Journal of Applied and Computational Mechanics, Vol. 3, No. 1, (2017), 80-91
DOI: 10.22055/jacm.2017.21576.1108

Shahid Chamran University

Journal of
Applied and Computational Mechanics

A DWT and SVM based method for rolling element bearing fault diagnosis and its comparison with Artificial Neural Networks

Sunil Tyagi[1], Sashi Kanta Panigrahi[2]

[1] Department of Mechanical Engineering, Defense Institute of Advanced Technology, Girinagar, Pune - 411025, India, suniltyagi@tyagination.com
[2] Department of Mechanical Engineering, Defense Institute of Advanced Technology, Girinagar, Pune - 411025, India, panigrahi.sk@gmail.com

Received March 02 2017; revised April 03 2017; accepted for publication April 06 2017.
Corresponding author: Sashi Kanta Panigrahi, panigrahi.sk@gmail.com

Abstract

A classification technique using Support Vector Machine (SVM) classifier for detection of rolling element bearing fault is presented here. The SVM was fed from features that were extracted from of vibration signals obtained from experimental setup consisting of rotating driveline that was mounted on rolling element bearings which were run in normal and with artificially faults induced conditions. The time-domain vibration signals were divided into 40 segments and simple features such as peaks in time domain and spectrum along with statistical features such as standard deviation, skewness, kurtosis etc. were extracted. Effectiveness of SVM classifier was compared with the performance of Artificial Neural Network (ANN) classifier and it was found that the performance of SVM classifier is superior to that of ANN. The effect of pre-processing of the vibration signal by Discreet Wavelet Transform (DWT) prior to feature extraction is also studied and it is shown that pre-processing of vibration signal with DWT enhances the effectiveness of both ANN and SVM classifiers. It has been demonstrated from experiment results that performance of SVM classifier is better than ANN in detection of bearing condition and pre-processing the vibration signal with DWT improves the performance of SVM classifier.

Keywords: Artificial Neural Network (ANN), Discreet Wavelet Transform (DWT), Fault Diagnosis, Rolling Element Bearing, Support Vector Machine (SVM).

1. Introduction

This Rolling element bearings are widely used in plant industry, propulsion systems and automobile. Bearings fault is the foremost cause of machinery breakdown as they are the most common machine element and they work under harsh operating conditions [1]. Proper functioning of machinery depends, to a great extent, on early detection of bearing faults. If not detected well in time, the bearing defect would causes malfunction that may even lead to catastrophic failure of the machinery.

Rolling element bearings endure heavy loads under industrial operating conditions and structural faults; such as wear, pitting, or spall may occur after a long period of running [1]. Components that often fail in rolling element bearing are outer-race, inner-race and the ball. The conventional method of machinery fault detection is to look for peak at the characteristic defect frequency in the frequency spectrum [2]. However, it is not feasible to detect the bearing fault in traditional manner as the bearing defects generate a series of impact vibrations every time a running ball passes over the surfaces of the defects. The resultant vibration in time domain are characterized by sharp peaks and these impact vibrations distributes their energy over wide range of frequencies; the bearing's defect frequency contains low energy [3] and hence can be easily masked by noise and other low frequency effects.

To overcome this problem, both time and frequency domain methods have been developed. They involve methods that usually involve indices sensitive to impulsive oscillations, such as peak level, rms value, crest factor, kurtosis and shock pulse etc.[1]. The passage of rolling element over the fault causes an impact due to sudden change of contact stresses. This excites one or more resonant frequencies of the bearing. Typically the resonant frequencies lie in the high frequency range (> 5 kHz) [3]. Accordingly, many techniques that employ high frequency vibrations in various ways have been developed for bearing fault detection [2]. However, all these techniques are not able to detect the bearing faults with high success especially when the defect is at incipient stage. In recent years soft computing has made great progress and has been applied effectively for machinery fault detection [4]. Various researchers have applied Artificial Neural Networks (ANN) for detection of rolling element faults [5] – [7]. SVM is another potent classification algorithm that has, also been employed by researchers for diagnosis of bearing fault [8]–[10]. As both of these algorithms have been used effectively for bearing fault detection hence a comparative study of ANN and SVM has been done in this paper to identify the best algorithm.

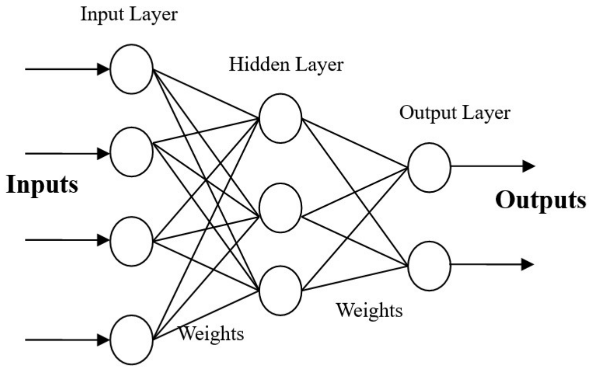

Fig. 1. Schematic representation of Artificial Neural Network

The vibration signals from bearings are inherently non stationary and they can be analyzed very effectively by Discreet Wavelet Transform (DWT) that provides both time and frequency information [11].Wavelet transform also provides a multi resolution analysis of the signal, it gives good time resolution in high frequency range and good frequency resolution at low frequency which makes it ideal for bearing fault detection [12]. Due to these reasons DWT has been used in this paper.

In the present work, the vibration signals were obtained from bearing in normal condition and bearings induced with faults. Features are obtained from vibration signals of bearing running in good and faulty conditions. These features are subsequently used as inputs to the ANN and SVM classifier to train these classifiers to distinguish features of good bearing and defective bearings. In the present approach, sets of normalized features are used so that even if the signals change in magnitude due to the change in speed or quality of sensor mounting, the diagnostic results are unaffected as long as the signal patterns remain unchanged. The effects of different types of features are presented in the paper. Effect of pre-processing by DWT on performance of the classifiers are studied and presented. The performance of classifier to identify different types of bearing faults is also presented in the paper. Finally a procedure is described to correctly categorize the bearing conditions by using SVM and DWT. The procedure is illustrated using the vibration data of a rotating shaft-line with normal and defective bearings.

Section 5 describes the experimental setup used in present research. How features are extracted for training the classifiers and how to the trained classifier is tested is explained in Section 6. Diagnosis of bearing condition using ANN and SVM classifiers are presented in Section 7 and 8 respectively. Section 9 explains the effect of pre-processing the vibration signal by DWT on efficacy of the ANN and SVM classifiers.

2. Artificial Neural Networks

The Artificial neural networks (ANN) are simplified artificial models based on the biological learning p rocess of the human brain [13]. ANNs have been used very extensively in recent years for different application s such as prognosis, classification, function approximation, control filter, pattern recognition etc [14]. Various r esearchers have used ANN for machinery fault detection [15]. The AANs have been successfully used for of b earing fault detection by using them as classifier to separate vibration signals of bearing in good condition and defective bearings[5] – [7]. The ANN is made up of a number of interconnected artificial processing nodes that are called neurons. The neurons are connected together in layers forming a network. A typical ANN is schematically illustrated in Fig. 1 [14]. There are three types of layers; namely input layer, hidden layer and output layer. The number of nodes within the input and output layers are dictated by the nature of the problem to be solved and the number of input and output variables needed to define the problem. The number of hidden layers and the nodes within each hidden layer is usually a trial and error process.

As illustrated in Fig. 2 each node in a layer (except the ones in the input layer) provides a threshold of a single value by summing up their input value $\mathbf{p_i}$ with the corresponding weight value $\mathbf{w_i}$. Then the neuron's net input value **n** is formed by adding up this weighted value (sum), with the bias term b. The bias is added to shift the sum relative to the origin. The net input value then goes into transfer function ***f***, which produces the neuron output ***a***.

$$a = f\left(\sum_{i=1}^{r} w_i \cdot p_i + b\right) \tag{1}$$

The transfer function f that transforms the weighted inputs into the output ***a*** is usually a non-linear function. The sigmoid (S-shaped) or logistic function is the most commonly used transfer function which restricts the nodes output between 0 and 1.

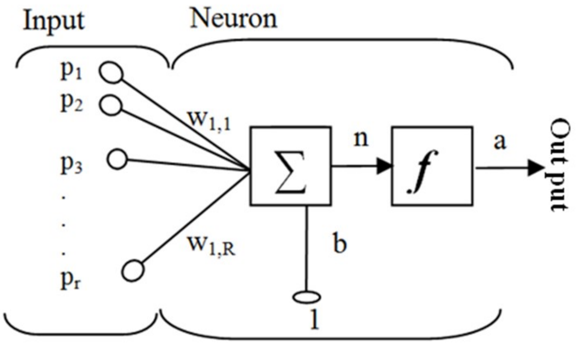

Fig. 2. Processing of input by neuron

Simplest and most common type of ANN is the feed forward network [16]. This is a supervised method of learning mainly used to train multilayer neural networks. In supervised learning, a set of inputs are applied to the network, then the resultant outputs produced by the network are compared with that of the desired targets. If the network is provided with following set of examples for proper behavior:

$$\{p_1,t_1\}\ ,\ \{p_2,t_2\}\ ,\ \ldots\ldots\ ,\ \{p_Q,t_Q\}, \tag{2}$$

where $\boldsymbol{p_Q}$ is an input to network and $\boldsymbol{t_Q}$ is corresponding target. The normalized mean square error (MSE) is calculated and propagated backwards via the network. Back propagation network (BPN) uses it to adjust the value of the weights on the neural connection in the multiple layers. This process is repeated until the MSE is reduced to an acceptably low value, which would be suitable to classify the test set correctly. The mean square error function ***F(x)*** at iteration ***k*** is given by:

$$F(\mathrm{x}) = \left[\left(\mathrm{t}_k - a_k\right)^2\right] \tag{3}$$

BPN uses steepest descent method to adjust the weights and biases. The adjusted weights and biases of $\boldsymbol{m^{th}}$ layer at iteration k are estimated by:

$$w_{i,j}^{m}(k+1) = w_{i,j}^{m}(k) - \alpha \frac{\partial F}{\partial w_{i,j}^{m}} \tag{4}$$

$$b_i^{m}(k+1) = b_i^{m}(k) - \alpha \frac{\partial F}{\partial b_i^{m}} \tag{5}$$

where $\boldsymbol{\alpha}$ is learning rate and $\boldsymbol{w_{i,j}}$ represents weights of connection between neuron ***i*** and neuron ***j***. After the ANN is successfully trained, it should be ready to test data not seen previously. Various algorithms are available to implement back-propagation network most common amongst them is Levenberg-Marquardt algorithm [16] which has been used in this paper.

3. Support Vector Machines (SVMs)

ANNs have proven to be good classifiers but they require large number of samples for training, which is not always true in practice [17]. Support vector machines (SVMs) are based on statistical learning theory and they specialise for a smaller sample number. SVMs have better generalization than ANNs and guarantee the local and global optimal solution similar to that obtained by ANN [18]. In recent years, SVMs have been found to be remarkably effective in many real-world applications [19], [20]. As it is hard to obtain sufficient fault samples in practice, SVMs have been applied for machinery fault diagnosis by various researchers in recent times [21]. SVM has also been successfully used by various researches for detection of ball bearing faults [8]– [10].

SVM is developed from the optimal separation plane under linearly separable condition. Its basic principle can be illustrated in two-dimensional way as Fig. 3 [22]. Fig.3 shows the classification of a series of points for two different classes of data, class ***A*** (circles) and class ***B*** (pentacles). The SVM tries to place a linear boundary ***H*** between the two classes and orients it in such way that the margin is maximized, namely, the distance between the boundary and the nearest data point in each class is maximal. The nearest data points are used to define the margin and are known as support vectors. Suppose there is a given training sample set G={(xi, yi), i=1...l }, each sample $\mathbf{x_i} \in R^d$ belongs to a class by $y \in \{+1, -1\}$. The boundary can be expressed as follows:

$$\omega \cdot x + b = 0 \tag{6}$$

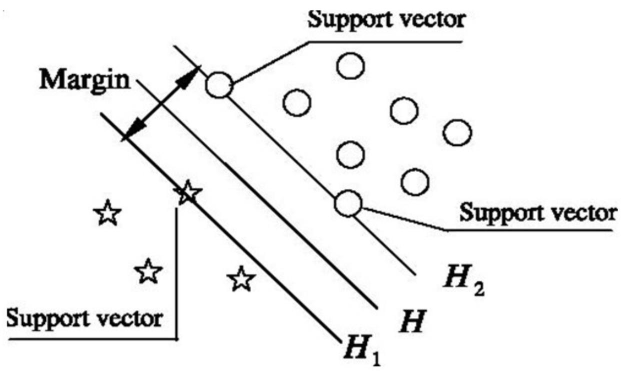

Fig. 3. Classification of data by SVM

where **ω** is a weight vector and **b** is a bias. So the following decision function can be used to classify any data point in either class **A** or **B**:

$$f(x) = \operatorname{sgn}(\omega \cdot x + b) \tag{7}$$

The optimal hyperplane separating the data can be obtained as a solution to the following constrained optimization problem:

$$\text{Minimise } \frac{1}{2}\|\omega\|^2 \tag{8}$$

$$\text{Subjet to } y_i\left[(\omega \cdot x + b)\right] - 1 \geq 0, \quad i = 1, \ldots l \tag{9}$$

Introducing the Lagrange multipliers $\alpha_i \geq 0$, the optimization problem can be rewritten as:

Maximize:

$$L(\omega, b, \alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \tag{10}$$

Subject to:

$$\sum_{i=1}^{l} \alpha_i y_i = 0 \tag{11}$$

The decision function can be obtained as follows

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^{l} \alpha_i y_i (x_i \cdot x) + b\right) \tag{12}$$

If the linear boundary in the input space ***s*** is not enough to separate into two classes properly, it is possible to create a hyperplane that allows linear separation in the higher dimension. In SVM, it is achieved by using a transformation **Φ(x)** that maps the data from input space to feature space. If a kernel function:

$$K(x, y) = \Phi(x) \cdot \Phi(y) \tag{13}$$

is introduced to perform the transformation, the basic form of SVM can be obtained

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^{l} \alpha_i y_i K(x, x_i) + b\right) \tag{14}$$

Among the kernel functions in common use are linear functions, polynomials functions, radial basis functions multi layered perceptron and sigmoid functions.

4. DWT and Multi-resolution analysis

4.1 Discreet Wavelet Transform

DWT have found wide applications in various types of machinery fault diagnosis [23], owing to its ability to treat the transient signals by generating time and frequency representation [11] and also due its capability to give multi resolution analysis [12]. Recently, wavelet transform has been applied for rolling element bearing fault diagnosis [24]-[25]. The wavelet transform is a tool that cuts up data, functions or operators into different frequency components, and then studies each component with solution matched to its scale. The use of wavelet transform is appropriate to analyze non-stationary signal since it gives the information about the signal both in frequency and time domains [28]. Let *x(t)* be the signal. The continuous wavelet transform (CWT) of *x(t)* is defined as:

$$W_\psi(\tau, s) = \int_{-\infty}^{+\infty} x(t) \cdot \psi^*_{\tau,s}(t)dt \tag{15}$$

where $\psi^*_{\tau,s}(t)$ is conjugate of $\psi_{\tau,s}(t)$, that is the scaled and shifted version of the transforming function, called a "mother wavelet", which is defined as:

$$\psi_{\tau,s}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-\tau}{s}\right) \tag{16}$$

The transformed signal is a function of τ and s, the translation and scale parameters respectively. The mother wavelet is a prototype for generating the other wavelet (window) functions. The scale parameter performs scaling operation on the mother wavelet. Each scale represents a frequency band. The term translation corresponds to time information in the transform domain; it shifts the wavelet along the time axis to capture the time information contained in the signal. The DWT is derived from discretization of $W_\psi(\tau, s)$ given by:

$$DWT(j,k) = \frac{1}{\sqrt{2^j}} \int_{-\infty}^{\infty} x(t)\psi\left(\frac{t-2^j k}{2^j}\right) dt \tag{17}$$

An efficient way to implement this scheme was developed by Mallat [29]. The basic working of DWT algorithm is illustrated in Fig.4. The DWT is performed by process of decomposition in which the discreet signal ***x*** is convolved with a low pass filter ***L*** and a high pass filter ***H***, resulting in two vectors ***A*** and ***D***. The vectors ***A*** and ***D*** are down sampled to obtain ***cA1*** called approximate coefficient and ***cD1*** called detail coefficient. In down sampling the odd indexed elements of filtered signals are omitted so that the numbers of coefficients produced in decomposition are equal to the number of elements in the discreet signal ***x(t)***.

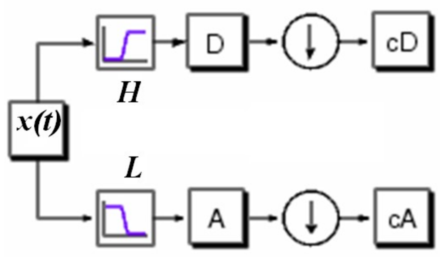

Fig. 4. Decomposition of wavelet Transform

4.2 Multi-Resolution Analysis

The decomposition process can be repeated using approximate coefficients ***cA*** to obtain DWT coefficients at different levels (scale) as per the desired resolution. The process is schematically depicted in Fig. 5.

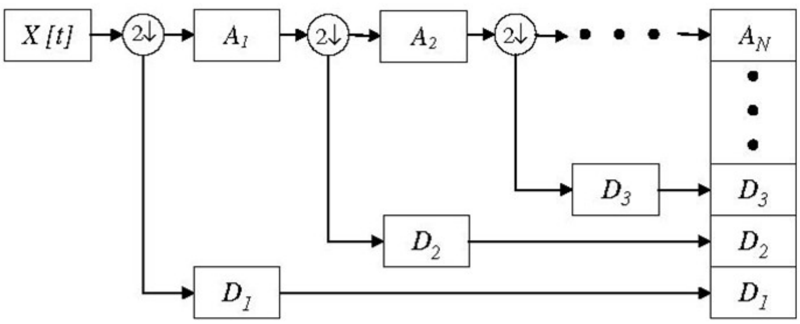

Fig. 5. Decomposition of signal at different levels by wavelet

5. Experimental setup

The test rig shown in Fig. 6 was composed of a variable speed AC motor driving a shaft rotor assembly through flexible couplers; shafts were rested on two ball bearings. A rotor was used for balancing. The bearings under analysis (type MB 204) were placed at load end side for ease of replacement. The load on the system can be adjusted by a manually adjustable magnetic brake, which was driven via a belt drive. Vibration signals were acquired by accelerometer stud mounted on the bearing housing. The faults were artificially introduced to the bearings. The types of faults included a defective outer-race, a defective inner-race, and a defective roller.

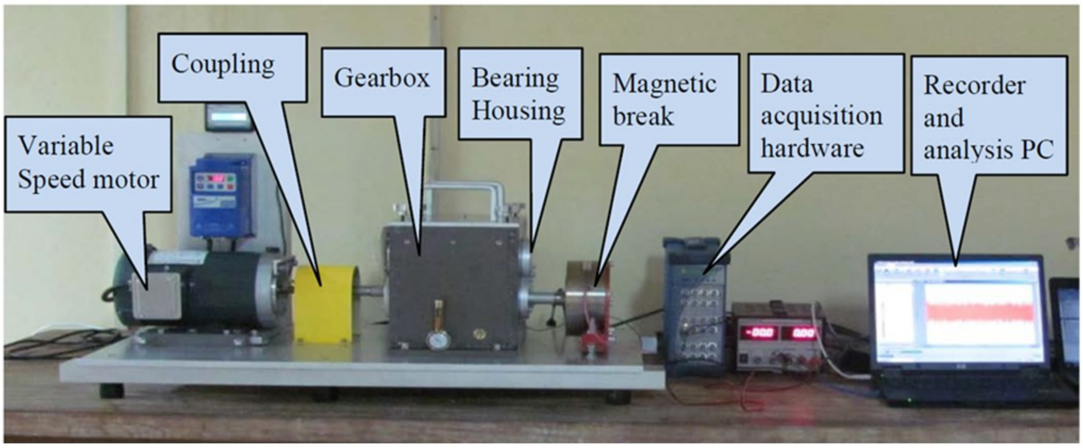

Fig. 6. Experimental setup.

The shaft was made to rotate at 25 Hz and vibration signals were collected at sampling rate of 51.2 KSa/s. The numbers of samples collected were 102400 for duration of 2 s. Following four signals were collected:

1. Bearing in normal condition,
2. Bearing with Outer Race fault (ORF),
3. Bearing with Inner Race fault (IRF), and
4. Bearing with Roller fault (BF).

6. Features and Creation of Training/Test Vectors

6.1 Feature Selection

Each signal of 102400 samples was divided in 40 non overlapping bins of 2560 samples (yi). Ten features were extracted from these 40 bins as follows:

Feature 1-5	- First five highest peaks
Feature 6	- Highest peak of power spectral density (PSD).
Feature 7	- Standard deviation σ.
Feature 8	- Skewness γ3 (third central moment).
Feature 9	- Kurtosis γ4 (fourth central moment).
Feature 10	- Sixth central moment γ6.

The features 6–10 were extracted using:

$$\sigma = \sqrt{E\left\{\left(y_i - \mu\right)^2\right\}}, \gamma_3 = \frac{E\left\{\left(y_i - \mu\right)^3\right\}}{\sigma^3}, \gamma_4 = \frac{E\left\{\left(y_i - \mu\right)^4\right\}}{\sigma^4}, \ \gamma_6 = \frac{E\left\{\left(y_i - \mu\right)^6\right\}}{\sigma^6} \tag{18}$$

where $\mu = E\left(y_i\right)$ is the mean value and *E* is represents the expected value of the function. These features extracted from vibration signals with or without the bearing fault were used for training the ANN and SVM classifier for diagnosis of the bearing condition and testing these classifiers post training. Figure 7 displays all ten features extracted from bearing having outer race fault (ORF), inner race fault (IRF) and ball fault (BF) is plotted against features extracted from rolling element bearing in good condition. It can be seen from Fig. 7 that there is a good separation between features obtained from defective bearings and features obtained from bearing in good condition. Hence these features are good features for classifying the good and bad bearings using the SVM and ANN classifier.

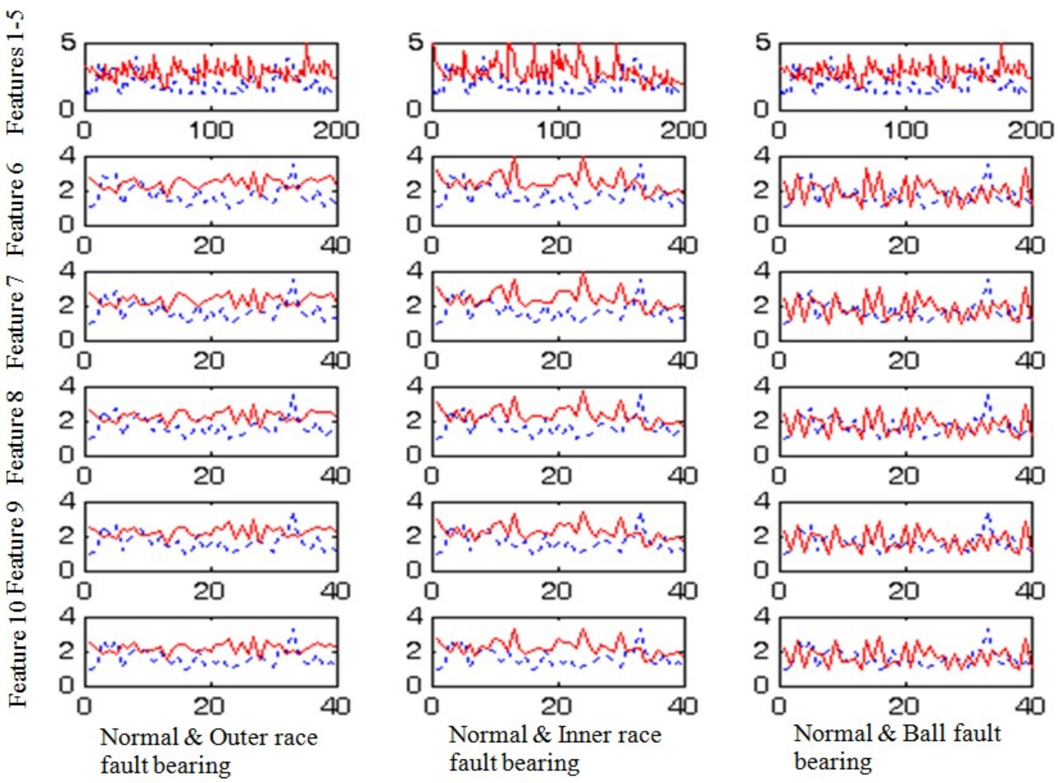

Fig. 7. Features of acquired Vibration signals, ______ defective, _ _ _ _ normal

Table 1: Effect of bearing fault type on identification of machine condition

Case	Input signals	ANN		SVM	
		Training success	Test success	Training success	Test success
1	ORF	48/48 (100%)	74/96 (77.1%)	48/48 (100%)	77/96 (80.2%)
2	IRF	45/48 (93.8%)	74/96 (77.1%)	48/48 (100%)	83/96 (86.5%)
3	BF	45/48 (100%)	85/96 (88.5%)	48/48 (100%)	82/96 (85.4%)
4	ORF, IRF	93/96 (96.9%)	73/96 (76 %)	96/96 (100%)	84/96 (87.5%)
5	ORF, BF	96/96 (100%)	89/96 (92.7%)	96/96 (100%)	85/96 (88.5%)
6	IRF, BF	93/96 (96.9%)	78/96 (81.3%)	96/96 (100%)	88/96 (91.7%)
7	ORF, IRF,BF	141/144 (97.9%)	81/96 (84.4%)	144/144 (100 %)	90/96 (93.8%)

6.2 Creation of training and test vectors

The vibration signal 1–4 each with 102400 samples were divided into 40 bins each having length of 2560 samples. The lengths of bins were selected so that each would contain sufficient number (>5) of impacts caused by passing of the rolling element over the fault. Out of these bins 24 bins were used for training the ANN and SVM classifier. The remaining 16 bins which have not been seen by the ANN and SVM classifier were used for testing. The training sets were created by features extracted from defective signal bins and normal bearing signal bins alternately. Thus three sets of 48 training vectors outer race fault (ORF), inner race fault (IRF) and ball fault (BF), were created for outer race fault, inner race fault and ball fault respectively. Similarly three sets of 32 test vectors were also created. As there were 10 features, therefore a training matrix of 10X144 and test matrix of size 10X96 were created. The diagnostic capability of ANN and SVM classifiers for different faults were also studied by adding/omitting the training sets of respective signal. The numbers of features were also varied to measure their effect.

Table 2: Effect of Input features on identification of bearing conditions

SVM	Test success (Max. 96)	68	54	76	50	59	55	87	75	80	95	89	89	90
SVM	Training success (Max. 144)	129	85	123	94	106	106	141	144	144	144	144	144	144
ANN	Test success (Max. 96)	64	48	73	39	42	91	91	64	80	94	75	84	81
ANN	Training success (Max. 144)	125	72*	122	124	123*	116	140	142	143	142	141	138	141
	Input features	1-5	6	7	8	9	10	6,7,8, 9,10	1-5,7, 8,9,10	1-5,6,8, 9,10	1-5,6,7, 9,10	1-5,6,7, 8,10	1-5,6,7, 8,9	1-5, 6,7, 8,9,10
	Case	8	9	10	11	12	13	14	15	16	17	18	19	7

7. Diagnosis of Bearing Condition using ANN

7.1 Effects of Bearing Defect Type

Table1 shows the results of training and testing the diagnostic capability of the ANN for different input vectors representing different type of bearing faults individually as well as in groups. All ten features were used to study the roles of the bearing defect type. Good training success (93.8 – 100%) was achieved in all cases studied. However the test success varied from 76 – 92.7%. The results of case 1 – 3 indicates that ANN is able to classify correctly for all type of bearing defects even if it is trained with features of only one type of defect (along with features of normal bearing). This indicates to similar nature of impact vibrations produced by different kind of bearing faults. Although 100% training success was achieved when ANN was input with ORF and BF signals; the test success was higher when trained with BF. Results of case 4-6 clearly shows that that the contribution of ball fault signal is most significant for identification of bearing condition as both test and train success was lower (case 4) when features from BF signal was omitted. Case 7 gave best performance when all types of bearing defects were used to train the network.

7.2 Effects of Bearing Defect Type

Table 2 shows the relative importance of signal features for identification of machine condition. For cases 8-19, all three input signals ORF, IRF and BF were used for training. The table presents results using all ten features, namely, the first five highest peaks, highest peak of PSD, standard deviation (σ), skewness (γ_3), kurtosis (γ_4) and sixth central moment (γ_6) either alone or in combination. In cases 8–13, the ANN was trained with only one feature, the contribution of feature 10 i.e. sixth central moment (γ_6) was found most significant as it gave best test success of 94.8 % (91/96) the success with training set was also high (80.6%). In cases 9 and 12 i.e. when network was trained with only peak of PSD (feature 6) and kurtosis (γ_4) (feature 9) the performance goal could not be reached. However, these features were still retained as using them in combination with other features gave good results. Low test results were obtained in case 15 and 18 when feature 6 and 9 were omitted. The use of central moments of order more than six did not have any significant effect on the diagnosis results. The third central moment γ_3 was found to be not a good feature as poor test success of only 40.6 % was obtained when the network was with only γ_3(case 11). Further, good test (97.9 %) and training success (98.6 %) was obtained in case 17 when γ_3 was omitted. In case 19 the training and test success were even better then case 7 which made use of all features. It is thus proposed that all features except feature 8 (γ_3) should be used to train the ANN.

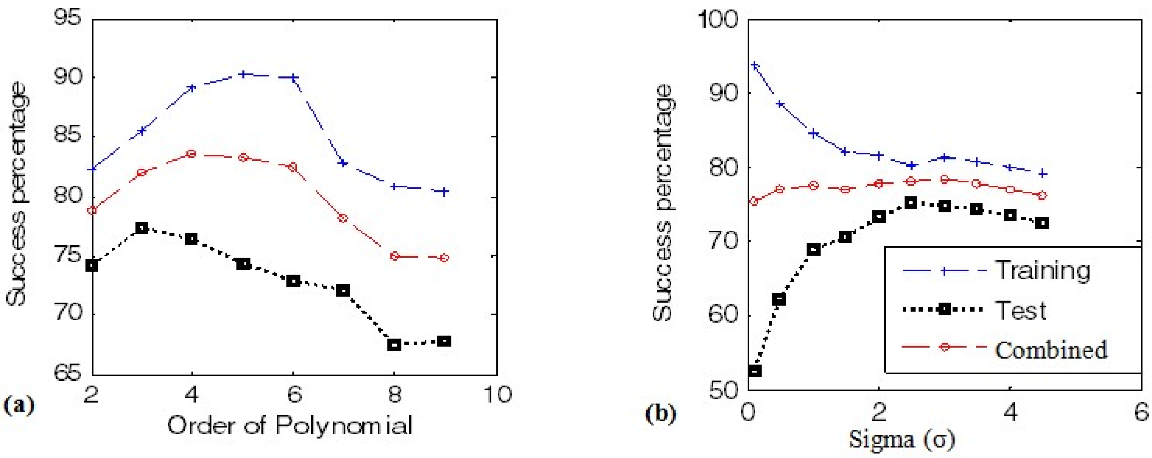

Fig. 8. SVM training results (a) With polynomial kernel (b) With RBF kernel

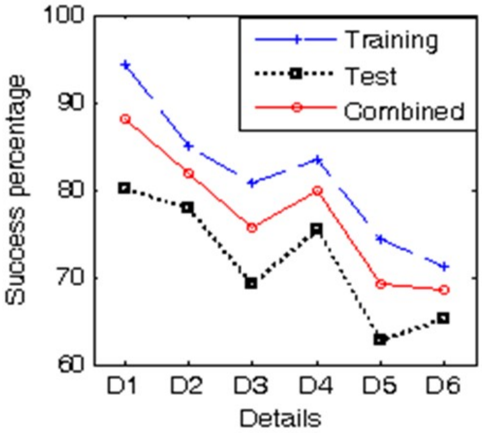

Fig. 9. Effect of DWT details at various levels

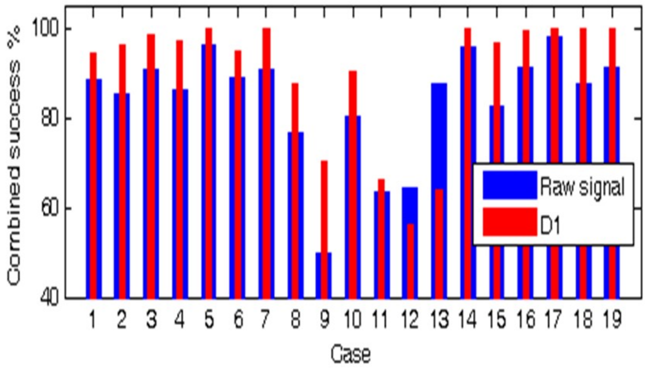

Fig. 10. ANN classification success

8. Diagnosis of Bearing Condition using SVM classifiers

The SVM classifiers was designed for same training and test vectors as used for ANN. Various kernel functions such as Linear, Quadratic, Multilayer Perceptron, Gaussian Radial Basis Function (RBF) and Polynomial kernels were used for all 19 cases. The Linear, Quadratic and Multi layer perceptron kernels could not achieve convergence in many cases and training was stopped after maximum number of iterations (200000) was reached. The RBF and polynomial kernels achieved convergence in all 19 cases. Fig. 8 presents the training, test and combined (test + train) success achieved by Polynomial and RBF kernel functions. The Fig. 8 presents the results by accumulating all 19 cases. The results are presented in percentages combining all 19 cases thus total 2304 training and 1824 test vectors were presented to the SVMs. Fig. 8 (a) presents the sccess achieved by polynomial kernal with respect to the order of polynomial. The order of polynomial for polynomial kernel was varied from 2 to 10 to find out the most optimum order. High combined (i.e. test plus training) success; more than 80 % was achieved when order of polynomial was set between 3 to 6. The test and train success decreased as order of polynomial was decreased below 3. When the order was increased beyond 3 the SVM started to over fit; the training percentage continued to increase whereas the test success percentage fell. The 4th order of polynomial showed good balance of train and test success.

Fig. 8 (b) presents the sccess achieved by RBF kernal with respect to the different values of sigma parameter. The penalty parameter C was kept equal to 1. In case of training with RBF kernel the RBF sigma (σ) was varied from 0.1 to 5 to select the most optimum scaling factor. The test and train percentage increased as sigma was decreased from 5 to 2.5. However, the SVM started to over fit as the sigma was decreased below 2.5. The cumulative percentages in case of RBF kernel was slightly lower as compared to the polynomial kernel. Hence SVM classifier with polynomial kernel of the order of 4 was thus selected for remaineder of this paper.

Three methods were tried to find the separating hyperplane namely Quadratic Programming, Least-Squares and Sequential Minimal Optimization method. The Sequential Minimal Optimization method was selected as it gave best results amongst the three choices. The test and training results for 19 cases using SVM with 4th order polynomial kernel are presented at Table 1 and 2. It is evident from results that more accurate classification of bearing condition is achieved by using SVM classifiers as compared to ANN. Both the training as well as the test successes were higher in case of SVM classifier as compared to ANN for 18 out of total 19 cases (case 13 was only exception). In case 10 and 11 the train success was lower than that achieved by ANN but success with test vectors were higher.

Another significant aspect of using SVM classifiers is the speed of training. The time taken for SVM to train is far less in comparison to the time taken by the ANN. Maximum time taken by the SVM classifier to train was 4.6 sec (case 13). Amongst all cases the ANN achieved fastest training in case 1 where it took 12 sec to train the network. Thus even the slowest case of SVM was much faster than ANN. It is proposed that the SVM classifiers be used for bearing condition classification.

9. Effects of Pre-processing with Discreet Wavelet Transform

The effectiveness of pre-processing the acquired vibration signals by DWT is discussed in this section. Instead of using the raw signal as was the case in previous sections, the vibration signals were pre-processed with DWT using Daubechies wavelet of order 44 (Db44) at level 6 to obtain the low frequency approximate at level six (A6) and the high frequency detail signals at level 1 to 6 (D1-D6). Frequency range of details were in the descending order, i.e. D1 had highest frequency content (12-25.6 kHz), and D6 had the lowest frequency content (0.3–1.2 kHz) whereas frequency content of D2 ranged from 4.6 – 10 kHz. The test and train vectors were created out of these details (D1 - D6) instead the raw signal. The selection of features and creation of training/test vectors were done as per section VI.

Fig. 9 shows the results of using various details (D1 – D6) by the SVM classifier. Total 2304 training and 1824 test vectors created (by accumulating all 19 cases) for each detail and were presented to the SVM classifier. Fig. 9 shows the training, test and combined (training + test) success percentages achieved. High test and training success was achieved when details D1 and D2 were used, indicating the influence of the bearing defect in high frequency range > 5 kHz (as reported by Tandon [3]). The performance of components D3 – D6 outside this frequency range were not very satisfactory. Best results were obtained when D1 was used to extract features, in this case the cumulative (case 1 – 19) training success was 94.3% and test success was 80 %. The Fig. 10 presents (case wise) the combined (training plus test) success obtained by ANN when the network was input with raw signal and with signals pre-processed with DWT (using level 1 details D1). Pre-processing with DWT has improved the success (combined) of classifying the bearing condition in 17 cases out of total 19 cases. The maximum increase was in case 9 where combined success increased from 120 (50%) to 170(70.8%). By pre-processing with DWT the success reduced in only two case i.e. case 12 and case 13.

Pre-processing with DWT also improves the classification of bearing condition by the SVM classifiers. Fig 11 presents the combined (training + test) success achieved by the SVM classifier when input with features extracted from raw signal and from pre-processed signal (D1). The combined success increased in 14 out of 19 cases. The maximum increase was in case 9, whereas four cases i.e.12 – 14, 17 and 18 showed decrease. Another significant effect of pre-processing the vibration signal with DWT was on the number of iterations (epochs) performed by the ANN and SVM classifier to train. Table 3 presents the number of iterations performed by ANN and SVM classifier to train when they were input with features extracted from raw signal and with signal pre-processed with DWT. When ANN was input with DWT processed signal the number of iterations required for training had reduced considerably for all cases except for case 8 and cases 12, 13 (where the training had stopped at 15000 epochs without reaching the performance goal). There has been more than 50% reduction (in training time) in six cases and maximum reduction was in case 10 where the number of epochs required for training reduced by about 90 %.

Similarly when SVM was fed with features of signal preprocessed with DWT there is a reduction in number of iterations performed in training. The reduction was observed in 13 out of 19 cases. The number of training iterations had increased in cases 14 – 19 where the SVM classifier was fed with one feature less. However, when all ten features were fed (case 7) the number of iterations reduced by about 82 %.

Table 3: Number of iterations performed during training

	ANN		SVM	
Case	**Raw Signal**	**D1**	**Raw Signal**	**D1**
1	12	9	128	58
2	18	14	365	88
3	33	16	789	204
4	27	13	532	71
5	32	20	647	113
6	27	25	506	159
7	32	27	550	101
8	106	152	6555	4433
9	15000*	3578	2614	158
10	2275	154	64	133
11	6760	4675	494	108
12	15000 *	15000 *	1567	142
13	11930	15000 *	6668	146
14	56	26	857	23536
15	64	32	6523	17429
16	37	32	821	49833
17	45	27	1154	36153
18	40	42	1023	26232
19	44	24	537	51587

* Training was terminated at maximum number of epochs

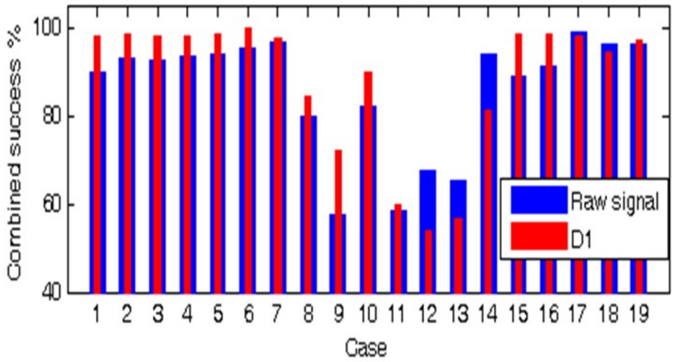

Fig. 11. SVM classification success

10. Conclusion

A method is presented to identify bearing condition by using simple features such as five highest peaks and statistical central moments of time domain vibration signal together with peak of Power Spectral Density. It is shown that using these simple features the bearing condition can be correctly classified with high accuracy with the help of ANN or SVM classifiers. 19 different cases were created to test the efficacy of ANN and SVM classifier in different conditions and it was found that SVM classifier performs better than ANN in almost all cases. Preprocessing with DWT improves the performance of both ANN and SVM classifiers. The test and train success increased in most cases when features extracted from details at level 1 (D1) were used to train ANN and SVM classifier. The DWT preprocessing also significantly lowers the number of iterations (epochs) required to train the ANN and SVM classifiers.

In practice it is difficult to obtain vibration signatures arising out of all kinds of bearing faults such as outer race fault, inner race fault or ball fault. In proposed method the vibration signals from any one type of bearing fault is sufficient to diagnose the bearing condition that may have other type of defect. It is perhaps because the proposed method does not attempt to make use of bearing defect frequency or time domain features; it focuses upon the peaky nature of impact vibrations by using highest peaks and statistical features such as central moments.

The present procedure is used to classify the status of the machine in the form of normal or faulty bearings. There is a scope for its extension to identify fault types and severity levels. Since the SVM classifier training is quite fast, training and test may be done on-line. These issues are subjects for further study.

References

[1] Randall RB, Antoni J. Rolling element bearing diagnostics—a tutorial. Mechanical systems and signal processing. 2011 Feb 28; 25(2):485-520.
[2] Norton MP, Karczub DG. Fundamentals of noise and vibration analysis for engineers. Cambridge university press; 2003 Oct 16.
[3] Tandon N, Choudhury A. A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings. Tribology international. 1999 Aug 31; 32(8):469-80.
[4] Holmes DE, Jain LC. Innovations in machine learning. Springer-Verlag Berlin Heidelberg; 2006.
[5] Bangalore P and Tjernberg LB. An Artificial Neural Network Approach for Early Fault Detection of Gearbox Bearings. IEEE Transactions on Smart Grid 2015, 6(2): 980-987.
[6] Yu Y, Junsheng C. A roller bearing fault diagnosis method based on EMD energy entropy and ANN. Journal of sound and vibration. 2006 Jun 27; 294(1):269-77.
[7] Samanta B, Al-Balushi KR. Artificial neural network based fault diagnostics of rolling element bearings using time-domain features. Mechanical systems and signal processing. 2003 Mar 1; 17(2):317-28.
[8] Yang J, Zhang Y, Zhu Y. Intelligent fault diagnosis of rolling element bearing based on SVMs and fractal dimension. Mechanical Systems and Signal Processing. 2007 Jul 31; 21(5):2012-24.
[9] FernáNdez-Francos D, MartíNez-Rego D, Fontenla-Romero O, Alonso-Betanzos A. Automatic bearing fault diagnosis based on one-class ν-SVM. Computers & Industrial Engineering. 2013 Jan 31; 64(1):357-65.
[10] Wu SD, Wu PH, Wu CW, Ding JJ, Wang CC. Bearing fault diagnosis based on multiscale permutation entropy and support vector machine. Entropy. 2012 Jul 27; 14(8):1343-56.
[11] Daubechies I. The wavelet transform, time-frequency localization and signal analysis. IEEE transactions on information theory. 1990 Sep; 36(5):961-1005.
[12] Peng ZK, Chu FL. Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography. Mechanical systems and signal processing. 2004 Mar 31; 18(2):199-221.
[13] Jain AK, Mao J, Mohiuddin KM. Artificial neural networks: A tutorial. IEEE computer. 1996 Mar 1; 29(3):31-44.
[14] Demuth HB, Beale MH, De Jess O, Hagan MT. Neural network design. Martin Hagan; 2014 Sep 1.
[15] Jardine AK, Lin D, Banjevic D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mechanical systems and signal processing. 2006 Oct 31; 20(7):1483-510.
[16] Hagan MT, Menhaj MB. Training feedforward networks with the Marquardt algorithm. IEEE transactions on

Neural Networks. 1994 Nov; 5(6):989-93.
[17] Zacksenhouse M, Braun S, Feldman M, Sidahmed M. Toward helicopter gearbox diagnostics from a small number of examples. Mechanical Systems and Signal Processing. 2000 Jul 1; 14(4):523-43.
[18] Foody GM, Mathur A. The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by a SVM. Remote Sensing of Environment. 2006 Jul 30; 103(2):179-89.
[19] Guo G, Li SZ, Chan KL. Support vector machines for face recognition. Image and Vision computing. 2001 Aug 1; 19(9):631-8.
[20] Barzilay O, Brailovsky VL. On domain knowledge and feature selection using a support vector machine. Pattern Recognition Letters. 1999 May 31; 20(5):475-84.
[21] Yan W, Shao H. Application of support vector machine nonlinear classifier to fault diagnoses. InIntelligent Control and Automation, 2002. Proceedings of the 4th World Congress on 2002 (Vol. 4, pp. 2697-2700). IEEE.
[22] Vapnik VN, Vapnik V. Statistical learning theory. New York: Wiley; 1998 Sep 16.
[23] Tse PW, Yang W and Tama HY. Machine fault diagnosis through an effective exact wavelet analysis. Journal of Sound and Vibration 2004; 277: 1005–1024.
[24] Peng YH, Yam R. Wavelet analysis and envelope detection for rolling element bearing fault diagnosis—their effectiveness and flexibilities. Journal of Vibration and Acoustics, Transactions of the ASME. 2001; 123:303-10.
[25] Prabhakar S, Mohanty AR, Sekhar AS. Application of discrete wavelet transform for detection of ball bearing race faults. Tribology International. 2002 Dec 31; 35(12):793-800.
[26] Nikolaou NG, Antoniadis IA. Rolling element bearing fault diagnosis using wavelet packets. Ndt & E International. 2002 Apr 30; 35(3):197-205.
[27] Samar VJ, Bopardikar A, Rao R, Swartz K. Wavelet analysis of neuroelectric waveforms: a conceptual tutorial. Brain and language. 1999 Jan 31; 66(1):7-60.
[28] Mallat SG. A theory for multiresolution signal decomposition: the wavelet representation. IEEE transactions on pattern analysis and machine intelligence. 1989 Jul; 11(7):674-93.