Springer

KSME

Journal of Mechanical Science and Technology 33 (7) (2019) 3095~3103
www.springerlink.com/content/1738-494x(Print)/1976-3824(Online)
DOI 10.1007/s12206-019-0604-7

Vibration signal based condition monitoring of mechanical equipment with scattering transform †

P. S. Ambika[1,*], P. K. Rajendrakumar[1] and Rijil Ramchand[2]

[1]*Department of Mechanical Engineering, National Institute of Technology, Calicut, 673601, India*
[2]*Department of Electrical Engineering, National Institute of Technology, Calicut, 673601, India*

(Manuscript Received November 1, 2018; Revised January 29, 2019; Accepted April 8, 2019)

Abstract

Scattering transform is proposed using machine learning to extract translational, rotational and deformation invariant information for the first time from vibration signals obtained from rolling element bearings (REBs). The core idea of scattering transform lies in the construction of a scattering network which is formed from a stack of signal processing layers of increasing width. Each layer is formed from the association of a linear filter bank with a non-linear operator. It uses a cascade of wavelet filter bank, modulus rectifiers and averaging operators to build a deep convolution network and computes multi-scale co-occurrence coefficients which are invariant to translation in time, rotation and deformation. The scattering transform coefficients are extracted as features from seven stages of a vibration signal prognosis data repository which are then input to a support vector machine (SVM) classifier. Vibration signals from the intelligent management system (IMS) bearing data centre are used to validate the proposed algorithm. Test results analysis and solution show that scattering transform can be used to obtain distinguishing features from seven bearing health stages with an average accuracy of 99 %. The results were compared with other feature extraction strategies on the same data and were found to be superior.

Keywords: Bearing fault diagnosis; Scattering convolution network; Scattering transform; Support vector machine

1. Introduction

The health condition of mechanical equipment is identified through continuous or periodic measurement and interpretation of time-domain data that help in determining the need for maintenance. Vibration, sound and temperature measurements are often key indicators of the state of the machine [1, 2]. Any increase in the amplitude of vibration, sound or temperature indicates the presence of a fault. Condition monitoring is the field of study primarily involved in diagnostics of faults and failure detection. The sooner the diagnosis, the chances for machinery breakdown and equipment downtime could be greatly reduced. Vibration signal analysis is the most common method for condition monitoring of mechanical equipment, especially for rotating machines. Conventional vibration analysis instruments use FFT (fast Fourier transform) which is a special case of discrete Fourier transform and also wavelet-based methods [3, 4]. But even though the wavelet transforms mitigates the disadvantages of FFT its lack of translation invariance affects the accuracy of the classification problem. Feature selection plays a vital role in the success of pattern classification. The non-stationary nature of real-time signals in faulty operating conditions makes it difficult to extract relevant information necessary for classification. Under such conditions, conventional measurement of statistical parameters may not be useful [5, 6]. In this context, researchers are forced to pay their attention to signal processing methods that improve fault classification tools [7, 8]. The performance of the classifier often depends not only on the algorithm but also on the selected features. Since the monotonicity property of a neural network assumes that increasing the number of features does not decrease the classifier's performance, one tends to use more features than needed in a classification problem [9, 10]. However, it has been observed in practice that in some cases [11] more features may reduce the classifier performance, especially when the number of training samples is small compared to the number of selected features. It is evident that a small number of features reduce the measurement cost and memory requirements. Moreover, the Watanabe's ugly duckling theorem [11] states that two arbitrary patterns can be made similar by translating them into a sufficiently large feature matrix through a cautious choice of feature vectors. It is hence essential to cleverly select among all features with redundant information, the best subset for practical application. This selection procedure should improve the performance of a

*Corresponding author. Tel.: +91 9446413119
E-mail address: ambikaps89@gmail.com
† Recommended by Associate Editor Gyuhae Park

classifier by promoting the strong feature and diminishing those with little information while at the same time reducing the correlation between features. There are two requirements for a good feature selection algorithm:

(i) There must be a well-defined criterion for judgment on the suitability of the subset of features, and

(ii) an efficient procedure for searching through the feature space.

Finding a suitable kernel to classify the signal optimally is another major challenge in any classification problem. Mapping a signal to a higher dimensional space helps to differentiate the signals by bringing a hyperplane to separate the different classes. Non-linear mapping done to a certain problem will not work with all cases when choosing a kernel. Structured support vector machines (SVMs) builds these types of kernels. In structured SVM, the signal x undergoes a succession of operations namely convolutions, rectification, normalization, regrouping several times and the filters in convolution operation undergoes learning to obtain the representation which is most appropriate for classification.

This paper aims to provide a novel feature extraction method proposed by S. Mallat for accurate fault identification of rolling element bearings through scattering transformation. Scattering wavelet transform convolution network is first used to derive a dimensionally reduced feature matrix in the field of REB fault diagnosis and condition monitoring. The features thus obtained are translation and rotation invariant and stable to deformations. The output thus obtained are multiscale spectral co-occurrence coefficients that retain all non-stationary spectral information, transients and time-dependent variations which are otherwise lost by fixed basis-functions. SVM classifier is used to train the features from various fault conditions occurring in the REBs which include faulty inner race, faulty outer race, faulty roller etc. Scattering transform coefficients for a new set of signals are later given to the SVM classifier to test the accuracy of classification. The ratio of number of signals correctly classified to the total number of test signals gathered is considered as the accuracy percentage. The overall computational complexity is of the order of ($N \log N$) where N is the size of the input vector. The software used for the proposed work is MATLAB 2018a and an 8 GB RAM workstation computer is employed for all the computations.

This work is organised as follows: Sec. 2 details the experimental setup from which the data were taken. Here detailed explanation of the scattering transform algorithm used is presented. Sec. 3 describes the classification methodology used in the proposed work. Sec. 4 mentions the overall procedure incorporated in the present work to reach the conclusions presented. It explains the dataset used and the results obtained. Comparisons with the existing literature are carried out in the Sec. 5 of this manuscript. The discussions based on the results obtained are summarised in Sec. 6. The work is concluded in Sec. 7.

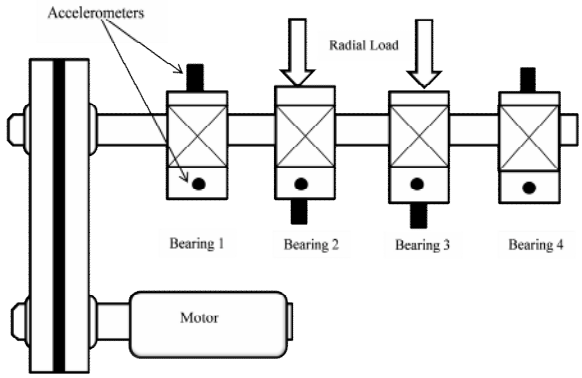

Fig. 1. The schematic of experimental setup for bearing run-to-failure analysis.

2. Materials and methods

2.1 Materials

The IMS bearing database [12] consists of three sets of four and eight-channel 1-second vibration signal snapshots recorded from four Rexnord ZA – 2115 double row bearings over durations of 30-35 days. Each data set is a test-to-failure experiment on the bearings with individual files of 20480 data points recorded at a sampling rate of 20 kHz. Each file is recorded at specific intervals, and the file name indicates the time when the data was collected.

2.2 Methods

A stable signal representation retains information about the signal under various invariances like translation, deformation and scaling. The scattering transform developed by the team under Mallat and Bruna [13], provides such coefficients from operations performed using a cascade of:

(i) Wavelet-based linear filtering as convolution,

(ii) modulus operator for nonlinear operation (rectification),

(iii) local averaging (averages the amplitude of the wavelet coefficients), and subsampling.

This process is repeated many times. At the end of this operation, a representation of scattering coefficients is obtained which will be most appropriate for classification [14]. The non-linear modulus operation is performed through convolution network architecture to create invariant wavelet coefficients [15]. Convolution networks can bring invariant coefficients in large-scale variabilities. Bruna et al. [16] mathematically studied the significance of multiple layers of nonlinear operation in the neural net, optimal properties, pooling operations, etc. Also they address the issue by scattering transform based deep convolution network.

A schematic representation of the scattering transform network is shown in Fig. 2. This network retains the relevant information and discards the unwanted information. The input signal can be recovered by inverting the wavelet modulus averaging operators. The translation invariant representation can also be achieved through Fourier transform (FT) modulus,

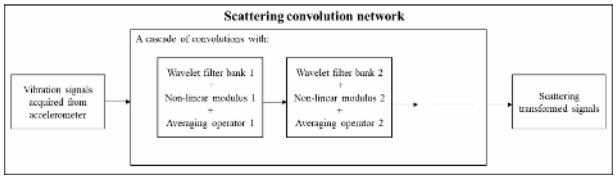

Fig. 2. A schematic representation of a scattering convolution network.

autocorrelation operation and registration. The Fourier transform (FT) of $x(t)$ is obtained as:

$$X(\omega)=\int x(u)e^{-i\omega u}du\,. \tag{1}$$

The short term FT of $x_{t,T}(u)=x(u)\mathcal{W}_T(u-T)$ is:

$$X_{(t,T)}(\omega)=\int x_{(t,T)}(u)e^{(-i\omega u)}du \tag{2}$$

where $\mathcal{W}_T$ is the window function with length T. But these methods fail to capture the stable invariant information under deformations. The sinusoidal waves that make up the Fourier transformed signal fail to capture the local variations due to stability issues under deformations. The FT operation can be made translation invariant by taking its modulus operation but it kills the phase.

Mel-frequency spectral coefficients (MFSC's) [17] are obtained as:

$$M_T x(t,j)=\frac{1}{2\pi}\int\left|X_{t,T}(\omega)\right|^2\left|\psi_j(\omega)\right|^2 d\omega \tag{3}$$

i.e, averaging the spectrogram $|X_{(t,T)}(\omega)|^2$ over Mel-frequency intervals $\psi_j(\omega)$ indexed by j.

By Parseval's theorem, Eq. (3) can be simplified as:

$$M_T x(t,j)=\int\left|x_{t,T}(u)\otimes\psi_j(u)\right|^2 \tag{4}$$

which translates to the energy of x of size T in a neighbourhood of t in j. However, the MFSC's cannot capture non-stationary structures of duration shorter than T (which is as small as 23 ms since any non-stationary signal is considered to be stationary around this interval [17]) like transients and time-dependent variations. This information can be retained by scattering transformation as multiscale spectral co-occurrence coefficients. The MFSC coefficients are the squared energy wavelet coefficients at the scale a^j in a neighbourhood of t. If the coefficients are obtained with logarithm, the squared amplitudes are no longer significant and Eq. (4) can be calculated by averaging the wavelet coefficient amplitudes of x as $\left|x\otimes\psi_j\right|\otimes\phi_J(t)$ over the neighbourhood of t in the interval covered by ψ_j of duration $T=a^J$. As the value of T increases, more information is lost by averaging. To recover this lost information, we try to bring about a link between MFSC coefficients and wavelet transforms. The term $\left|x\otimes\psi_j\right|\otimes\phi_J(t)$ can be thought to be similar to the low frequency component of the wavelet transform of $\left|x\otimes\psi_{j1}\right|$:

$$W_J\left|x\otimes\psi_{j1}\right|(t)=\begin{pmatrix}\left|x\otimes\psi_{j1}\right|\otimes\phi_J(t)\\ \left|x\otimes\psi_{j1}\right|\otimes\psi_{j2}(t)\end{pmatrix}_{j_2<J+P}. \tag{5}$$

Wavelet transform is invertible, and therefore, the information lost by convolution with $\phi_J(t)$ is recovered by obtaining averaged measurements. But in the case of the wavelet, as the wavelet coefficients of the signal and its translated form are different, it gives translation covariance. In scattering transform, translation invariance is introduced by non-linearity. Wavelet transforms utilises translations hence not translation invariant [16]. If K is a linear or non-linear operator which utilises translations, then $\int K\,x(u)du$ is translation invariant. Similarly, if $Kx=x\otimes\psi_\lambda$ then $\int x\otimes\psi_\lambda(u)du$ is translation invariant for all x, $\because\int\psi_\lambda(u)du=0$. Wavelet operations are localized operations on the signal which tells about how sensitive it can be, during translation [17]. This issue is handled in the scattering method through modulus operation and averaging the signal in time domain. Averaging operation averages all the wavelet coefficients in different positions. This process certainly loses resolutions (it removes all the high-frequency components), but it gains stability and solves the invariance problem. The scattering method uses the multilayer architecture in the deep network to recover this lost information [18]. In this network, different wavelet-based filters are used on which the high-frequency components are scattered, thereby it is possible to recover the lost information. Since the wavelet coefficient of the signal and its shifted version will be different, wavelets don't exhibit the translation invariance property. The scattering transform achieves translation invariance by performing a non-linear operation through modulus and averaging the result. The modulus of the wavelet coefficients are filtered with a low-pass filter to obtain the averaged coefficients:

$$\left|\left|x\otimes\psi_{j1}\right|\otimes\psi_{j2}(t)\right|\otimes\phi_J(t) \tag{6}$$

which are the co-occurrence coefficients at scales a^{j1} and a^{j2}. They are referred to as co-occurrence coefficients because they compute the interferences of signal x with two successive wavelets ψ_{j1} and ψ_{j2}. These coefficients are also called scattering coefficients. They assess the time-dependent variations that $\left|x\otimes\psi_{j1}(t)\right|$ incurs in the wavelet domain interval spanned by ψ_{j2}. This aids scattering trans-

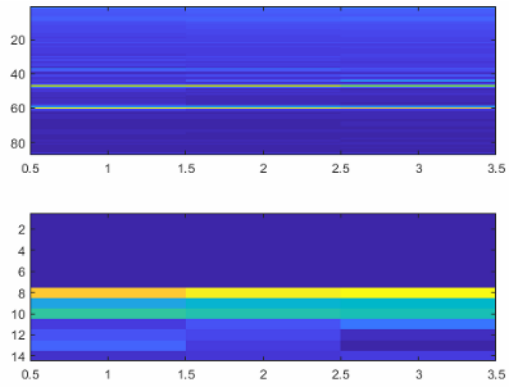

Fig. 3. First-order scattering coefficients of vibration signal sampled at 20 KHz and the co-occurrence coefficients $\left\|x\otimes\psi_{j_1}\right|\otimes\psi_{j_2}\left|\otimes\phi_J\right.$ for a scale a^{j1}.

form to achieve translation invariance. More specifically, by performing a non-linear operation through modulus and averaging the result, translation invariance is achieved in scattering transformations. First-order scattering coefficients of the vibration signal obtained from REBs sampled at 20 KHz and the co-occurrence coefficients $\left\|x\otimes\psi_{j_1}\right|\otimes\psi_{j_2}\left|\otimes\phi_J\right.$ for a scale a^{j1} are shown in Fig. 3.

2.2.1 Scattering wavelets

The modulus of convolution operation in the scattering transform can be denoted by operator U. This operator computes the modulus of complex wavelet coefficients while keeping the phase of $x\otimes\phi_J$ [18, 19]:

$$U_J x(t)=\begin{pmatrix} x\otimes\phi_J(t) \\ \left|x\otimes\psi_J(t)\right| \end{pmatrix}_{j<J+P} \tag{7}$$

where '$\otimes$' operator refers to convolution. As wavelets have zero mean, the translated invariance measure extracted from each wavelet sub-band lambda, produced a non-zero average value through the non-linear operation. 'U' is a wavelet modulus operator at the j^{th} sub-band level and P is the number of wavelet filters. A scattering transform first computes $U_J x$ and outputs the low-frequency signal $x\otimes\phi_J$. At the next layer, U_J retransforms each $\left|x\otimes\psi_{j_1}\right|$ to $\left|x\otimes\psi_{j_1}\right|\otimes\phi_J$ and computes $\left\|x\otimes\psi_{j_1}\right|\otimes\psi_{j_2}\left|\right.$. These coefficients are again transformed by U_J, which outputs $\left\|x\otimes\psi_{j_1}\right|\otimes\psi_{j_2}\left|\otimes\phi_J\right.$ and thus third-order wavelet signals are computed which are again sub-decomposed by U_J and so on.

This transformation is repeated for levels and removing the coefficients not filtered by ϕ_J yields a scattering vector of size $m+1$ at time t:

$$S_J x(t)=\begin{pmatrix} x\otimes\phi_J(t) \\ \left|x\otimes\psi_{j_1}\right|\otimes\phi_J(t) \\ \left\|x\otimes\psi_{j_1}\right|\otimes\psi_{j_2}\left|\otimes\phi_J(t)\right. \\ \vdots \\ \ldots\left|x\otimes\psi_{j_1}\right|\ldots\left|\otimes\psi_{j_m}\right|\otimes\phi_J(t) \end{pmatrix}_{j_1,j_2,\ldots<J+P} . \tag{8}$$

2.2.2 Scattering convolution network

This scattering transform is a cascade of modulated filter banks and non-linear rectifications. It has architecture similar to that of convolution network. The output is taken from all the layers. The energy in different layers decreases as the number of layers increase. Deep inside the layer, the energy is very minimal.

From the figure, it can be understood that the input signal (here 'x') is distributed over the cascade of wavelet modulus operators. The stages of distribution are denoted as U. In the operation $\left|x\otimes\psi_{j_1}\right|$ and $\left\|x\otimes\psi_{j_1}\right|\otimes\psi_{j_2}\left|\right.$ the high frequency information is extracted using the wavelet operation. This will induce deformation, which are handled by modulus and averaging operation, given as $\left\|x\otimes\psi_{j_1}\right|\otimes\psi_{j_2}\left|\otimes\phi_J\right.$. The scattering coefficients of the signal 'x' obtained in each stage are: $S[0]x$, $S[j_1]x$, $S[j_1\,j_2]x$, $\cdots$, $S[j_1\,j_2\ldots j_m]x$ where $[j_1\,j_2\ldots j_m]$ denotes the path of frequencies 'P'. Therefore the scattering coefficients in the various paths are as follows:

The first scattering coefficient is the average of the signal obtained by convolving the signal with an averaging filter ϕ_J as:

$$S[0]x=x\otimes\phi_J(t). \tag{9}$$

The next layer is obtained by applying magnitude of the wavelet transforms at different scales and orientations and convolving with a low-pass filter ϕ_J as:

$$S[j_1]x=\left|x\otimes\psi_{j_1}\right|\otimes\phi_J(t). \tag{10}$$

The next layer is obtained by convolving the modulus of the coefficients obtained from the previous layer with a filter as:

$$S[j_1\,j_2]x=\left\|x\otimes\psi_{j_1}\right|\otimes\psi_{j_2}\left|\otimes\phi_J(t)\right.. \tag{11}$$

The next layer is obtained in a similar manner as:

$$S[j_1\,j_2\ldots j_m]x=\left\|\ldots\left|x\otimes\psi_{j_1}\right|\ldots\left|\otimes\psi_{j_m}\right|\otimes\phi_J(t)\right\| \tag{12}$$

and so on.

The resulting scattering operator is as shown in Eq. (8). A

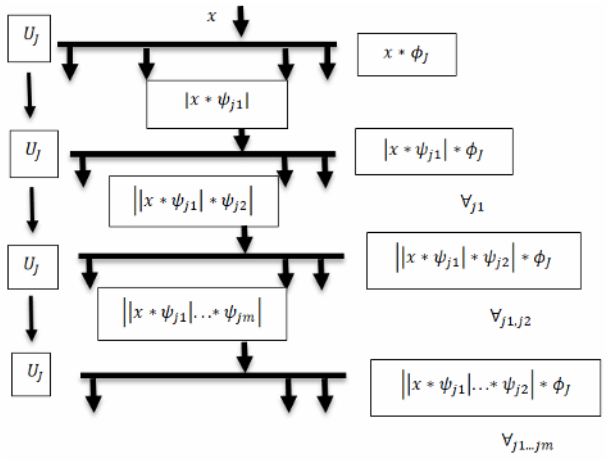

Fig. 4. A scattering operator showing cascade of wavelet modulus operators which outputs convolutions with ϕ_j [14].

spectrogram-like visualization of the second order translation and rotation-invariant scattering transform called scattergram of the rolling element bearing signals of IMS dataset is shown in Fig. 3. The top image is made of first-order coefficients, organized in time-scale for a scale a^{j1}. Very low frequencies, i.e., large scales, appear at the bottom of the image. The second image shows re-normalized second-order co-occurrence coefficients $\left||x \otimes \psi_{j_1}| \otimes \psi_{j_2}\right| \otimes \phi_J$.

3. Classification using support vector machine classifier

Let $X_1 = \{x_1, x_2, \cdots, x_n\}$ be the input and $Y_1 = \{y_1, y_2, \ldots, y_n\}$ be the output label and let X and Y be their sample spaces. In SVM^{struct}, the labels can be any structured spaces like strings or lattices or graphs, not just intergers or real spaces. The objective here is to generalise multi-class SVM formulation such that features are extracted from both inputs and outputs. This is performed with the help of powerful optimization techniques for learning an algorithm that maps input vectors to their corresponding outputs. The feature function Ψ is used to map a pair of (x_n, y_n) to a computationally simple form.

$$F : X \text{ X } Y \rightarrow \mathbb{R}^d \tag{13}$$

In other words, the aim is learning a problem which learns a function:

$$f : X \rightarrow Y \tag{14}$$

such that the training sample of $(x_1, y_n), (x_2, y_n), \ldots, (x_n, y_n) \epsilon X \times Y$. This is performed by learning a discriminant function:

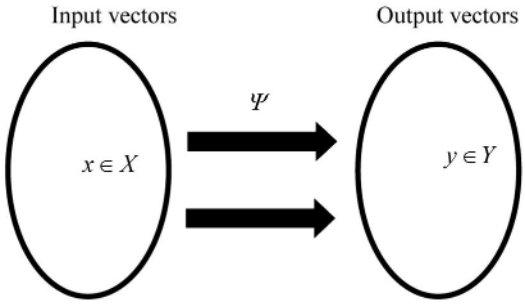

Fig. 5. A graph depicting the simplified version of the learning algorithm employed in SVM^{struct}.

$$F : X \times Y \rightarrow \mathbb{R}. \tag{15}$$

The hypothesis f is:

$$f(x, w) = arg \max_{y \in Y} F(x, y; w) \tag{16}$$

where w is a parameter vector. The prediction f is derived from (x_n, y_n), the input-output pairs by maximising F over the response variable for a specific given input x. F is linear in some combined feature representation of input-output pair with w as the combining vector often called weight matrix in a space $\Psi(x, y)$, often named kernel space. In other words,

$$F(x, y; w) = \langle w, \Psi(x, y) \rangle. \tag{17}$$

Thus the prediction f can be calculated by finding the structure $y \in Y$ which maximises $F(x, y; w)$.

The classifier is trained by solving the convex optimization problem:

$$\min_{\omega} \omega^2 + C \sum_n \max_y \left(\Delta(y_n, y) + \omega . \left(\Psi(x_n, y) - \Psi(x_n, y_n) \right) \right) \tag{18}$$

where $(x_1, y_n), (x_2, y_n), \ldots, (x_n, y_n)$ is the training set of only positives and C is the regularisation constant [20-23].

4. Scattering transform-based fault detection

Scattering transform is a machine learning based approach to rotating machine fault diagnosis. The transformed data has three properties:

- Translational invariance
- Rotational invariance
- Deformation stability

The scattering transform based algorithms have found applications in music genre recognition, image and texture classification, and fetal heart rate characterization.

4.1 Experimental device and datasets

The run-to-failure bearing vibration signals are obtained

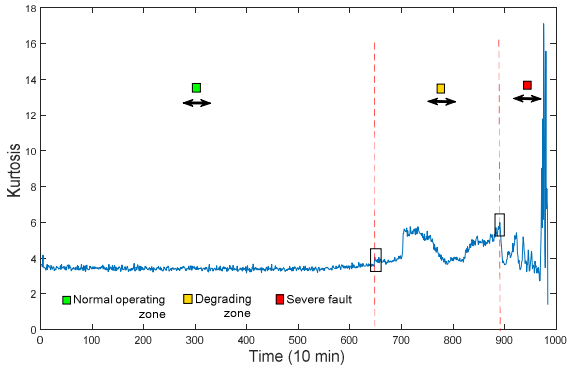

Fig. 6. The kurtosis plot of bearing 3 for the whole life cycle.

from the IMS bearing data centre [12]. The bearing data set have been validated in many research works [24-30] and is a standard for the study on bearing remaining useful life prediction.

The bearing type used in this experiment is Rexnord ZA-2115 double row ball bearings. Four such bearings were run on a maximum radial load of 6000 lbs at 2000 rpm. Three such tests were carried out with an accelerometer placed on the bearing holders recording the vibration signals from each bearing. An oil circulation system was installed to lubricate the bearings. A magnetic plug is placed in the oil feedback pipe to collect the debris. The test is terminated by an electric switch when the debris collected by the magnetic plug exceeds a certain level. Since different bearings fail differently, the fault patterns of the bearings once dismantled are recorded, and their fault patterns are studied. The schematic of the experimental setup is shown in Fig. 1.

The three tests carried out with the test mentioned above setup provided vibration signals from three different fault types: Inner race fault, rolling element defect, and outer race fault. Bearing 3 and 4 of test 1 displayed outer race and rolling element defects, respectively. Bearing 1 of test 2 showed outer race defect. The kurtosis plot of each of these defect cases shows an increasing trend in amplitude when plotted against time in the x-axis. The kurtosis curve of bearing 3 is shown in Fig. 6. This plot is divided into three regions based on their kurtosis values: normal operating zone, degrading zone, severe fault zone. Thus seven conditions were obtained: Normal (N), degrading inner race (DIR), degrading outer race (DOR), degrading roller (DR), faulty inner race (FIR), faulty outer race (FOR) and faulty roller (FR).

4.2 Classification result analysis

Traditional classification methods normally use statistical features such as mean, standard deviation, kurtosis, skewness, root mean square and so on. These features are extracted from the vibration signals obtained from the accelerometer to form a feature matrix. Here, the feature matrix is constructed from a stack of signal processing layers of increasing width called

Table 1. Detailed description of the samples and accuracy obtained.

	Raw sample length	No. of testing samples/ training samples	Number of samples incorrectly classified	Time taken (s)	Accuracy obtained (%)
All health conditions	700	92/608	0	0.8129	100
		174/526	1	0.9887	99.41
		234/466	2	0.9684	99.58
		296/404	0	0.8874	100
		299/401	1	0.8654	99.51
		357/343	2	0.8995	99.54
		478/222	2	0.8147	99.58
		505/195	4	0.9982	99.21
		536/164	2	0.9986	99.62

scattering network. The layers are formed by the association of a linear filter bank w_{op} with a non-linear complex modulus operator. Each operator w_{op} performs two actions: energy averaging by a low-pass filter ϕ which includes averaging the modulus coefficients obtained from the previous layer, and energy scattering by a band-pass filter ψ. The input signal x is decomposed using linear operators ψ_{op} and modulus operators, creating scattering invariants S and intermediate covariant coefficients U_j. Thus the scattering transform of an input signal x is defined as the set of all paths that x might take from layer to layer. In this way, the architecture of a scattering network resembles a convolutional deep network.

In our experiment, 100 samples of each of the seven fault type are provided as input to the scattering network. The feature matrix is composed of 380 scattering transform coefficients obtained from each of the 100 samples. The 20 kHz sampled signal was thus reduced to 380 samples in the scattering domain. The scattering transform coefficients of all health conditions are considered as raw sample data. From this sample, a random number of samples (using rand function, MATLAB) are grouped as test data and provided as input to the SVM classifier. The results obtained using different classifiers may vary. The proposed work used SVM as the classifier because it has been proved to be better than ANN and other classifiers in the literature. Moreover, overlooking the fact that there are different kernel functions inside SVM classifier that maps the input variables into kernel space, the most efficient classifier has been selected on a trial and error basis. The classifier has been trained and tested 100 times and the average classification accuracies have been calculated. The output from the SVM classifier is the ratio of number of correctly classified test data to the total number of test data from each health condition. The accuracy percentages were obtained from 99.21 % - 100 % for all combinations. The corresponding confusion matrices are also shown in Tables 2-7.

Table 2. Confusion matrix 1.

	N	DIR	DOR	DR	FIR	FOR	FR
N	4	0	0	0	0	0	0
DIR	0	7	0	0	0	0	0
DOR	0	0	3	0	0	0	0
DR	0	0	0	8	0	0	0
FIR	0	0	0	0	20	0	0
FOR	0	0	0	0	0	29	0
FR	0	0	0	0	0	0	21

Table 3. Confusion matrix 2.

	N	DIR	DOR	DR	FIR	FOR	FR
N	8	0	0	0	0	0	0
DIR	0	13	0	0	0	0	0
DOR	0	0	6	0	0	0	0
DR	0	0	0	15	0	0	0
FIR	0	0	0	0	38	0	0
FOR	0	0	0	0	0	54	0
FR	0	0	0	0	0	0	40

Table 4. Confusion matrix 3.

	N	DIR	DOR	DR	FIR	FOR	FR
N	11	0	0	0	0	0	0
DIR	0	17	0	0	0	0	0
DOR	0	0	8	0	0	0	0
DR	0	0	0	20	0	0	0
FIR	0	0	0	0	51	0	0
FOR	0	0	0	0	0	73	0
FR	0	0	0	0	0	0	54

Table 5. Confusion matrix 4.

	N	DIR	DOR	DR	FIR	FOR	FR
N	14	0	0	0	0	0	0
DIR	0	21	0	0	1	0	0
DOR	0	1	9	0	0	0	0
DR	0	0	0	24	1	0	0
FIR	0	0	0	0	65	0	0
FOR	0	0	0	0	0	92	0
	N	DIR	DOR	DR	FIR	FOR	FR

5. Comparisons with existing literature

The case study performed using the scattering transform coefficients and SVM is evaluated by comparing with seven intelligent diagnosis methods in Table 8.

Except for the first method [24] the other six use adaptive signal processing strategies like empirical mode decomposition. The first method [24] shows better classification accuracy than scattering transform. But they have considered only three conditions (N, FIR, FOR) of bearing degradations. Scat-

Table 6. Confusion matrix 5.

	N	DIR	DOR	DR	FIR	FOR	FR
N	25	0	0	0	0	0	0
DIR	0	13	0	0	0	0	0
DOR	0	0	44	0	0	0	0
DR	0	0	0	126	0	0	0
FIR	0	0	0	0	119	0	0
FOR	0	0	0	0	0	15	0
FR	0	0	0	0	0	1	29

Table 7. Confusion matrix 6.

	N	DIR	DOR	DR	FIR	FOR	FR
N	114	0	0	0	0	0	0
DIR	0	39	0	0	0	0	0
DOR	0	0	32	0	0	0	0
DR	0	0	0	26	0	0	0
FIR	0	0	0	0	75	0	0
FOR	0	0	0	0	0	76	0
FR	0	0	0	0	0	0	49

Table 8. Comparisons of existing feature extraction from literature for IMS bearing dataset.

Reference	Feature extraction	Classification method	Classes	Accuracy obtained (%)
Gryillias & Antoniadis, 2012 [24]	Time-domain statistics + Frequency domain features	SVM	N, FIR, FOR	100
Yu & Junsheng, 2006 [25]	Empirical mode decomposition	NN	N, FIR, FOR	93
Liu, Cao, Chen, He, & Shen, 2013 [26]	Time domain statistics	SVM	N, FIR, FOR, FR	97.5
Xu & Chen, 2013 [27]	Empirical mode decomposition	SVM	N, FIR, FOR	97
Ben Ali et al., 2015 [28]	Empirical mode decomposition + Time domain statistics	NN	N, DIR, DOR, DR, FIR, FOR, FR	93
Zhang et al., 2018 [29]	Automatic learning	Softmax	N, DIR, FIR, DR, FR, DOR, FOR	99.66
Berredjem & Benidir, 2018 [30]	Wavelet packet coefficients related statistical features	Fuzzy expert system	FIR, FR, FOR	96.08
This work	Scattering transform	SVM	N, DIR, DOR, DR, FIR, FOR, FR	99.21-100

tering transform produces testing accuracies of 100 % (92/700; the ratio of test samples to total samples), 99.41 % (174/700), 99.58 % (234/700), 100 % (296/700), 99.51 % (299/700), 99.54 % (357/700), 99.58 % (478/700), 99.21 % (505/700) and 99.62 % (536/700). The overall classification accuracies obtained were ranging from 99.21 – 100 %. This suggests that the proposed work almost accurately classifies all the health categories in bearing fault diagnosis.

6. Discussion

The scattering transform operation uses a scattering propagator operator $U_J x(t)$ at each layer. On comparing scattering convolution network with other convolution networks, this method is a deep convolution network. It outputs co-occurrence coefficients $S_J x(t)$ at all layers not alone the last layer. As the levels propagate, the energy of the deepest layer reduces to zero and majority energy is concentrated at the layers near $|J| \le 3$. Moreover, the filters are designed from predefined wavelets and not from the data. This leads to information loss that may have helped differentiate classes [24]. The high frequency information lost during averaging is recovered at the very next layer by applying wavelet transform on the modulus of wavelet coefficients from the previous layer in a wavelet scattering convolution network [11]. Moreover, the scattering transform provides more discrimination capability as it provides information on higher order moments [26]. It eliminates high frequency instability as it preserves Lipschitz stability of wavelets. In addition, very few parameters need be manipulated depending on the nature of the input dataset. This saves a lot of computational complexity.

7. Conclusion

The present study proposes a machine learning algorithm called scattering transform for the first time in bearing fault diagnosis. The main contributions are the development of scattering coefficient features that are invariant to translational and rotational variance. The architecture of a scattering network closely resembles a deep convolutional network. The proposed method is evaluated on IMS bearing dataset for three rolling element bearings. The current work is a small attempt in realising an invariant representation of diagnosable discriminative features that can achieve better accuracy than intelligent diagnosis methods. The proposed methodology can be extended to predict the remaining useful life of all kinds of rolling element bearings. This is considered in an upcoming work.

Acknowledgements

Many people have directly or indirectly contributed to the progress of this research. Among all, I would like to particularly express my heartfelt gratitude to Dr K. P. Soman, Professor and Head, Centre for Excellence in Computational Engineering and Networking (CEN), Amrita Vishwa Vidyapeetham, Ettimadai, Coimbatore, Tamil Nadu, India, my mentor and guide who introduced me to Scattering transforms. In addition I am also grateful to Ms. Neethu Mohan and Mr. Sachin Kumar both Research Scholars at CEN for their valuable advice and related resources on this research work.

References

[1] C. W. de Silva, *Vibration Monitoring, Testing, and Instrumentation*, CRC Press (2007).

[2] R. B. Randall, *Vibration-based Condition Monitoring*, Willey, John & Sons, Inc. (2011).

[3] S. Jin and S.-K. Lee, Journal Bearing fault detection utilizing group delay and the Hilbert-Huang transform, *J. of Mechanical Science and Technology*, 31 (3) (2017) 1089-1096.

[4] B. Allison, Validation of single ball rolling contact fatigue machine dynamics, *Journal of Mechanical Science and Technology*, 31 (1) (2017) 37-39.

[5] D. C. D. Oguamanam, H. R. Martin and J. P. Huissoon, On the application of the beta distribution to gear damage analysis, *Applied Acoustics*, 45 (3) (1995) 247-261.

[6] U. Benko, J. Petrov, D. Juricic, J. Tavcar and J. Rejec, An approach to fault diagnosis of vacuum cleaner motor based on sound analysis, *Mechanical Systems and Signal Processing*, 19 (2005) 427-445.

[7] Q. Xiong, Y. Xu, Y. Peng, W. Zhang, Y. Li and L. Tang, Low-speed rolling bearing fault diagnosis based on EMD denoising and parameter estimate with alpha stable distribution, *Journal of Mechanical Science and Technology*, 31 (4) (2017) 1587-1601.

[8] S. Wan, X. Zhang and L. Dou, Compound fault diagnosis of bearings using improved fast spectral kurtosis with VMD, *Journal of Mechanical Science and Technology*, 32 (11) (2018) 5189-5199.

[9] Y. Li, W. Zhang, Q. Xiong, D. Luo, G. Mei and T. Zhang, A rolling bearing fault diagnosis strategy based on improved multiscale permutation entropy and least squares SVM, *Journal of Mechanical Science and Technology*, 31 (6) (2017) 2711-2722.

[10] H. Yuan, J. Chen and G. Dong, An improved initialization method of D-KSVD algorithm for bearing fault diagnosis, *Journal of Mechanical Science and Technology*, 31 (11) (2017) 5161-5172.

[11] A. K. Jain, R. P. W. Duin and J. Mao, Statistical pattern recognition: A review, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22 (1) (2000) 4-37.

[12] H. Qiu, J. Lee, J. Lin and G. Yu, Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics, *Journal of Sound and Vibration,* 289 (4-5) (2006) 1066-1090.

[13] J. Bruna and S. Mallat, Invariant scattering convolution networks, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 35 (8) (2013) 1872-1886.

[14] Y. LeCun, K. Kavukcuoglu and C. Farabet, Convolutional

networks and applications in vision, *ISCAS* (2010) 253-256.
[15] J. Wang, J. Zhang and J. Zhao, Texture classification using scattering statistical and cooccurrence features, *Mathematical Problems in Engineering* (2016).
[16] J. Bruna and S. Mallat, Invariant scattering convolution networks, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 35 (8) (2013) 1872-1886.
[17] J. Bruna and S. Mallat, Classification with scattering operators, *CVPR* (2011) 1561-1566.
[18] S. G. Mallat, A theory for multiresolution signal decomposition: The wavelet representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 11 (7) (1989) 674-693.
[19] J. Andén and S. Mallat, Multiscale scattering for audio classification, *ISMIR* (2011) 657-662.
[20] T. Joachims, The maximum-margin approach to learning text classifiers: Methods theory, and algorithms, *Ausgezeichnete Informatikdissertationen 2001* (2003).
[21] I. Tsochantaridis, T. Joachims, T. Hofmann and Y. Altun, Large margin methods for structured and interdependent output variables, *Journal of Machine Learning Research,* 6 Sep (2005) 1453-1484.
[22] L. Sifre and S. Mallat, *Rigid-motion Scattering for Image Classification,* Ph.D. Thesis (2014).
[23] K. Manjusha, M. A. Kumar and K. P. Soman, Integrating scattering feature maps with convolutional neural networks for Malayalam handwritten character recognition, *International Journal on Document Analysis and Recognition (IJDAR),* 21 (3) (2018) 187-198.
[24] K. C. Gryllias and I. A. Antoniadis, A support vector machine approach based on physical model training for rolling element bearing fault detection in industrial environments, *Engineering Applications of Artificial Intelligence,* 25 (2) (2012) 326-344.
[25] Y. Yu and C. Junsheng, A roller bearing fault diagnosis method based on EMD energy entropy and ANN, *Journal of Sound and Vibration,* 294 (1-2) (2006) 269-277.
[26] Z. Liu, H. Cao, X. Chen, Z. He and Z. Shen, Multi-fault classification based on wavelet SVM with PSO algorithm to analyze vibration signals from rolling element bearings, *Neurocomputing,* 99 (2013) 399-410.
[27] H. Xu and G. Chen, An intelligent fault identification method of rolling bearings based on LSSVM optimized by improved PSO, *Mechanical Systems and Signal Processing,* 35 (1-2) (2013) 167-175.
[28] J. B. Ali, N. Fnaiech, L. Saidi, B. Chebel-Morello and F. Fnaiech, Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals, *Applied Acoustics,* 89 (2015) 16-27.
[29] Y. Zhang, X. Li, L. Gao and P. Li, A new subset based deep feature learning method for intelligent fault diagnosis of bearing, *Expert Systems with Applications* (2018).
[30] T. Berredjem and M. Benidir, Bearing faults diagnosis using fuzzy expert system relying on an improved range overlaps and similarity method, *Expert Systems with Applications,* 108 (2018) 134-142.

Ambika P. S. is a research scholar in mechanical engineering, National Institute of Technology, Calicut, Kerala, India. She is pursuing her Ph.D. degree in the area of rolling element bearing fault diagnosis and prognosis. Her research interests include vibrations, linear algebra and signal processing.

P. K. Rajendrakumar is a Professor and Head at Mechanical Engineering Department, National Institute of Technology, Calicut, Kerala, India. He obtained his Ph.D. in tribology from Indian Institute of Science, Bengaluru. His research interests include contact mechanics, friction, wear, lubrication, biomechanics.

Rijil Ramchand is an Associate Professor at Department of Electrical Engineering, National Institute of Technology, Calicut, Kerala, India. He obtained his Ph.D. degree from Indian Institute of Science, Bengaluru. His research interests include power electronics, Induction motor drives, multilevel inverters, current controlled inverters.