Rolling bearings fault diagnosis and classification based on optimal wavelet subband selection and deep neural network

Rakesh Kumar Jha (✉ rakesh.20db@gmail.com)

Rajiv Gandhi Proudhyogiki Vishwavidyalaya <https://orcid.org/0000-0002-6568-8083>

Preety D Swami

Rajiv Gandhi Proudhyogiki Vishwavidyalaya

Research Article

Keywords: Autoencoder, fault diagnosis, Shannon entropy, softmax classifier, stationary wavelet transform.

Posted Date: March 23rd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-342393/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Rolling bearings fault diagnosis and classification based on optimal wavelet subband selection and deep neural network

Rakesh Kumar Jha^{1, a *}, Preety D Swami^{1, b}

¹ Department of Electronics & Communication Engineering, RGPV, Bhopal- 462033, India

^a email-address: rakesh.20db@gmail.com

^b email-address: preetydswami@gmail.com

Abstract

Time-frequency analysis plays a vital role in fault diagnosis of nonstationary vibration signals acquired from mechanical systems. However, the practical applications face the challenges of continuous variation in speed and load. Apart from this, the disturbances introduced by noise are inevitable. This paper aims to develop a robust method for fault identification in bearings under varying speed, load and noisy conditions. An Optimal Wavelet Subband Deep Neural Network (OWS-DNN) technique is proposed that automatically extracts features from an optimal wavelet subband selected on the basis of Shannon entropy. After denoising the optimal subband, the optimal subbands are dimensionally reduced by the encoder section of an autoencoder. The output of the encoder can be considered as data features. Finally, softmax classifier is employed to classify the encoder output. The vibration signals were recorded on a machinery fault simulator setup for various combinations of speed and load for healthy and faulty bearings. The signals were subjected to various noise levels and the deep neural network was trained. The achieved experimental results reveal high accuracy in fault classification as compared to other techniques under comparison.

Keywords Autoencoder, fault diagnosis, Shannon entropy, softmax classifier, stationary wavelet transform.

1. Introduction

Machines play vital role in our lives and have become a part of our daily activities. All machines in industrial production rely on bearings for their functioning. To support the axial and radial loads, rolling bearings are employed. Bearing faults are responsible in 40-45% [1, 2] cases of failures in machines. Condition based machine health monitoring [3] system is best suited to prevent sudden breakdown of machine due to failure of rotating elements. Online machine health condition monitoring scheme uses vibration or acoustic signals [1, 3] generated by the rolling elements in operation. Much research, in the field of fault diagnosis and analysis of real time machine

health prognostics has been carried out. Earlier researchers used the statistical characteristics of extracted vibration signals separately in the time and frequency domains and then in the combined time-frequency domain. Time domain approach incorporates mainly statistical moment approach, time synchronous averaging (TSA) [4,5] and auto regression approaches. TSA is employed for unwanted noise suppression and unwanted periodic events which are synchronous to the signal. Frequency domain analysis relies on extraction and analysis of frequency domain features. Envelope detection [6] and cepstral analysis [7] are well known frequency domain techniques. Envelope detection relies

on demodulation of optimal band for efficient signal analysis. Selection of optimal demodulation subband is quite difficult. Antony [8] developed an algorithm named spectral kurtosis, which suggests most impulsive frequency band of the signal for demodulation. These methods are very simple and easy but less sensitive to nonlinear and nonstationary signals. To analyse nonlinear and nonstationary signals efficiently, the researchers then started to employ the Hilbert–Huang Transform (HHT) based empirical mode decomposition (EMD) [9-10] and Ensemble Empirical Mode Decomposition (EEMD) techniques [11-12]. Wavelets possess the property of adaption with the signal shape hence Wavelet Transform (WT) is a perfect tool for analysis of nonlinear and nonstationary signals. WT [13-20] is a multiresolution transform which decomposes the signals into wavelet subbands and furnishes the signal information in the time-frequency domain. The statistical features of wavelet coefficients in the form of central moments can be utilized to analyse and differentiate the conditions of healthy and faulty machinery components. Apart from that WT is also used for noise suppression. Thus WT has become an important tool for diagnosis of flaws and machine health monitoring. Several researchers have proposed flaws' diagnostics using entropy based feature parameters such as Approximation Entropy (ApEn) [12,21,22], Permutation Entropy (PE) [23-24], Sample Entropy (SaEn) [25] and Spectral Entropy (SE) [26] in the time-frequency domain. ApEn measures the regularity and can quantify the unpredictability in the rapid changes in a time-series (T-S) dataset. SaEn is a modification of ApEn and it overcomes two disadvantages of ApEn; it is independent of data length and has a simpler implementation. PE entropy measures the complexity based on permutation patterns in T-S data. SE is the entropy of normalized power spectral density of a signal

and can quantify the randomness of power spectrum computed for some duration.

Nowadays, the application of intelligent fault detection [14-20, 27-30] is widely used in machinery health monitoring systems. Machine learning approaches due to their easy and accurate fault diagnosis are widely acceptable for fault detection. K-nearest-neighbour (KNN), Support vector machines (SVM), Artificial Neural Networks (ANN) and Convolutional Neural Networks (CNN) are widely exploited for classification problems. Authors in [20] used higher order cumulants and WT for feature vectors subjected to KNN for classification of noisy vibration signals. K-mean clustering on PCA reduced statically features of the signal was used in [29] for classification of flaws. Authors in [6, 11, 29] followed the SVM approach of data classification. Authors in [19] employed the CNN approach for flaws' classification of signals with heavy noise. ANN [22, 28-29], a machine learning algorithm, is a data driven approach and is trained by the features of data sets extracted from the machine running in different conditions. The traditional ANN [30] works efficiently only when the features crafted for training are separable to a large extent. The feature selection part of ANN puts a constraint on its performance. Deep neural networks (DNN) [31-33] are used to eliminate this problem of manual selection of features. Hence, deep neural networks are widely adopted for classification problems.

The performance of above mentioned techniques degrades in harsh industrial environment. The poor signal to noise ratio (SNR) of incipient faults and lack of susceptibility to atmospheric noise deteriorates the performance of the algorithms. Considering such realistic problems, this paper proposes a data driven approach for fault detection of rolling bearings. The purpose of this paper is to design a fault classification algorithm that is

robust to variations in speed, load and noise. The proposed approach extracts information from an optimal subband of the signal decomposed using Stationary Wavelet Transform (SWT). The optimal subband is chosen on the basis of minimum Shannon entropy.

The 1D optimal subbands are converted to 2D matrices on the basis of sampling interval and rotation speed. The 2D matrices are then dimensionally reduced by the encoder section of the autoencoder. The output of the encoder is basically the hidden layer information of the signal and can be considered as the signal's features. The encoder

output is fed to the softmax classifier. The classifier finally distinguishes between the healthy and the faulty signals. The high degree of classification even for data corrupted with heavy noisy advocates the success of the algorithm.

The remaining part of the paper is organized as follows. The proposed OWS-DNN method of fault diagnosis is described in section 2. Section 3 reports the experiments and results. Validation of the proposed method is done in section 4 and performances are also compared with few other techniques. Section 5 concludes the paper.

2. Proposed OWS-DNN Method

Fig. 1. Framework of OWS-DNN algorithm

The framework of the proposed OWS-DNN fault diagnosis methodology is shown in Fig. 1. The aim is to

design a robust fault diagnosis algorithm that is independent of changes in speed, load and interfering

noise levels. This is done by mapping the signal to the frequency domain and finding some robust features from an optimal frequency band. All the further processing will be then carried out on the coefficients of this optimal frequency band only. The description of the proposed method is provided in the following subsections.

2.1 SWT decomposition

For the decomposition of signal into various subbands, SWT is employed. The conventional Discrete Wavelet Transform (DWT) suffers from aliasing problem due to lack of shift invariance. SWT overcomes this problem and serves as a strong candidate for denoising of signals. SWT is nondecimated DWT that does not downsample the sequence at each decomposition level. This results in better resolution but at the cost of bandwidth. For SWT decomposition, appropriate low pass filters (H) and high pass filters (G) are applied to the data at each level so that two sequences are produced at the next level. The approximation coefficients (CA) and the detailed coefficients (CD) are the outputs of the low pass filters and the high pass filters respectively.

As decimation is not applied, the length of the new sequence generated is equal to that of the original sequence and the length does not reduce as in DWT. The filters at each level are modified by zero padding or upsampling as shown in Fig. 2(a). For any signal s , for decomposition levels $j = 1, 2, \dots, J$, the sequences obtained after each level of decomposition can be defined in a recursive manner as [34]

$$CA_{j+1} = H_j CA_j \text{ and } CD_{j+1} = G_j CA_j \quad (1)$$

Fig. 2. (a) Upsampling of low-pass and high-pass filters

Fig. 2. (b) Signal decomposition at the first level

Fig. 2. (c) Signal decomposition at the j^{th} level.

The decomposition of a signal s at the first level is shown in Fig. 2(b) while decomposition at the j^{th} level is displayed in Fig. 2(c).

2.2 Selection of Optimal SWT Subband

For the selection of optimal subband the idea is to extract the subband that contains the maximum amount of information. For this the criterion of Shannon entropy is

considered. The entropy of a signal measures the amount of information embedded in it

For a randomly distributed event $X = \{x_i, i = 1, 2, 3 \dots n\}$ with probability distribution $p(x_i)$, the amount of information [35] is $1/\ln(p(x_i))$. The average information of the event [35-36] is mathematically represented as in Eq. (2).

$$H(X) = - \sum_{i=1}^n (p(x_i) \ln p(x_i)) \quad (2)$$

The term $H(X)$ is the Shannon entropy of random distribution of X which is a function of probability distribution $p(x_i)$.

The amount of information in a signal is dependent on noise. Adding noise N to a signal X ($S = X + N$) furnishes different entropies for different noise levels. For a signal X subjected to AWGN noises $N1$ and $N2$ having standard deviation $\sigma1$ and $\sigma2$, ($\sigma1 < \sigma2$), the resultant signals $S1$ and $S2$ with their respective entropies $H(s1)$ and $H(s2)$ hold the inequality relation $H(s1) > H(s2)$ [37]. Mathematically,

$$S1 = X + N1(\sigma1) \quad (3)$$

$$S2 = X + N2(\sigma2) \quad (4)$$

$$\text{If } \sigma1 < \sigma2 \text{ then } H(s1) > H(s2) \quad (5)$$

This behaviour can be observed from Fig. 3 which depicts the relation between entropies of optimal subbands, i.e., the minimum entropy amongst all the subbands for various noise levels.

For each of the healthy, inner race (IR), outer race (OR) and ball faults, with increase in noise standard deviation the minimum entropy goes on decreasing. This feature can efficiently separate the noisy data acquired and is thus helpful in classification for fault diagnosis.

Fig. 3. Entropy vs. SNR for bearings

Fig. 4. Entropies of approximation and detail subbands.

In the proposed work, the selection of number of levels of wavelet decomposition and thereafter selecting an optimal subband from the decomposed subbands is an important step. The subband that has the minimum entropy i.e. the maximum information is chosen to be the optimal subband. To decide the number of levels of wavelet decomposition, the entropies of all subbands are computed. It is observed that with the increase in decomposition level, initially the entropies of the subbands go on decreasing and after a certain level the entropy starts to increase. This level at which the entropy is minimum is then decided as the number of levels up to which the vibration signal is to be decomposed using SWT.

On the basis of experimentations done on vibration signals, selection of four decomposition levels (eight subbands) is decided for this work. This can be seen in Fig. 4 that shows the behaviour of the entropy values of approximation and detail subbands at five decomposition

levels. For the approximation subbands, the entropies are decreasing but for the detail subbands initially decrease in the entropy values is observed up to level four and after that at level five, entropy has increased.

2.3 Denoising the optimal subband

The wavelet coefficients of the optimal subband will be used to generate the code using autoencoder for training the neural network. Denoising of this optimal subband helps in better classification of the faults. Thresholding of only those optimal subbands is done for which the noise standard deviation exceeds a threshold that is calculated empirically. Small amount of noise present improves classification results; hence, the optimal subbands having noise standard deviation below the threshold value are not denoised. For the coefficients C_{opt} , corresponding to the optimal subband, the noise standard deviation, σ_{TH} , is estimated based on median absolute deviation (MAD) [38] and can be defined as

$$\sigma = \frac{\text{Median Absolute Deviation } (C_{opt})}{0.6745} \quad (6)$$

Choice of MAD over mean standard deviation is due to its ability to get rid of outliers in non-uniform data sets. The threshold value is computed using universal thresholding and given by [38]

$$\lambda = \sigma \sqrt{2 * \ln(n)} \quad (7)$$

Where σ is median absolute deviation and n is the signal length.

The wavelet coefficients are denoised using hard thresholding of the optimal subband coefficients. Hard thresholding sets the coefficients of a signal to zero whose value is less than the threshold.

Hard thresholding can be represented mathematically as

$$x = \begin{cases} 0, & \text{for } x < \lambda \\ x, & \text{otherwise} \end{cases} \quad (8)$$

The parameter λ is the threshold value for the signal to be denoised.

2.4 Autoencoder

Autoencoder is an unsupervised neural network technique designed to efficiently compress the input data and learn the encoded data. Further it decodes the encoded data to reconstruct the replica of input data at the output with very little losses.

Fig. 5 shows the block diagram representation of an autoencoder. An encoder compresses the input data into a lower dimension called bottleneck or code. The code is used to train the neural network and the decoder reproduces the approximated input signal at the output.

The network is trained to minimize the loss function $\mathcal{L}(x, \hat{x})$ which is the difference between the original signal and the approximate signal.

A single layer autoencoder is shown in Fig. 6. The encoder, hidden layer and decoder are the primary components of an autoencoder

Fig. 5. Block diagram of Autoencoder

Fig. 6. Single hidden layer autoencoder

Let ψ be the activation function of the encoder which maps the random input vector $X \in \mathcal{R}^n$ to hidden layer $h \in \mathcal{R}^p$ such that

$$h = \psi(W.X + b) \quad (9)$$

Where h represents a vector comprising of latent variables, W is the weight matrix such that $W \in \mathcal{R}^{p \times n}$, X is the input matrix and b is the bias vector. The decoder having activation function φ , maps the code h into the approximated output signal $\hat{X}$.

$$\hat{X} = \varphi(W'.h + b') \quad (10)$$

Where $W' \in \mathcal{R}^{n \times p}$ is the weight matrix and b' is the bias vector of the output layer.

The activation function ψ and φ are intentionally chosen in such a way that the reconstructed signal should not be the exact replica of the original signal. The reconstructed signal should be approximation of the original input signal which removes the overfitting problem encountered in ANN. The cost error function can be expressed in terms of loss function and regularization [31] and are represented in eq. (11).

$$\phi(e) = \frac{1}{2} \|X - \hat{X}\|^2 + \Omega(h, X) \quad (11)$$

Here the first term is loss function $\mathcal{L}(X, \hat{X})$ and the second term $\Omega(h, X)$ is regularization which discourages memorization or overfitting. The regularization [32, 33] may be expressed as in eq. (11).

$$\Omega(h, X) = \lambda \sum_{i=1}^p \|\nabla_{x_i} \cdot h_i\|^2 \quad (12)$$

Where λ is the weight decay parameter and the difference operator ∇_x defined over random variable x of the i^{th} node.

In the proposed work, single hidden layer autoencoder was employed to reduce the dimension of the optimal wavelet subband signals. The code generated by the encoder was used as feature vectors and was fed to softmax classifier to distinguish and classify the healthy and faulty signals.

2.5 Softmax classifier

Softmax classifier, a linear classifier, predicts the output in terms of probability distribution of the classes. It uses the cross entropy function to update the weight of the layers. Cross entropy is a loss function used to measure the error between the actual and the predicted output. For probabilistic distributions the cross entropy between two distributions is given by eq. (13)

$$\mathcal{L}(X, \hat{X}) = - \sum_{i=1}^n x^{(i)} \log(\hat{x}^{(i)}) \quad (13)$$

Where, X and $\hat{X}$ are the original input signal and its approximated output signal. n is the total number of sample points of the original and reconstructed output signal.

For K classes of the input data X and weight W , the probabilistic output produced by the softmax classifier is given in eq. (14).

$$P(Y = K|X) = \frac{\exp(W_K^T \cdot X)}{\sum_i^N \exp(W_i^T \cdot X)} \quad (14)$$

Where $(\cdot)$ indicates the dot product and N be the total number of classes.

The softmax classifier was used in the last stage of DNN and it classifies the input signal using dimensionally reduced code of the autoencoder.

3. Experimental Results

3.1 Vibration data description

In this work, datasets from two machineries are taken for experimentation. First one is the machinery fault simulator in laboratory and the second is the Case Western Reserve University (CWRU) [39] bearing dataset available online. Details of these datasets are provided in the following subsections.

3.1.1 Experimental test rig

The original experimental vibration signal data was obtained at a sampling frequency of 25.6 KHz from accelerometers (model 356A16) mounted on a machinery fault simulator. The vibration data captured by accelerometer were recorded in personal computer using 4-channel data acquisition system (DAQ). Experimental setup of the machinery fault simulator is shown in Fig. 7. The faults were seeded on the rolling bearings. Measurements were recorded by reinstalling each faulty bearing on the motor shaft and it was then run at speed 1200 rpm with 3 load conditions corresponding to no load, medium load and heavy load.

For varying the loads, a 450 gm rotor was installed that acts as medium load and a 5 Kg rotor was installed as heavy load. The vibration signals of a healthy bearing and 3 faulty bearings (IR, OR, Ball faults) at no load and SNR levels 0 dB and -20 dB are shown in Fig. 8 and Fig. 9 respectively.

Fig. 7. Experimental setup of machinery fault simulator.

Fig. 8 Signal plot for healthy and faulty bearings at no load and $SNR = 0 \text{ dB}$

Fig. 9 Signal plot for healthy and faulty bearings at no load and $SNR = -20 \text{ dB}$

3.1.2 CWRU data set

The significance of the method proposed is highlighted by comparing it with some previous fault diagnosis methods developed by authors in [18-20]. These authors have validated the outcomes of their methods on the data set from the popular bearing data center of CWRU that is available online. Hence, for the purpose of comparison of

Fig. 10 Experimental test bench setup CWRU [39]

the proposed results with the existing available techniques, the proposed algorithm is also tested on the CWRU dataset. Fig. 10 shows the CWRU test bench setup.

The bearing dataset is acquired for bearing with model number 6205. Faults were seeded on the inner race, outer race and ball of the bearing by electro discharge machining. In this work, the bearing data that is considered is the one from the drive end under four load conditions acquired at a rate of 12K samples per second. Data files corresponding to fault size 0.007 inches are considered and their details are listed in the Appendix.

3.2 Generation of noisy data

Noise is inevitable in the real world measurements obtained from industries. For testing the proposed method's accuracy under noisy environment, the signals were contaminated with additive white Gaussian noise of different levels. The noise level is usually measured by the ratio of signal power to noise power defined as SNR and expressed as

$$SNR_{dB} = 10 \log_{10} \left(\frac{P_{signal}}{P_{noise}} \right) \quad (15)$$

The proposed algorithm is tested with noisy vibration signals having SNR in the range of 0 dB to -20 dB.

3.3 Signal decomposition using SWT and optimal subband selection

The heart of the proposed algorithm lies in the selection of one optimal subband from the many subbands obtained upon SWT decomposition of the vibration signal. It has been observed experimentally that decomposition upto four levels using *bior2.2* wavelet is sufficient for obtaining accurate results [40]. Decomposing a signal at four levels results in four approximation subbands ($CA1, CA2, CA3$, and $CA4$) and four detail subbands ($CD1, CD2, CD3$, and $CD4$). The optimal subband is selected on the basis of minimum entropy (maximum information). These subbands along with their entropies for healthy and faulty bearings under the working conditions of 1200 rpm, heavy load and 0 dB SNR are illustrated in Fig. 11 to Fig. 14. As shown in Fig. 11, amongst the eight subbands, the entropy of the third detail subband ($CD3$) is minimum. So, this subband is the optimal subband for the healthy bearing for the above mentioned working conditions. Similarly, for the faults in inner race, outer race and ball, the minimum entropies as shown in Fig. 12 to Fig. 14 can be observed in $CD4, CA4$, and $CA4$ respectively.

Fig. 11. Approximation and detail subbands of a healthy bearing.

Fig. 12. Approximation and detail subbands of a bearing with fault in the inner race.

Fig. 13. Approximation and detail subbands of a bearing with fault in the outer race

Fig. 14. Approximation and detail subbands of a bearing with ball fault.

Likewise, the investigation of vibration signals belonging to healthy or faulty bearings, at varying loads and at different noise levels can be done in a similar manner by selecting the optimum subband obtained from wavelet decomposition.

3.4 Fault diagnosis and classification

The autoencoder can be used for dimension reduction since the encoder compresses the input data and generates the lower dimensional encoded output which is referred to as the code. The code can be considered as features to train the classifier. For feature extraction using autoencoder, the one dimensional optimal subband is first denoised. Denoising is not required for signals with low level noise and it is observed that removing noise from such signals degrades the classification results. Only those subbands are denoised whose estimated standard deviation is greater than some threshold which is empirically computed as 0.2. The denoised subband is then normalized and converted to a two dimension matrix having dimension $R \times C$. The values of R and C depend on the length of the signal, rotating speed of the machine

and the sampling frequency. For a vibration signal of length l , rotating at S rpm, the shaft rotation frequency is $f_r = S/60$ Hz. Considering the sampling frequency to be f_s samples/sec, the number of samples per rotation is $n = f_s/f_r$. Number of columns of the matrix C is the number of samples per rotation and hence $C = n$ and number of rows is thus $R = l/n$. Each signal in this work is acquired at a sampling frequency of 25.6 KHz. Thus the length of the signal corresponding to a duration of 10 sec is $l = 256000$ samples. For a shaft rotating at $S = 1200$ rpm, the rotation frequency f_r is 20 Hz leading to a matrix of dimension 200×1280 .

This matrix is fed to the autoencoder having one hidden layer that consists of 100 nodes. The encoder converts the input matrix into corresponding code of size 200×100 as depicted in Fig.15. For each test condition, one healthy and 3 faulty cases (IR, OR and ball faults) are considered. So the input data matrix will be of size $4 * [200, 1280] = [800, 1280]$. The encoder yields the code of size $[800, 100]$ representing the complete test inputs.

Fig. 15. Fault classification using softmax classifier

The hidden layer is selected in such a way that its output is approximately 10% of the input data. If we reduce the data below this value, performance of the classifier deteriorates.

For CWRU data set, each healthy and faulty input matrix is of size $[300, 400]$. Hence the data input matrix in this case is $4 * [300, 400] = [1200, 400]$. The input data subjected to the encoder produces the output data of size $[1200, 40]$. Hidden layer size is kept 40 to reduce the input data by 90%. Thus, the output of the encoder is 10% of the input matrix.

The reduced column vectors can be considered as features of the classifier. The output of encoder is applied to the input of the softmax classifier as discussed in section 2.5.

For a rolling bearing the total number of different outputs are 4, representing healthy and faulty conditions as can be seen from Fig. 15.

4. Validation of the proposed OWS-DNN method

4.1 Validation of OWS-DNN on the experimental test rig

The fault detection performance of rolling bearings was evaluated over many scenarios that include varying loads and varying noise levels. 12 scenarios corresponding to healthy and faulty bearings rotating at speed 1200 rpm under no load, medium load and heavy load conditions and noise levels ($SNR = 0\text{ dB}$ to -20 dB) were considered.

For experimental test rig data sets, the size of the input data is $[800, 1280]$. The encoder reduced its size and produces an output matrix of size $[800, 100]$. 80% data

was used for training purpose and remaining 20% data was used for testing. Thus the size of training, testing data sets are $[640, 100]$ and $[160, 100]$ respectively.

Table 1 shows the classification results of the classifier at various loads and noise conditions. The accuracy of the classifier is 100% at no-load condition and noise levels upto -15 dB while at $SNR = -20\text{ dB}$ the accuracy rolls down to 95.6%. The classifier maintains 100% classification accuracy at medium load (450 gm) and heavy load (5 kg) with $SNR = 0\text{ dB}$. The classification accuracy decreases with increasing loads and noise levels as mentioned in Table 1. Fig. 16 shows the classification accuracy at no load and $SNR = 0\text{ dB}$.

4.2 Validation of OWS-DNN on CWRU data set

Apart from testing the algorithm on the experimental test rig, it is also tested on the CWRU data set to justify its universality.

For CWRU data set, the encoder produces the output matrix of size $[1200, 40]$. As done with the experimental test rig data sets, 80 percent of the encoded data was used for training purpose and remaining 20% of the encoded data was used for testing purpose. Thus the training matrix is of size $[960, 40]$ and testing data sets is of size $[240, 40]$. Table 2 shows the classification accuracies of the classifier under different load and varying noise conditions. Fig. 17 depicts the classification accuracy at no load (0 hp) and $SNR = 0\text{ dB}$.

The testing classification accuracy is 99.6% at no-load (0 hp) and $SNR = 0\text{ dB}$., while it is 97.1% at 1 hp load and $SNR = 0\text{ dB}$.. The classification accuracies are 95.0% and 85.0% at 2 hp and 3 hp loads with $SNR = 0\text{ dB}$ respectively. The performance of the classifier deteriorates with increasing load and noise levels.

		Confusion Matrix				
Output Class	Healthy	40 25.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	IR Fault	0 0.0%	40 25.0%	0 0.0%	0 0.0%	100% 0.0%
	OR Fault	0 0.0%	0 0.0%	40 25.0%	0 0.0%	100% 0.0%
	Ball Fault	0 0.0%	0 0.0%	0 0.0%	40 25.0%	100% 0.0%
		100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%
		Healthy	IR Fault	OR Fault	Ball Fault	Target Class

Fig. 16. Confusion matrix for rolling bearings at 1200 rpm under no load and $SNR = 0\text{ dB}$.

		Confusion Matrix				
Output Class	Healthy	60 25.0%	1 0.4%	0 0.0%	0 0.0%	98.4% 1.6%
	IR Fault	0 0.0%	59 24.6%	0 0.0%	0 0.0%	100% 0.0%
	OR Fault	0 0.0%	0 0.0%	60 25.0%	0 0.0%	100% 0.0%
	Ball Fault	0 0.0%	0 0.0%	0 0.0%	60 25.0%	100% 0.0%
		100% 0.0%	98.3% 1.7%	100% 0.0%	100% 0.0%	99.6% 0.4%
		Healthy	IR Fault	OR Fault	Ball Fault	Target Class

Fig. 17. Confusion matrix for rolling bearings at 1797 rpm under 0 hp load and $SNR = 0\text{ dB}$.

Table 1: Fault classification results of experimental test rig dataset

Load Conditions	Classification Accuracy (%)			
	$SNR = 0\text{dB}$	$SNR = -10\text{dB}$	$SNR = -15\text{dB}$	$SNR = -20\text{dB}$
No load	100	100	100	95.6
Medium load	100	94.4	86.9	91.9
Heavy load	100	92.8	80.0	73.8

Table 2: Fault classification results of CWRU bearing dataset

Load Conditions	Classification Accuracy (%)			
	$SNR = 0\text{dB}$	$SNR = -10\text{dB}$	$SNR = -15\text{dB}$	$SNR = -20\text{dB}$
Load-0hp	99.6	84.2	80.4	76.3
Load-1hp	97.1	80.0	77.5	73.8
Load-2hp	95.0	81.3	76.3	71.3
Load-3hp	85.0	74.6	71.2	64.6

Table 3. Proposed method's performance comparison with available results

<i>Scheme</i>	<i>Classification Accuracy (%)</i>		
	<i>Methodology used</i>	<i>SNR = −10dB</i>	<i>SNR = −15dB</i>
Goumas et al. [14]	DWT ,LVQ ANN , Fuzzy gradient and Bayesian classifier	33.33	-
Lou et al.[15]	Wavelet transform and Fuzzy inference	33.33	-
Samanta et al. [27]	Statistical parameters, ANN	33.33	-
Malhi et al. [28]	Statistical parameters, ANN, PCA	38.09	-
Seker et al. [16]	DWT, Multi resolution analysis	78.25	-
Li et al. [17]	DWT, WPT higher order statistics	86	-
Yaqub et al. [18]	WT and cumulant orders, KNN	91.23	-
Zhao et al. [19]	WPT, Permutation entropy, CNN	94.20	-
Amar et al. [20]	Vibration spectral imaging, ANN	96.90	85.15
Proposed OWS-DNN	Optimal wavelet subband, ANN	100	100

Fig. 18. Comparison of OWS-DNN with other methods at (a) $SNR = -10 \text{ dB}$ (b) $SNR = -15 \text{ dB}$.

The results are also compared with some well-established techniques of flaw classification under noisy atmosphere and are tabulated in Table 3 and presented pictorially in Fig. 18. The algorithm works well even at high SNR such as -20 dB where almost all other schemes either cease working or give very poor performance. The results

advocate the robustness of the proposed algorithm over others in severe noisy environment.

Conclusion

This paper proposes a technique named OWS-DNN for machinery fault diagnosis using optimal wavelet subband and DNN. The algorithm is designed to diagnose and

efficiently classify the machinery faults under robust noisy atmosphere at which most of the schemes cease to work or give very poor performance. Experiments have been performed on healthy and faulty bearings in a machinery fault simulator and the faults were seeded using electro discharge machining. To test the universality of the proposed algorithm, it was also tested on the CWRU data set. Various combinations of machinery faults at different loads and SNR levels have been tested under the proposed scheme. The selected optimal wavelet subband undergoes dimension reduction by employing the encoder section of autoencoder and the output is classified by the softmax classifier with high accuracy. For experimental test rig data set, the classification accuracy varies from 73.8 % to 100% at different loads

Appendix

CWRU data used in the proposed work

Table A: 12K Drive End Bearing Fault Data

S.No	Approx. Motor Speed (rpm)	Load (HP)	File number			
			Healthy	IR Fault	OR Fault	Ball Fault
1	1797	0	97.mat	105.mat	118.mat	130.mat
2	1772	1	98.mat	106.mat	119.mat	131.mat
3	1750	2	99.mat	107.mat	120.mat	132.mat
4	1730	3	100.mat	108.mat	121.mat	133.mat

Declarations:

1. Conflict of interests

The authors certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or

and SNR levels. For CWRU data sets the classification accuracy varies from 64.6% to 99.6%. Upon comparing with well-established works of few researchers it is found that the proposed method provides high classification accuracies even at noise levels as high as -20 dB.

Acknowledgement

The authors express sincere gratitude to Dr. Anand Parey, Professor, Department of Mechanical Engineering, Indian Institute of Technology, Indore to allow us for carrying out experiments on machinery fault simulator in his solid mechanics laboratory. It is his profuse support and suggestions due to which we could capture the real time vibration signals for various fault simulations.

beliefs) in the subject matter or materials discussed in this manuscript

2 Funding

We further declare that no funding has been received for the conduct of this research work or preparation of the manuscript.

3. Competing interest

We the authors declare that we have no known competing financial interests or personal relationships that could

have appeared to influence the work reported in this paper.

4. Authors contribution

Experiments were carried out on machinery fault simulators by *Rakesh Kumar Jha* in solid mechanics laboratory at IIT, Indore. Technical discussions and algorithm designing is done by both the authors. Project has been implemented using MATLAB programing by first author. Manuscript prepration is done by the second author, *Preeti D Swami*.

5. Consent to publish

We further declare that the manuscript is the original work of us and we have not published it elsewhere and it has not been submitted elsewhere. We transfer the publication right to Springer, if it would be accepted by the journal for publication.

6. Availability of data and materials

The data and code may be shared if asked after publication.

7. Ethical approval – Not applicable.

8. Consent to participate – Not applicable.

References

[1] Henriquez P, Alonso JB, Ferrer MA, Travieso CM (2014) Review of automatic fault diagnosis systems using audio and vibration signals. In: IEEE Trans. on sys., man and cybernetics: systems 44 (5): 642-652. <https://doi.org/10.1109/TSMCC.2013.2257752>.

[2] Rai A., Upadhyay SH, A review on signal processing techniques utilized in fault diagnosis of rolling element bearings, Tribology International 96 (2016) 289-306. <https://doi.org/10.1016/j.triboint.2015.12.037>.

[3] Duan, Z, Wu T, Guo S. et al.(2018) Development and trend of condition monitoring and fault diagnosis of multi-

sensors information fusion for rolling bearings: a review. Int J Adv Manuf Technol 96: 803–819. <https://doi.org/10.1007/s00170-017-1474-8>

[4] Randall RB, Antoni J (2011) Rolling element bearing diagnostics—A tutorial. Mechanical Systems and Signal Processing. 25: 485–520. <https://doi.org/10.1016/j.ymssp.2010.07.017>.

[5] Pandya Y, Parey A (2014) Spur gear tooth root crack detection using time synchronous averaging under fluctuating speed. Measurement 52:1–11. <https://doi.org/10.1016/j.measurement.2014.02.029>

[6] Yang Y, Yu D, Cheng J (2007) A fault diagnosis approach for roller bearing based on IMF envelope spectrum and SVM. Measurement 40:943–950. <https://doi.org/10.1016/j.measurement.2006.10.010>

[7] Ibarra-Zarate D, Tamayo-Pazos O, Vallejo-Guevara A(2019) Bearing fault diagnosis in rotating machinery based on cepstrum pre-whitening of vibration and acoustic emission. Int J Adv Manuf Technol 104: 4155–4168. <https://doi.org/10.1007/s00170-019-04171-6>

[8] Antoni J(2006) The spectral kurtosis: a useful tool for characterizing non-stationary signals. Mechanical Systems and Signal Processing. 20: 282–307. <https://doi.org/10.1016/j.ymssp.2004.09.001>

[9] Abdelkader R, Kaddour A, Derouiche Z (2018) Enhancement of rolling bearing fault diagnosis based on improvement of empirical mode decomposition denoising method. Int J Adv Manuf Technol 97: 3099–3117. <https://doi.org/10.1007/s00170-018-2167-7>

[10] Hoseinzadeh MS, Khadem SE, Sadooghi MS (2019) Modifying the Hilbert-Huang transform using the nonlinear entropy-based features for early fault detection of ball bearings. Applied Acoustics 150: 313–324. <https://doi.org/10.1016/j.apacoust.2019.02.011>

[11] Amarouayache IIE, Saadi MN, Guersi N. et al. (2020) Bearing fault diagnostics using EEMD processing and

convolutional neural network methods. *Int J Adv Manuf Technol* 107: 4077–4095.
<https://doi.org/10.1007/s00170-020-05315-9>

[12] Amirat Y, Benbouzid MEH, Wang T, Bacha K, Feld G (2018) EEMD-based notch filter for induction machine bearing faults detection. *Applied Acoustics* 133:202–209.
<https://doi.org/10.1016/j.apacoust.2017.12.030>

[13] Moumene I, Ouelaa N (2016) Application of the wavelets multiresolution analysis and the high-frequency resonance technique for gears and bearings faults diagnosis. *Int J Adv Manuf Technol* 83, 1315–1339.
<https://doi.org/10.1007/s00170-015-7436-0>

[14] Goumas SK, Zervakis ME, Stavrakakis GS (2002) Classification of washing machines vibration signals using discrete wavelet analysis for feature extraction. In: *IEEE Trans. Instrum. Measurement* 51 (3) :497–508. DOI :10.1109/TIM.2002.1017721.

[15] Luo X, Loparo KA(2004) Bearing fault diagnosis based on wavelet transform and fuzzy inference. *Mech. System and Signal Processing* 18: 1077–1095.
[https://doi.org/10.1016/S0888-3270\(03\)00077-3](https://doi.org/10.1016/S0888-3270(03)00077-3).

[16] Seker S, Ayaz E (2003) Feature extraction related to bearing damage in electric motors by wavelet analysis. *J. Franklin Inst.* 340: 125–134.
[https://doi.org/10.1016/S0016-0032\(03\)00015-2](https://doi.org/10.1016/S0016-0032(03)00015-2).

[17] Li F, Meng G, Ye L, Chen P (2008) Wavelet transform-based higher-order statistics for fault diagnosis in rolling element bearings. *J. Vib. Control*, 14 (11): 1691–1709. Doi /10.1177/ 1077546308091214

[18] Yaqub MF, Gondal I, Kamruzzaman J (2012) Inchoate Fault Detection Framework: Adaptive Selection of Wavelet Nodes and Cumulant Orders. In: *IEEE Tran. on Instr. and Meas.* 61(3): 685–695. DOI: 10.1109/TIM.2011.2172112.

[19] Zhao L-Y, Wang L, Yan R-Q (2015) Rolling Bearing Fault Diagnosis Based on Wavelet Packet Decomposition

and Multi-Scale Permutation Entropy, *Entropy*. 176447-6461, doi:10.3390/e17096447.

[20] Amar M, Gondal I, Wilson C (2015) Vibration Spectrum Imaging: A Novel Bearing Fault Classification Approach. In: *IEEE Trans. on Industrial Electronics*, 62 (1) : 494–502. <https://doi.org/10.1109/TIE.2014.2327555>.

[38] Jha RK, Swami PD (2020) Intelligent fault diagnosis of rolling bearing and gear system under fluctuating load conditions using image processing technique. *J Mech Sci Technol* 34(10): 4107–4115.
<https://doi.org/10.1007/s12206-020-0903-z>

[21] Yan R, Gao RX (2007) Approximate Entropy as a diagnostic tool for machine health monitoring. *Mech Syst Signal Processing* 21:824–39.
<https://doi.org/10.1016/j.ymssp.2006.02.009>

[22] He Y, Huang J, Zhang B (2012) Approximate entropy as a nonlinear feature parameter for fault diagnosis in rotating machinery. *Meas Sci Technology* 23 : 1-15 .
<https://doi.org/10.1088/0957-0233/23/4/045603>.

[23] Tian Y, Wang Z, Lu C(2019) Self-adaptive bearing fault diagnosis based on permutation entropy and manifold-based dynamic time warping. *Mech Syst Signal Process* 114:658–673.
<https://doi.org/10.1016/j.ymssp.2016.04.028>.

[24] Yan R, Liu Y, Gao RX (2012) Permutation entropy: a nonlinear statistical measure for status characterization of rotary machines. *Mech Syst Signal Processing* 29: 474–84. <https://doi.org/10.1016/j.ymssp.2011.11.022>

[25] Cheng G, Chen X, Li H P Li et al. (2016) Study on planetary gear fault diagnosis based on entropy feature fusion of ensemble empirical mode decomposition. *Measurement* 91:140–54.
<https://doi.org/10.1016/j.measurement.2016.05.059>.

[26] Pan YN, Chen J, Li XL (2009) Spectral entropy: A complementary index for rolling element bearing performance degradation assessment. *Proc Inst Mech*

- Eng Part C J Mech Eng Sci. 223: 1223–31.
<https://doi.org/10.1243/09544062jmes1224>.
- [27] Samanta B, Al-Balushi KR (2003) Artificial neural network based fault diagnostics of rolling element bearings using time-domain features. *Mech. Systems and Signal Proc.* 17(2): 317–328.
<https://doi.org/10.1006/mssp.2001.1462>
- [28] Malhi A, Gao RX (2004) PCA-Based Feature Selection Scheme for Machine Defect Classification. *IEEE Trans. on Instrumentation and Measurement* 53 (6) : 1517-1524. DOI: 10.1109/TIM.2004.834070
- [29] Liu Z, Cao H, Chen X, He Z, Shen(2013) Z Multi-fault classification based on wavelet SVM with PSO algorithm to analyze vibration signals from rolling element bearings. *Neurocomputing* 99: 399–410.
<https://doi.org/10.1016/j.neucom.2012.07.019>.
- [30] Jha RK, Swami PD (2020) Intelligent fault diagnosis of rolling bearing and gear system under fluctuating load conditions using image processing technique. *J Mech Sci Technol* 34(10): 4107–4115.
<https://doi.org/10.1007/s12206-020-0903-z>
- [31] Sharma R, Pachori RB, Sircara P (2020) Seizures classification based on higher order statistics and deep neural network. *Biomedical Signal Processing and Control* 59: 101921.
<https://doi.org/10.1016/j.bspc.2020.101921>
- [32] Sun W, Shao S, Zhao R et al. (2016) A sparse autoencoder-based deep neural network approach for induction motor faults classification, *Measurement* 89 : 171-178
<https://doi.org/10.1016/j.measurement.2016.04.007>
- [33] Schmidhuber J (2015) Deep learning in neural networks: An overview. *Neural networks* 61:85-117.
<https://doi.org/10.1016/j.neunet.2014.09.003>.
- [34] Nason GP, Silverman BW, The Stationary Wavelet Transform and some Statistical Applications, . In: Antoniadis A., Oppenheim G.(eds) *Wavelets and Statistics*. Lecture Notes in Statistics, vol 103. Springer, New York. https://doi.org/10.1007/978-1-4612-2544-7_17.
- [35] Rioul O, CMAP (2018) This is IT: A Primer on Shannon's Entropy and Information. *L'Information, Seminaire Poincare*. 23 : 43 -77.
- [36] Djebbar F, Ayad B (2017) Energy and Entropy Based Features for WAV Audio Steganalysis. *Journal of Information Hiding and Multimedia Signal Processing*. 8(1) : 168-181.
- [37] Starck JL, Murtagh F, Gstaad R(1998) A New Entropy Measure Based on the Wavelet Transform and Noise Modeling, *IEEE Trans. On Circuits and Systems—II: Analog and Digital Signal Processing* 45(8) :1118-1124. DOI: 10.1109/82.718822.
- [38] Santos LS, Secchi AR, Biscaia Jr EC (2012) Wavelet Threshold Influence in Optimal Control Problems. *Computer Aided Chemical Engineering* 30 : 1222-1226.
<https://doi.org/10.1016/B978-0-444-59520-1.50103-2>
- [39] CWRU bearing data centre, seeded fault test data, <https://csegroups.case.edu/bearingdatacenter/pages/download-data-file>
- [40] Jha RK, Swami PD, Singh D (2018) Comparison and Selection of Wavelets for Vibration Signals Denoising and Fault Detection of Rotating Machines using Neighbourhood Correlation of SWT Coefficients. In: *International Conference on Advanced Computation and Telecommunication (ICACAT-2018)* Bhopal, India. DOI: 10.1109/ICACAT.2018.8933643.

Figures

Figure 1

Framework of OWS-DNN algorithm

Figure 2

(a) Upsampling of low-pass and high-pass filters (b) Signal decomposition at the first level (c) Signal decomposition at the jth level.

Figure 3

Entropy vs. SNR for bearings

Figure 4

Entropies of approximation and detail subbands.

Figure 5

Block diagram of Autoencoder

Figure 6

Single hidden layer autoencoder

Figure 7

Experimental setup of machinery fault simulator.

Figure 8

Signal plot for healthy and faulty bearings at no load and SNR=0 dB

Figure 9

Signal plot for healthy and faulty bearings at no load and SNR=-20 dB

Figure 10

Experimental test bench setup CWRU [39]

Figure 11

Approximation and detail subbands of a healthy bearing.

Figure 12

Approximation and detail subbands of a bearing with fault in the inner race.

Figure 13

Approximation and detail subbands of a bearing with fault in the outer race

Figure 14

Approximation and detail subbands of a bearing with ball fault.

Figure 15

Fault classification using softmax classifier

Confusion Matrix					
Output Class	Healthy	IR Fault	OR Fault	Ball Fault	
	40 25.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	0 0.0%	40 25.0%	0 0.0%	0 0.0%	100% 0.0%
	0 0.0%	0 0.0%	40 25.0%	0 0.0%	100% 0.0%
	0 0.0%	0 0.0%	0 0.0%	40 25.0%	100% 0.0%
					Target Class
					Healthy
					IR Fault
					OR Fault
					Ball Fault

Figure 16

Confusion matrix for rolling bearings at 1200 rpm under no load and SNR=0 dB.

		Confusion Matrix				
Output Class	Healthy	60 25.0%	1 0.4%	0 0.0%	0 0.0%	98.4% 1.6%
	IR Fault	0 0.0%	59 24.6%	0 0.0%	0 0.0%	100% 0.0%
	OR Fault	0 0.0%	0 0.0%	60 25.0%	0 0.0%	100% 0.0%
	Ball Fault	0 0.0%	0 0.0%	0 0.0%	60 25.0%	100% 0.0%
		100% 0.0%	98.3% 1.7%	100% 0.0%	100% 0.0%	99.6% 0.4%
		Target Class				
		Healthy	IR Fault	OR Fault	Ball Fault	

Figure 17

Confusion matrix for rolling bearings at 1797 rpm under 0 hp load and SNR=0 dB

Figure 18

Comparison of OWS-DNN with other methods at (a) SNR =-10 dB (b) SNR=-15 dB.