scheme based on the performance of the ANN and the SVM. Figure 1 shows the denoising schemes and the bearing diagnostic procedure employed in this study.

## 2. Wavelet Based Denoising

The characteristic vibration signals of the defective bearings are not generally readily available when collected by means

TABLE 1: List of thresholding schemes.

Thresholding scheme	Researcher/s
Conventional thresholding schemes	$s1$ Donoho [11]
	$s2$ Donoho [11]
Modified or improved thresholding schemes	$s3$ Huaigang et al. [3]
	$s4$ Fang and Huang [4]
	$s5$ Lin and Cai [5]
	$s6$ Zhang et al. [6]
	$s7$ Cai-lian et al. [7]

of a Data Acquisition (DAQ) system. This is mostly because the noise, influenced by the resonant frequency of the rotating system, masks the characteristic vibration signals. The noises are often stochastic signals whose frequency band will overlap with the interested signals. Therefore, it is difficult to eliminate the noise from the signals effectively by using the traditional filtering methods. Also, the traditional methods of denoising need the knowledge of the parameters which are difficult to be determined [8]. The wavelet based denoising has gained popularity due to its effectiveness and also that it overcomes the difficulties of the traditional denoising methods. The SNR must appreciably increase and the RMSE must become small on a successful application of a denoising method.

Suppose that a signal of interest  $f(n)$  has been corrupted by the noise  $z(n)$ , so that we get a signal  $g(n)$  as in (1) which resembles the raw signal collected by means of a DAQ system,

$$g(n) = f(n) + \sigma z(n), \quad (1)$$

where  $z(n)$  is a unit-variance, zero-mean Gaussian white noise and  $\sigma^2$  is the variance of the noise. The denoising is a way to recover  $f(n)$  from the samples of  $g(n)$  as properly as possible. The three-step procedure adopted in the wavelet based denoising is (i) decomposition of the raw signal using the wavelet transform to get the approximation and the detail coefficients, (ii) suppressing the detail coefficients by selecting a suitable threshold value and by applying a suitable thresholding rule, and (iii) reconstructing the signal by applying IWT to the original approximation coefficients and the suppressed detail coefficients to get the denoised signal [9, 10].

Several denoising schemes (step ii) have been proposed by researchers [3–7]. In this work, the denoising effectiveness of seven different denoising schemes has been compared. Table 1 gives the list of seven denoising schemes and the researchers who have proposed them.

**2.1. Conventional Denoising Schemes.** The wavelet denoising method focuses on the selection of the thresholding rules and the determination of the threshold value. Donoho [11] gave two thresholding rules, namely, hard-thresholding ( $s1$ ) and the soft-thresholding ( $s2$ ), which are considered to be the conventional wavelet based denoising schemes and they

are readily available functions in the Wavelet toolbox of MATLAB. The hard-thresholding scheme is expressed as

$$y_1(x) = \begin{cases} x & \text{if } |x| \geq \lambda, \\ 0 & \text{if } |x| < \lambda, \end{cases} \quad (2)$$

where  $x$  is the wavelet coefficient,  $y_1(x)$  is the corresponding suppressed wavelet coefficient by hard-thresholding, and  $\lambda$  is the threshold value.

The soft-thresholding scheme is expressed as

$$y_2(x) = \begin{cases} \text{sign}(x)(x - \lambda) & \text{if } |x| \geq \lambda, \\ 0 & \text{if } |x| < \lambda, \end{cases} \quad (3)$$

where  $y_2(x)$  is the suppressed wavelet coefficient obtained by soft-thresholding and the other terms have the same meaning as in (2).

**2.2. Modified Soft-Thresholding Schemes Proposed by Different Researchers.** A list of the denoising schemes  $s3$  to  $s7$  proposed by various researchers [3–7] is provided in Table 1. Huaigang et al. [3] have proposed an improved soft-thresholding function as given in (4). According to them, the conventional thresholding functions set the coefficients below the threshold value to zero, but, in their proposed method, these coefficients were tuned by a polynomial function. The coefficients that were below the threshold value and close to it were attenuated to a value less than the far coefficients. For important coefficients, the function was garrote-like, resulting in a more powerful function:

$$y_4(x) = \begin{cases} x + (k-1)\lambda - 0.5 \frac{k\lambda^m}{x^{m-1}} & \text{if } x > \lambda, \\ 0.5 \frac{k|x|^{[m+(2-k)/k]}}{\lambda^{[m+(2-2k)/k]}} \text{sign}(x) & \text{if } |x| < \lambda, \\ x - (k-1)\lambda - 0.5 \frac{k(-\lambda^m)}{x^{m-1}} & \text{if } |x| < -\lambda, \end{cases} \quad (4)$$

where  $m$  and  $k$  are tuning parameters and the other terms in (4) have the same meaning as in (2). By tuning the parameter  $k$ , the thresholding function can be between the hard- and the soft-thresholding functions. By tuning the parameter  $m$ , the near-optimum thresholding function is adjusted to the optimum one by applying small changes. As per [3], optimization of the parameter  $k$  works similar to a global search and optimization of the parameter  $m$  works like a local search in finding the best thresholding function. The authors in [3] have selected  $m = 2$  and  $4$  and  $k \in [0, 1]$ . In this work,  $m = 4$  and  $k = 0.8$  have been selected.

Fang and Huang [4] have proposed a wavelet trimmed thresholding scheme as given in (5) which was an improved version of the hard- and the soft-thresholding schemes:

$$y_5(x) = \begin{cases} x \left( \frac{|x|^\alpha - \lambda^\alpha}{|x|^\alpha} \right), & \text{if } |x| \geq \lambda, \\ 0, & \text{if } |x| < \lambda, \end{cases} \quad (5)$$

where  $\alpha$  is a parameter and the other terms in (5) have the same meaning as in (2). They suggested that with careful

tuning of the parameter  $\alpha$  for a particular signal, a best denoising effect could be achieved. When  $\alpha = 1$ , it was equivalent to the soft-thresholding and when  $\alpha \rightarrow \infty$ , it was equivalent to the hard-thresholding. Accordingly, in this work,  $\alpha = 5$  has been chosen.

Lin and Cai [5] proposed a new threshold function given in (6) which had the advantage of the nonnegative dead zone thresholding:

$$y_6(x) = \begin{cases} \sqrt{1 - \beta^2} \text{sign}(x)|x - \lambda| + \beta x, & \text{if } |x| \geq \lambda, \\ 0, & \text{if } |x| < \lambda, \end{cases} \quad (6)$$

$$\beta = \exp\left[-\frac{(|x| - \lambda)^2}{k}\right], \quad (6a)$$

where  $k$  is a positive number and the other terms in (6) have the same meaning as in (2). When  $x \rightarrow \lambda$ ,  $\beta \rightarrow 1$  and  $y_6(x) = x$ , which overcomes the disadvantage of the soft-thresholding, and when  $x \rightarrow \infty$ ,  $\beta \rightarrow 0$  and  $y_6(x) \rightarrow 0$ , which makes the signal smoother than the hard-thresholding function. A value of  $k = 0.5$  has been chosen in this work, for using this scheme.

Zhang et al. [6] have proposed an improved thresholding function given in (7):

$$y_7(x) = \begin{cases} \text{sign}(x)(|x|^u - \alpha\lambda^u)^{1/u}, & \text{if } |x| \geq \lambda, \\ 0, & \text{if } |x| < \lambda, \end{cases} \quad (7)$$

where  $u$  and  $\alpha$  are parameters whose proper tuning can provide an effective denoising while the other terms in (7) have the same meaning as in (2). The power  $u$  is used in order to enlarge the difference between the signal and the noise,  $u > 0$  ( $u = 2, 3, 4, \dots$ ). It can be observed that when  $\alpha = 0$ , (7) becomes the hard-thresholding function. When the value of  $\alpha$  is appropriately chosen between 0 and 1, the effectiveness of the denoising could be optimized. For using this scheme in this work,  $u = 10$  and  $\alpha = 0.6$  have been chosen.

In a new thresholding function proposed by Cai-lian et al. [7], suppression of the detail coefficients was done according to (8):

$$y_8(x) = x \times \left\{ \frac{1 - \exp\left[-\frac{(x/\lambda)^2}{k}\right]}{1 + \exp\left[-\frac{(x/\lambda)^2}{k}\right]} \right\}. \quad (8)$$

This thresholding function depends on the proper selection of the constant  $\lambda$ . It is continuous unlike the conventional soft-thresholding function and is easily differentiable, statistically very reliable, and robust, making it completely suitable for the discrete signal denoising. The optimum value of  $\lambda$  can be determined as proposed in [7], but, in the current work,  $\lambda = 0.8$  has been chosen by trial and error so as that the ANN's and the SVM's training and test accuracies were maximum.

### 3. Wavelet Based Denoising of a Synthetic Signal

The focus of the first part in this work was to apply the seven wavelet based denoising schemes listed in Table 1 to a

synthetic signal that represented a defective bearing vibration signal. In order to simulate the vibration signal of a defective bearing, a weak synthetic signal  $0.5e^{-500t} \sin(10000t)$  with a defect frequency of 50 Hz was considered. A sampling frequency of 48 kHz was used as the real-time bearing vibration signals were acquired at the same rate. A plot of the synthetic signal is shown in Figure 2(a). It was corrupted with a strong zero-mean Gaussian white noise. A plot of the corrupted signal is shown in Figure 2(b). The energy of the synthetic signal was 41.89 and that of the corrupted signal was 510.77. The expression for computing the signal energy is given in (9), whereas the expressions for computing the SNR and the RMSE for a denoised signal are given in (11), and (12) respectively:

$$E = \sum_{i=1}^N x_i^2, \quad (9)$$

where  $E$  is the energy of the signal  $x$  and  $N$  is the length of the signal:

$$\text{SNR} = 10 \log_e \left[ \frac{\sum_{i=1}^N x_i^2}{\sum_{i=1}^N (d_i - x_i)^2} \right], \quad (10)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (d_i - x_i)^2}, \quad (11)$$

where  $x$  is the corrupted signal,  $d$  is the denoised signal, and  $N$  is the length of the signal.

The corrupted signal was subjected to the seven wavelet based denoising schemes listed in Table 1. The corrupted signal was subjected to the DWT so as to decompose it into four levels using Daubechies 8 mother wavelet through a customized MATLAB program. According to Nyquist's rule, the maximum frequency of the vibration signal was set to 24 kHz because the sampling frequency was 48 kHz. The frequency bandwidths of the approximation and the detail coefficients of the wavelet decompositions are shown in Figure 3. For each level of the wavelet decomposition, the threshold value  $\lambda$  was determined as per Stein's Unbiased Risk Estimate (SURE), as SURE threshold selection rules are more conservative as expressed in [9]. This threshold value was used for all denoising schemes, except for *s7*, where it was selected by trial and error. Figure 4 shows the plots of the noise free synthetic signal and the denoised signals by different schemes. The values of  $E$ , SNR, and RMSE for the denoised signals are given in Table 2. The objective of signal processing was to increase the SNR and lower the RMSE of a corrupted signal. From Table 2, it can be observed that *s7* gave high SNR and low RMSE. The peaks in the original synthetic signal (representing the bearing fault impulses) were identifiable more clearly in *s7* denoising scheme when compared to the other schemes (refer to Figure 4).

### 4. Wavelet Based Denoising of Real-Time Vibration Signal

In the second part of the work in this paper, the objective was to apply the seven schemes of denoising discussed in

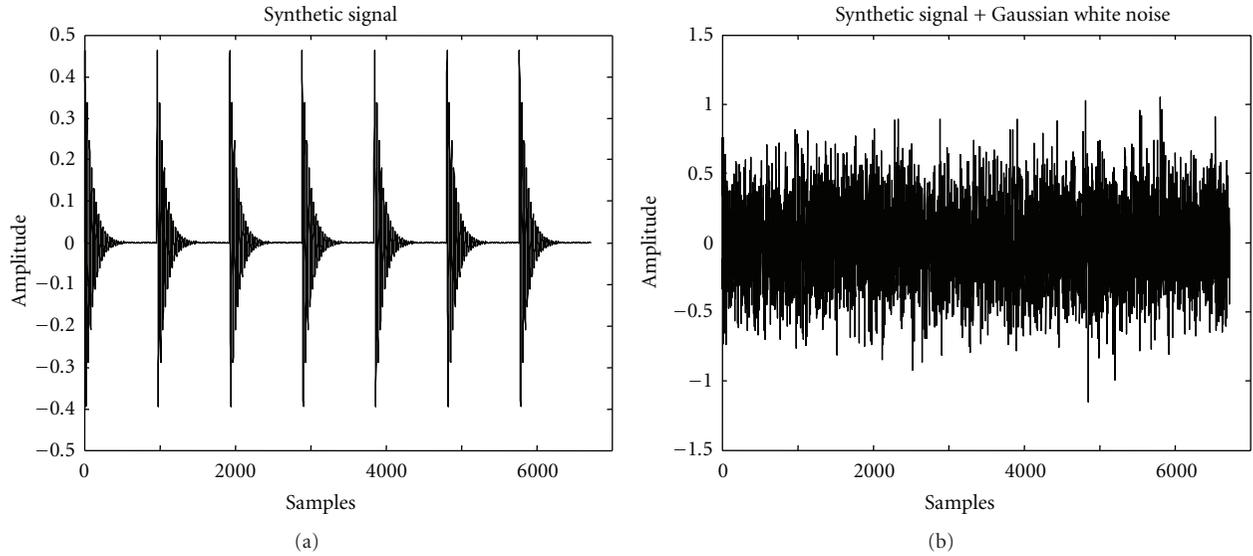


FIGURE 2: (a) Plot of a synthetic signal simulating a defective bearing vibration signal free of noise. (b) Plot of a signal corrupted with zero-mean Gaussian white noise.

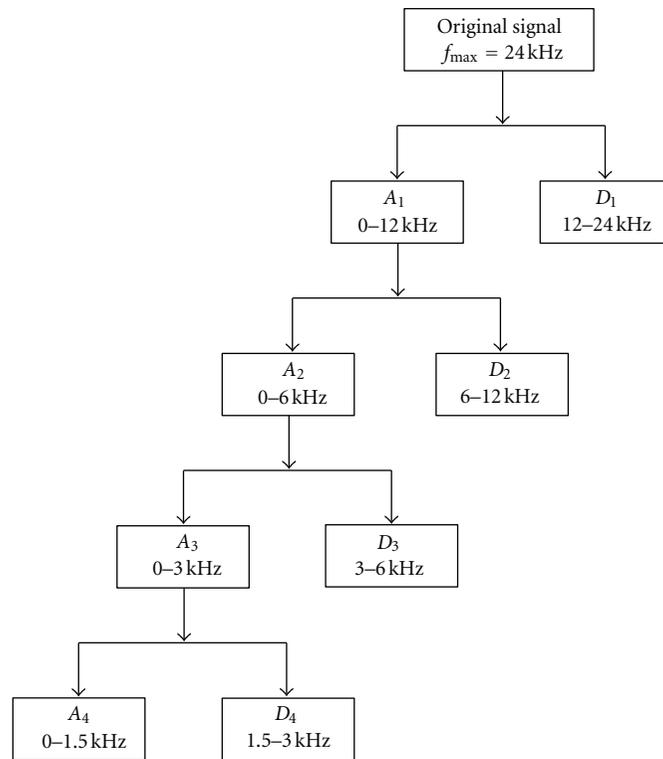


FIGURE 3: Discrete wavelet decomposition of bearing vibration signal.

the previous section to the vibration signals collected from a customized bearing test rig for four bearing conditions ( $N$ : normal bearing,  $IR$ : bearing with defect on inner race,  $B$ : bearing with defect on Ball, and  $OR$ : bearing with defect on an outer race). A schematic diagram of the customized test rig used for extracting the bearing vibration signals is shown in Figure 5. It consisted of a horizontal shaft

mounted on a support bearing (right) and a test bearing (left). A radial load on the test bearing was applied through a hydraulic loading arrangement. The accelerometers were mounted on the horizontal and the vertical surfaces of the test bearing housing ( $X$  &  $Y$ ). The vibration signals measured through the accelerometers were acquired at a rate of 48000 samples per second and stored in the computer through

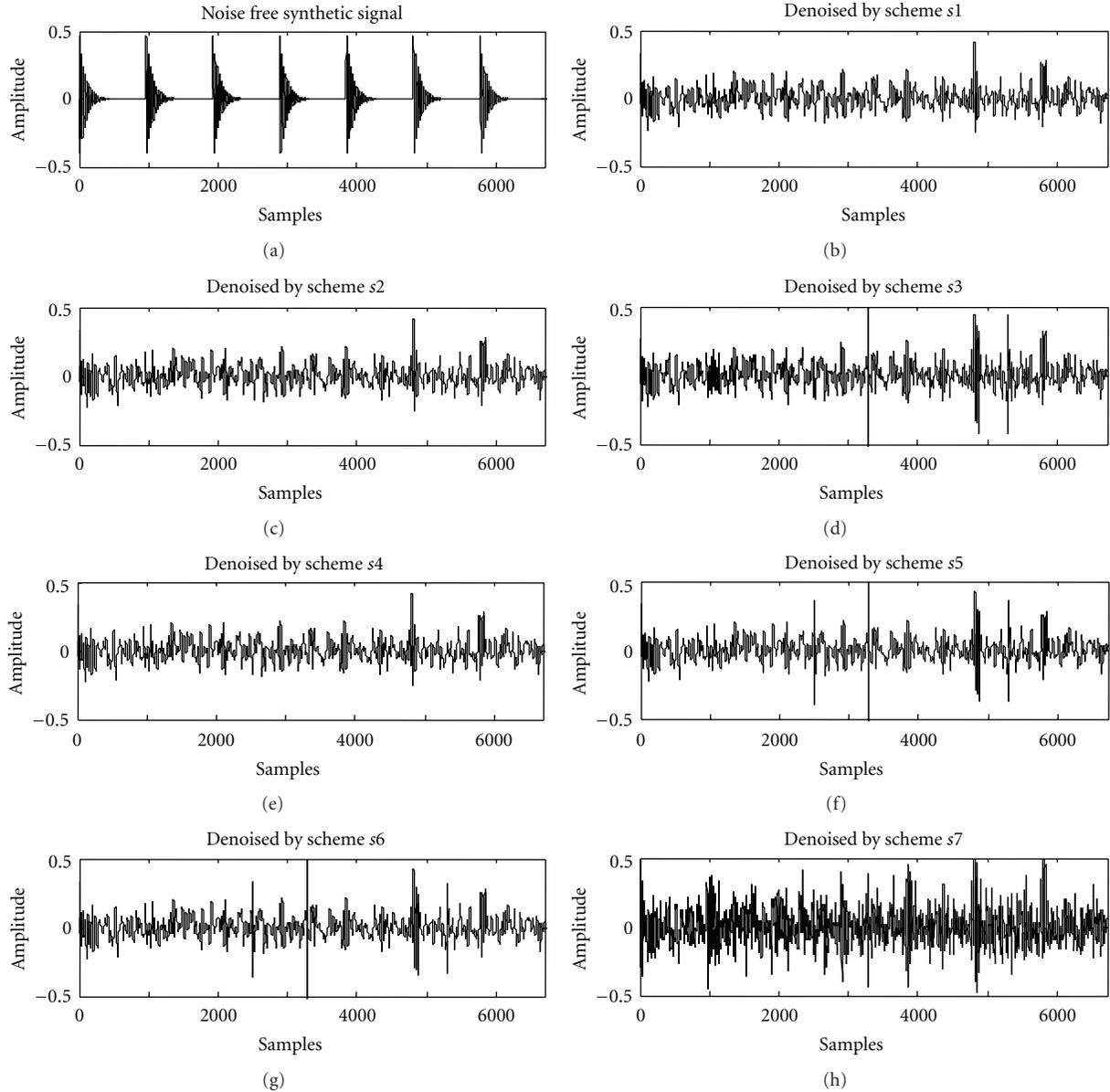


FIGURE 4: Plots of noise free synthetic signal and denoised signals by different schemes.

the DAQ system. The acceleration signals were collected for 5.08 seconds from a 6205 deep groove ball bearing under a radial load of 1.7 kN and shaft speeds of 356 and 622 rpm. The signals collected from accelerometer-*X* were considered for analysis, as the signals acquired in *Y*-direction were not very sensitive to the bearing condition. Each trial of the experiment resulted in a data vector of size  $250000 \times 1$ .

Figure 6(a) shows the raw vibration signal collected from a bearing with the OR defect for a load of 1.7 kN and a speed of 622 rpm and Figure 6(b) shows the plot of the denoised signal using scheme *s7*. It is clear from the figures that the selected denoising scheme has been effective in representing the original signal with reduced noise (SNR-1.9279, RMSE 0.0261).

## 5. Feature Extraction

Each denoised vibration signal ( $250000 \times 1$ ) was divided into 50 nonoverlapping bins each with 5000 data. From each bin, 30 features were extracted out of which features 1 to 17 ( $T_1$  to  $T_{17}$ ) were the statistical time domain features and features 18 to 30 ( $F_1$  to  $F_{13}$ ) were the statistical frequency domain features. This formed a single pattern. Hence, for four conditions of the bearing, two speed conditions and one load condition, a total of 400 patterns ( $50 \times 8$ ) were extracted. The feature set matrix consisted of 30 features  $\times$  400 patterns. Each feature was normalized, by dividing each element of the feature by the feature maxima, so as to attain values between 0 and 1. The patterns of the matrix were thoroughly mixed, out of which 300 patterns (75%) were

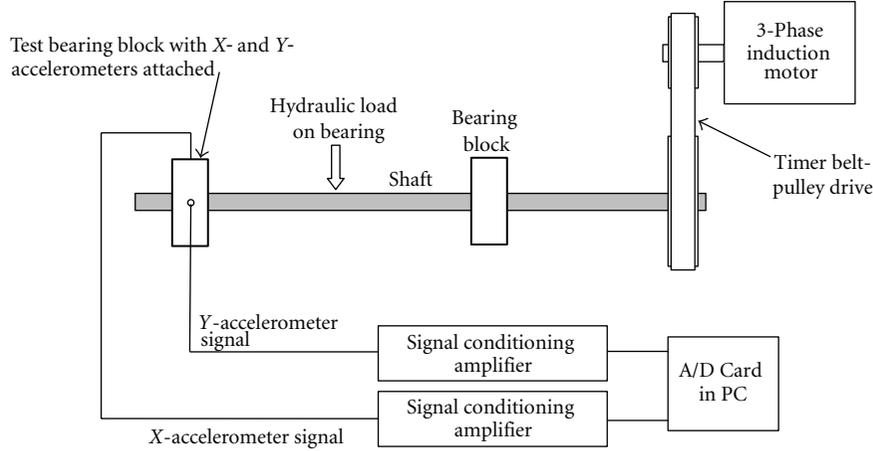
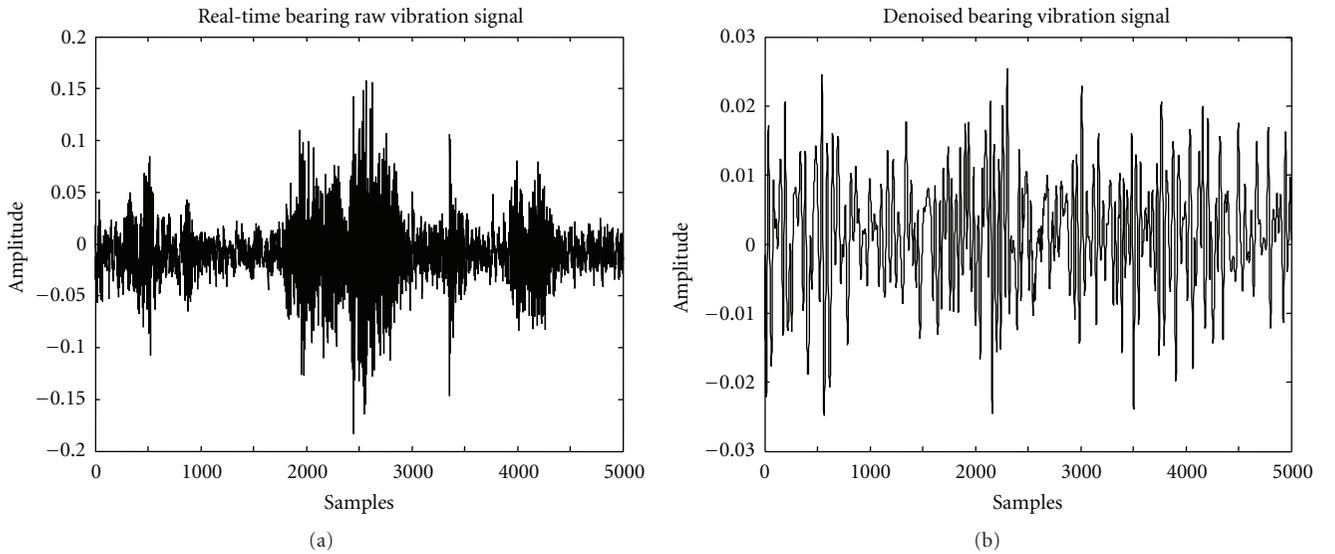


FIGURE 5: Schematic diagram of the test rig.


 FIGURE 6: (a) Raw vibration signal from bearing with OR defect. (b) Denoised vibration signal from bearing with OR defect using  $s7$ .

used in the training data set and the remaining 100 patterns (25%) in the test data set. A list of the features extracted from the denoised vibration signal is given in Table 3.

## 6. Feature Dimensionality Reduction

The FC has been used as a DRT in this work. The criterion for the FC is based on computing the “separation distance” between the two classes of interest and it depends upon the mean and the standard deviation of the two classes. The separation distance between the two classes as per the FC is given in (12) as suggested by Yen and Lin [20]:

$$J_k^{P,Q} = \frac{|\text{Mean}(t_k^P) - \text{Mean}(t_k^Q)|^2}{[\text{Std}(t_k^P)]^2 + [\text{Std}(t_k^Q)]^2}, \quad (12)$$

where  $J_k^{P,Q}$  is a measure of the Fisher’s Separation Distance between the two classes of the bearing  $P$  and  $Q$  for the  $k$ th

feature ( $P$  and  $Q$  each may be  $N$ ,  $IR$ ,  $B$ , and  $OR$  defect).  $\text{Mean}()$  and  $\text{Std}()$  are the mean and the standard deviation. For the four-class problem in this study, the summation of the pairwise combinations  $J_k^{N,IR}$ ,  $J_k^{N,OR}$ ,  $J_k^{N,B}$ ,  $J_k^{IR,OR}$ ,  $J_k^{IR,B}$ , and  $J_k^{OR,B}$  has been taken to estimate the Fisher’s Discriminant Power (FDP)  $F_k$  of a specific feature  $t_k$  [20]:

$$F_k = J_k^{N,IR} + J_k^{N,OR} + J_k^{N,B} + J_k^{IR,OR} + J_k^{IR,B} + J_k^{OR,B}, \quad (13)$$

where  $F$  is a vector of FDPs.

The features with higher values of the FDPs form sensitive inputs to the ANN/SVM classifiers. The FDPs computed for all the 30 features have been arranged in a descending order, resulting in a vector  $F^*$ . In this paper a new method of selecting the number of sensitive features based on a threshold value  $\theta$  has been proposed. The expression to compute  $\theta$  is given in (14):

$$\theta \approx \frac{\sum_{k=1}^s F_k^*}{\sum_{k=1}^f F_k}, \quad (14)$$

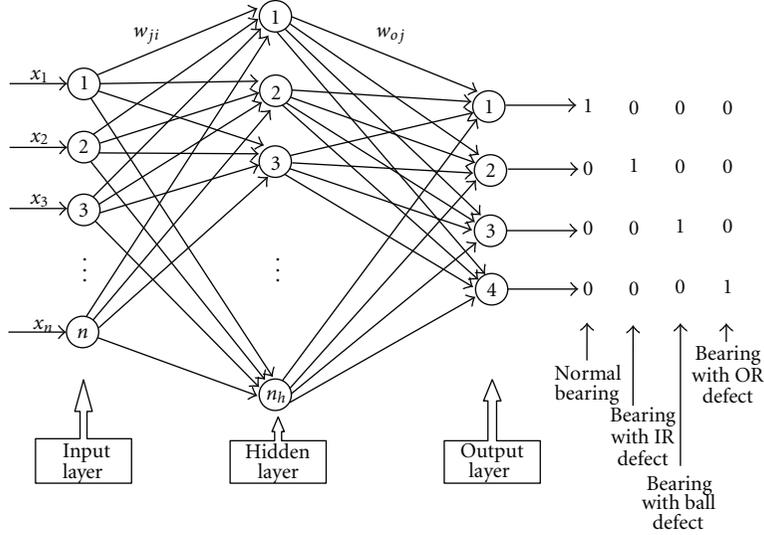


FIGURE 7: Structure of the MLPNN architecture.

TABLE 2: Energy, SNR, and RMSE of synthetic signal, corrupted signal, and denoised signals.

Signal type	Energy	SNR	RMSE
Synthetic signal	41.895919	—	—
Corrupted signal	510.773031	0.793049	0.264839
Signal denoised by $s1$	40.005523	0.824919	0.264418
Signal denoised by $s2$	40.005523	0.824919	0.264418
Signal denoised by $s3$	48.305217	1.539327	0.255139
Signal denoised by $s4$	40.005523	0.824919	0.264418
Signal denoised by $s5$	44.967623	0.930983	0.263019
Signal denoised by $s6$	44.136735	0.930165	0.263030
Signal denoised by $s7$	75.604166	4.712617	0.217706

where  $f$  is the total number of features extracted (30 in this work) and  $s$  is the number of selected features such that the sum of the  $s$  largest FDPs divided by the sum of all the FDPs is approximately equal to  $\theta$ . In this work, a threshold value of  $\theta = 0.85$  was chosen. Table 4 shows the FDPs for signals denoised by seven schemes. It can be seen that different denoising schemes have selected different numbers of features ( $s$ ) based on the threshold value of  $\theta = 0.85$ .

## 7. Performance of the Denoising Schemes Based on the ANN/SVM

In order to evaluate the performance of the different denoising schemes, the ANN/SVM classifiers were trained and tested using two types of inputs based on the features extracted from the denoised signals, namely, (i) the use of all the 30 features as inputs and (ii) the use of the features selected by the FC as inputs. A binary scheme of classification was used to define the bearing condition at the output of

both the ANN and the SVM, namely,  $N$  (1 0 0 0),  $IR$  (0 1 0 0),  $B$  (0 0 1 0), and  $OR$  defect (0 0 0 1) to denote the four-class condition of the bearings.

**7.1. ANN Classifier.** The ANN adopted here was the Multi-layer Perceptron Neural Network (MLPNN) which uses back propagation algorithm for training. Only one hidden layer with different numbers of neurons in hidden layer,  $n_h = 5, 10, 15, 20, 25$  and  $30$  were used. The sigmoid activation function is used in the hidden and the output layer. A mean square error of  $10^{-6}$ , a minimum gradient of  $10^{-10}$ , and maximum number of epochs of 500 are used. The training process would stop if any one of these conditions were met. The initial weights and biases of the network were fixed randomly. The MLPNN was implemented by using the MATLAB Neural Network Toolbox. Figure 7 shows the structure of the MLPNN used, where  $x_1, x_2, \dots, x_n$  are the inputs (features),  $n_h$  are the number of nodes in the hidden layer, and  $w_{ji}$  and  $w_{oj}$  are the connection weights between the input-hidden layers and the hidden-output layers, respectively.

The performance of the MLPNN classifier for the two types of inputs (all the features and the features selected by FC) extracted from the denoised signals is shown in Table 5. It is clear from the table that, for signals denoised using scheme  $s7$ , the accuracies on training and test data were higher compared to other schemes. The accuracies for different numbers of neurons in the hidden layer,  $n_h = 15, 20, 25$ , and  $30$  were comparatively lower and therefore have not been reported in the table.

**7.2. SVM Classifier.** The SVM classifier used in this work was based on the customized MATLAB tool box provided in [27]. In this work, the SVM was trained and tested for different values of the regularization parameter  $\gamma$  and the kernel width

TABLE 3: Time and frequency domain features extracted for the study.

Sl. number	Feature
(1)	$T_1 = n^{-1} \sum_{i=1}^n x_i$
(2)	$T_2 = \left[ n^{-1} \sum_{i=1}^n x_i^2 \right]^{1/2}$
(3)	$T_3 = \left[ (n-1)^{-1} \sum_{i=1}^n (x_i - T_1)^2 \right]^{1/2}$
(4)	$T_4 = \frac{1}{2} [\text{Max}(x_i) - \text{Min}(x_i)]$
(5)	$T_5 = \left( n^{-1} \sum_{i=1}^n \sqrt{ x_i } \right)^2$
(6)	$T_6 = n^{-1} \sum_{i=1}^n  x_i $
(7)	$T_7 = (n-1)^{-1} \sum_{i=1}^n \left( \frac{x_i - T_1}{T_3} \right)^3$
(8)	$T_8 = (n-1)^{-1} \sum_{i=1}^n \left( \frac{x_i - T_1}{T_3} \right)^4$
(9)	$T_9 = \frac{T_4}{T_2}$
(10)	$T_{10} = \frac{T_4}{T_5}$
(11)	$T_{11} = \frac{T_2}{T_6}$
(12)	$T_{12} = \frac{T_4}{T_6}$
(13)	$T_{13} = \frac{T_3}{T_6}$
(14)	$T_{14} = [\log(T_3)]^{-1} \sum_{i=1}^n \log( x_i  + 1)$
(15)	$T_{15} = \sqrt{\frac{T_5}{T_3}}$
(16)	$T_{16} = -\sum_{i=1}^n \log \left\{ [T_3 \sqrt{2\pi}]^{-1} \exp \left[ \frac{-(x_i - T_1)^2}{2T_3^2} \right] \right\}$
(17)	$T_{17} = -\sum_{i=1}^n \log \left[ \beta \eta^{-\beta}  x_i ^{\beta-1} \exp \left\{ \frac{ x_i }{\eta} \right\}^{\beta} \right],$

where  $\beta$  is the shape factor and  $\eta$  is the scale factor

(18)	$F_1 = N^{-1} \sum_{k=1}^N S(k)$
(19)	$F_2 = (N-1)^{-1} \sum_{k=1}^N [S(k) - F_1]^2$
(20)	$F_3 = N^{-1} F_2^{-1.5} \sum_{k=1}^N [S(k) - F_1]^3$

TABLE 3: Continued.

Sl. number	Feature
(21)	$F_4 = N^{-1} F_2^{-2} \sum_{k=1}^N [S(k) - F_1]^4$
(22)	$F_5 = \left[ \sum_{k=1}^N f_k S(k) \right] \left[ \sum_{k=1}^N S(k) \right]^{-1}$
(23)	$F_6 = \left\{ N^{-1} \sum_{k=1}^N [f_k - F_5]^2 S(k) \right\}^{1/2}$
(24)	$F_7 = \left\{ \left[ \sum_{k=1}^N f_k^2 S(k) \right] \left[ \sum_{k=1}^N S(k) \right]^{-1} \right\}^{1/2}$
(25)	$F_8 = \left\{ \left[ \sum_{k=1}^N f_k^4 S(k) \right] \left[ \sum_{k=1}^N f_k^2 S(k) \right]^{-1} \right\}^{1/2}$
(26)	$F_9 = \left[ \sum_{k=1}^N f_k^2 S(k) \right] \left\{ \sum_{k=1}^N S(k) \sum_{k=1}^N f_k^4 S(k) \right\}^{-1/2}$
(27)	$F_{10} = \frac{F_6}{F_5}$
(28)	$F_{11} = N^{-1} F_6^{-3} \sum_{k=1}^N (f_k - F_5)^3 S(k)$
(29)	$F_{12} = N^{-1} F_6^{-4} \sum_{k=1}^N (f_k - F_5)^4 S(k)$
(30)	$F_{13} = N^{-1} F_6^{-0.5} \sum_{k=1}^N (f_k - F_5)^{1/2} S(k)$

\* Features 1 to 17 ( $T_1$  to  $T_{17}$ ) are statistical features extracted from data in time domain and 18 to 30 ( $F_1$  to  $F_{13}$ ) are features extracted from data in the frequency domain.  $n$  is the number of data points in the time domain signal,  $x_i$  is the acceleration amplitude of  $i$ th data point in the time domain signal,  $N$  is the number of lines in the frequency spectrum,  $S(k)$  is the amplitude of the  $k$ th line in the frequency spectrum,  $f_k$  is the frequency value of the  $k$ th line in the frequency spectrum [18, 19].

$\sigma^2$ . Parameter  $\gamma$  was varied from 6 to 10 in steps of 1 and  $\sigma^2$  was varied from 3 to 4 in steps of 0.25.

The performance of the SVM classifier for the two types of inputs (all the features and the features selected by the FC) extracted from the denoised signals is shown in Table 6. It is clear from the table that, for signals denoised using scheme  $s7$ , the accuracies on the training and the test data were higher compared to the other schemes. The accuracies for the other values of  $\gamma$  and  $\sigma^2$  were comparatively lower and are therefore not reported in the table.

## 8. Discussion

The focus of this work is to evaluate seven different wavelet based denoising schemes. In order to ascertain the effectiveness of the schemes, a corrupted synthetic signal simulating the real-time bearing vibration signal is used. By observation and also based on the SNR and RMSE, scheme  $s7$  is found to be the most effective in denoising the corrupted signal. In order to evaluate its performance on a real-time bearing vibration signal, signals collected from bearings under four

TABLE 4: FDP and selected features by FC.

FDPs arranged in a descending order for signals denoised by seven schemes													
s1		s2		s3		s4		s5		s6		s7	
s = 5		s = 5		s = 6		s = 6		s = 4		s = 4		s = 9	
#	$F^*$	#	$F^*$	#	$F^*$	#	$F^*$	#	$F^*$	#	$F^*$	#	$F^*$
1	419.43	1	419.43	1	419.43	1	419.43	1	419.43	1	419.43	1	419.43
17	11.13	17	11.13	17	11.17	17	11.13	17	11.12	17	11.12	17	11.12
21	7.36	21	7.36	21	7.38	21	7.36	21	7.36	21	7.37	26	8.56
5	7.21	5	7.21	5	7.25	5	7.21	5	7.22	5	7.22	24	8.15
20	7.11	20	7.11	20	7.21	20	7.11					21	7.31
				6	6.33	6	6.31					5	7.16
												20	7.04
												22	6.94
												6	6.29
$\sum_{k=1}^{k=s} F_k^*$	452.24		452.24		458.77		458.55		445.13		445.14		482
$\sum_{k=1}^{k=f} F_k$	533.87		533.87		531.85		533.87		519.46		520.73		565.51
$\theta$	0.85		0.85		0.86		0.86		0.86		0.85		0.85

TABLE 5: Performance of the MLPNN classifier.

Scheme	Number of neurons in hidden layer	Epochs	All the 30 features as input			Features selected by FC as inputs		
			Training accuracy (%)	Test accuracy (%)	Epochs	Training accuracy (%)	Test accuracy (%)	
s1	5	30	100	95.67	82	100.00	88.75	
	10	35	100	92.67	78	100.00	89.75	
s2	5	36	100	93.67	200	98.83	92.25	
	10	47	100	94.67	63	100.00	89.25	
s3	5	41	100	94.33	38	100.00	95.25	
	10	61	100	90.33	163	100.00	90.50	
s4	5	15	100	96.00	153	99.92	91.75	
	10	42	100	94.00	102	100.00	86.50	
s5	5	61	100	93.67	157	99.92	92.50	
	10	182	100	86.00	92	100.00	88.00	
s6	5	207	99.11	90.00	152	99.92	92.25	
	10	80	100	87.33	143	100.00	89.25	
s7	5	22	100	97.67	18	100.00	95.50	
	10	52	100	92.33	124	100.00	91.25	

TABLE 6: Performance of SVM classifier.

Scheme	All the 30 features as input				Features selected by FC as inputs			
	$\gamma$	$\sigma^2$	Training accuracy (%)	Test accuracy (%)	$\gamma$	$\sigma^2$	Training accuracy (%)	Test accuracy (%)
s1	6	3.5	68.67	58.00	6	3	68.67	57.00
s2	6	3.5	68.67	58.00	6	3	68.67	57.00
s3	6	3.75	78.33	67.00	6	3	75.00	63.00
s4	6	3.5	68.67	58.00	6	3	68.67	57.00
s5	6	3.0	67.67	57.00	9	3	71.67	64.00
s6	6	3.0	67.67	57.00	10	3	71.67	64.00
s7	7	3.5	86.00	80.00	8	3	85.33	84.00

TABLE 7: The effectiveness of the denoising scheme  $s7$  using the ANN and the SVM.

	All feature set		Reduced feature set	
	Training accuracy (%)	Test accuracy (%)	Training accuracy (%)	Test accuracy (%)
ANN	100.00	97.67	100.00	95.50
SVM	86.00	80.00	85.33	84.00

conditions and single load and two speeds were subjected to the same denoising schemes. The denoised signals were used for extracting the features in the time domain and the frequency domain. The features extracted were subjected to dimensionality reduction using the FC. Both the reduced feature set selected by the FC and the complete feature set were used as inputs to the ANN and the SVM classifiers for comparing the performance of the different denoising schemes. Table 7 shows the performance of the ANN and the SVM classifier for the denoised vibration signal using the  $s7$  scheme. It is clear from the table that the denoising scheme  $s7$  resulted in more than 95% accuracy on the test data using the ANN and more than 80% accuracy on the test data using the SVM. Also for  $s7$  scheme, the reduced feature set obtained using FC resulted in a higher performance in terms of the test data accuracy and the number of epochs when compared to the use of all the features. The proposed method of selecting the number of sensitive features from the vibration data obtained using different denoising schemes based on a threshold  $\theta$  has been successful. Therefore, the DRT like the FC can be effectively used in improving the performance of the ANN and the SVM classifiers. Hence  $s7$  scheme is found to be effective in denoising the real-time vibration signals, when compared to the other denoising schemes.

## 9. Conclusions

This paper presents the evaluation of the effectiveness of the wavelet based denoising schemes using the ANN and the SVM classifiers applied to the bearing condition classification problem. Seven different denoising schemes selected from an extensive literature survey were used for denoising a synthetic corrupted signal resembling a bearing vibration signal. Based on the SNR and the RMSE, the best denoising scheme was selected. This scheme has been applied, along with the other schemes for denoising the real-time vibration signals collected from an REB test rig for four different bearing conditions, one load and two speeds. The features extracted from the denoised signal in the time domain and the frequency domain have been used as inputs to the ANN and the SVM classifiers. In order to reduce the dimension of the feature set, FC is used and the reduced feature set is also used as inputs to the ANN and the SVM. The proposed method of selecting the reduced number of features based on the threshold  $\theta$  is found to be effective. It is found that the best denoising scheme selected based on the synthetic signal performed better (in terms of the

classification accuracies and the number of epochs) with the feature set extracted from bearing vibration signals, when compared to the other denoising schemes. Hence, it can be concluded that  $s7$  scheme can be effectively used to denoise the bearing vibration signals for an efficient classification of its condition.

## Acknowledgment

The work reported in this paper is a part of a research project funded by the Visvesvaraya Technological University, Belgaum, Karnataka, India (Project number VTU/Aca-RGS/2008-2009/7200).

## References

- [1] N. Tandon and A. Choudhury, "Review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings," *Tribology International*, vol. 32, no. 8, pp. 469–480, 1999.
- [2] M. S. Patil, J. Mathew, and P. K. R. Kumar, "Bearing signature analysis as a medium for fault detection: a review," *Journal of Tribology*, vol. 130, no. 1, Article ID 014001, 2008.
- [3] Z. Huaigang, W. Zhibin, and Z. Ying, "Analysis of signal denoising method based on an improved wavelet thresholding," in *Proceedings of the 9th International Conference on Electronic Measurement and Instruments (ICEMI '09)*, pp. 1987–1990, IEEE, August 2009.
- [4] H. T. Fang and D. S. Huang, "Wavelet de-noising by means of trimmed thresholding," in *Proceedings of the 5th World Congress on Intelligent Control and Automation (WCICA '04)*, pp. 1621–1624, Hangzhou, China, June 2004.
- [5] Y. Lin and J. Cai, "A new threshold function for signal denoising based on wavelet transform," in *Proceedings of the International Conference on Measuring Technology and Mechatronics Automation (ICMTMA '10)*, pp. 200–203, IEEE.
- [6] G. Zhang, J. Wu, and Z. Cui, "Application of wavelet thresholding de-noising in DSA," in *Proceedings of the International Symposium on Information Science and Engineering (ISISE '08)*, pp. 130–134, IEEE, December 2008.
- [7] L. Cai-lian, S. Ji-xiang, and K. Yao-hong, "Adaptive image denoising by a new thresholding function," in *Proceedings of the 6th International Conference on Wireless Communications, Networking and Mobile Computing*, pp. 1–5, IEEE, September 2010.
- [8] Z. K. Peng and F. L. Chu, "Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography," *Mechanical Systems and Signal Processing*, vol. 18, no. 2, pp. 199–221, 2004.
- [9] M. Michel, Y. Misiti, G. Oppenheim, and J. M. Poggi, "Wavelet Toolbox 4 User's Guide," 2010.
- [10] X. Wang, Y. Zi, and Z. He, "Multiwavelet denoising with improved neighboring coefficients for application on rolling bearing fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 25, no. 1, pp. 285–304, 2011.
- [11] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, 1995.
- [12] C. C. Wang, Y. Kang, P. C. Shen, Y. P. Chang, and Y. L. Chung, "Applications of fault diagnosis in rotating machinery by using

- time series analysis with neural network,” *Expert Systems with Applications*, vol. 37, no. 2, pp. 1696–1702, 2010.
- [13] J. Zarei, “Induction motors bearing fault detection using pattern recognition techniques,” *Expert Systems with Applications*, vol. 39, no. 1, pp. 68–73, 2012.
- [14] P. K. Kankar, S. C. Sharma, and S. P. Harsha, “Fault diagnosis of ball bearings using machine learning methods,” *Expert Systems with Applications*, vol. 38, no. 3, pp. 1876–1886, 2011.
- [15] V. N. Vapnik, “An overview of statistical learning theory,” *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 988–999, 1999.
- [16] Y. Yang, D. Yu, and J. Cheng, “A fault diagnosis approach for roller bearing based on IMF envelope spectrum and SVM,” *Measurement*, vol. 40, no. 9-10, pp. 943–950, 2007.
- [17] V. Sugumaran, V. Muralidharan, and K. I. Ramachandran, “Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing,” *Mechanical Systems and Signal Processing*, vol. 21, no. 2, pp. 930–942, 2007.
- [18] B. Sreejith, A. K. Verma, and A. Srividya, “Fault diagnosis of rolling element bearing using time-domain features and neural networks,” in *Proceedings of the IEEE Region 10 Colloquium and 3rd International Conference on Industrial and Information Systems (ICIIS '08)*, vol. 409, pp. 1–6, Kharagpur, India, December 2008.
- [19] Y. Lei, Z. He, and Y. Zi, “A new approach to intelligent fault diagnosis of rotating machinery,” *Expert Systems with Applications*, vol. 35, no. 4, pp. 1593–1600, 2008.
- [20] G. G. Yen and K.-C. Lin, “Wavelet packet feature extraction for vibration monitoring,” *IEEE Transactions on Industrial Electronics*, vol. 47, no. 3, pp. 650–667, 2000.
- [21] M. J. Fuente, D. Garcia-Alvarez, G. I. Sainz-Palmero, and T. Villegas, “Fault detection and identification method based on multivariate statistical techniques,” in *Proceedings of the IEEE Conference on Emerging Technologies and Factory Automation (ETFA '09)*, esp, September 2009.
- [22] L. H. Chiang, M. E. Kotanchek, and A. K. Kordon, “Fault diagnosis based on Fisher discriminant analysis and support vector machines,” *Computers and Chemical Engineering*, vol. 28, no. 8, pp. 1389–1401, 2004.
- [23] X. C. Tang and Y. Li, “Monitoring and fault diagnosis using fisher discriminant analysis,” in *Proceedings of the 6th International Conference on Machine Learning and Cybernetics (ICMLC '07)*, pp. 1100–1105, Hong Kong, August 2007.
- [24] L. B. Jack and A. K. Nandi, “Fault detection using support vector machines and artificial neural networks, augmented by genetic algorithms,” *Mechanical Systems and Signal Processing*, vol. 16, no. 2-3, pp. 373–390, 2002.
- [25] B. Samanta, K. R. Al-Balushi, and S. A. Al-Araimi, “Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection,” *Engineering Applications of Artificial Intelligence*, vol. 16, no. 7-8, pp. 657–665, 2003.
- [26] A. Saxena and A. Saad, “Evolving an artificial neural network classifier for condition monitoring of rotating mechanical systems,” *Applied Soft Computing Journal*, vol. 7, no. 1, pp. 441–454, 2007.
- [27] J. A. K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle, “LS-SVMlab: a MATLAB/C toolbox for least squares support vector machines,” 2002, <http://www.esat.kuleuven.ac.be/sista/lssvmlab>.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

