Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 97 (2014) 1772 – 1783

Procedia Engineering

www.elsevier.com/locate/procedia

12th GLOBAL CONGRESS ON MANUFACTURING AND MANAGEMENT, GCMM 2014

# Bearing fault diagnosis using wavelet packet transform, hybrid PSO and support vector machine

Rajeswari.C[a]* ,Sathiyabhama.B[b],Devendiran.S[c] & Manivannan.K[d]

[a]*School of Information Technology and Engineering ,VIT University, Vellore, Tamil Nadu, India*
[b]*Department of Computer Science and Engineering, Sona College of Technology, Salem, Tamil Nadu,India*
[c,d]*School of Mechanical and Building Sciences ,VIT University, Vellore, Tamil Nadu, India*

**Abstract**

A new intelligent methodology in bearing condition diagnosis analysis has been proposed to predict the status of rolling bearing based on vibration signals by multi class support vector machine (MSVM), a classification algorithm. Wavelet packet transform (WPT) is used for signal processing and standard statistical feature extraction process. Feature reduction is a method used to deselect the irrelevant features acquired from the large dataset. Recent survey shows feature reduction is used widely in the field of machine learning to discover the knowledge with reduced features. Rough set is hybridized with particle swarm optimization (PSO), an population based stochastic optimization technique, to reduce the features. The efficiency of classification algorithm is compared based on their classification accuracy before and after feature reduction. Four states of bearing health conditions such as normal, defective inner race, defective outer race and defective ball conditions are simulated and used in this proposed work.


Selection and peer-review under responsibility of the Organizing Committee of GCMM 2014

*Keywords:* Experimental test rig ;Wavelet packet transform ; Statistical features;Particle swarm optimization (PSO); rough set ;support vector machine (SVM) and Fault diagnosis .

## 1. Introduction

The dynamic performance of rotating components is highly influential on performance of any rotating machine. Particularly bearings, which is consider as a heart of rotating machinery. Fault finding methodologies of rolling bearings had importance in preventing machinery from failures by giving forewarning. A number of works have

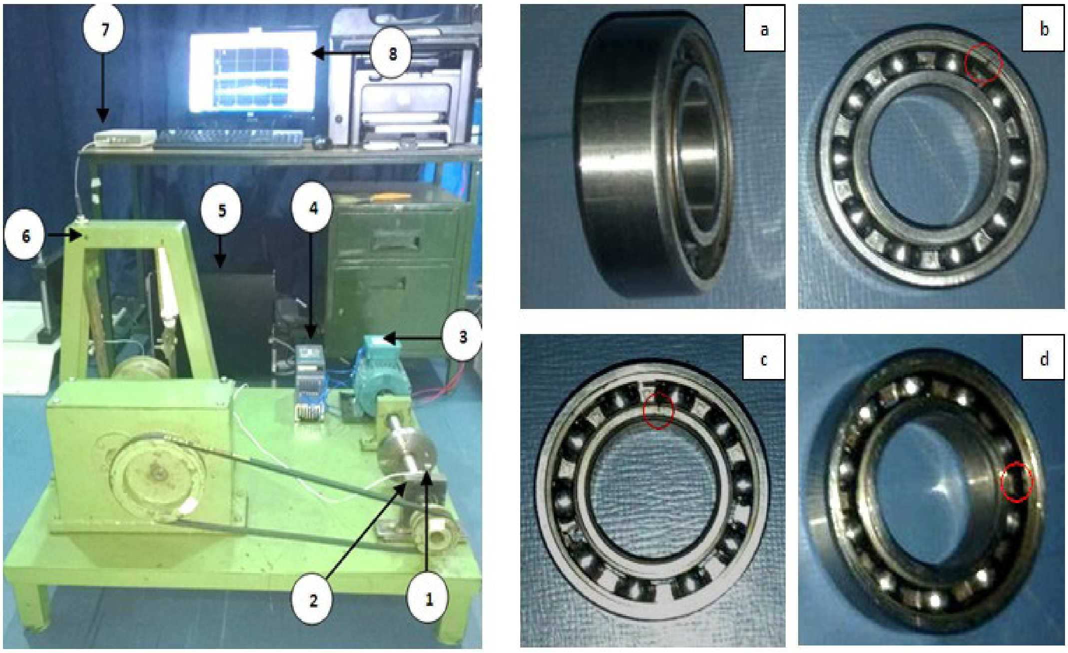
* Corresponding author. Rajeswari.C[a]* Tel.: +91 9600376711.
*E-mail address:* rajeswari.c@vit.ac.in


Selection and peer-review under responsibility of the Organizing Committee of GCMM 2014
doi:10.1016/j.proeng.2014.12.329

been done in this field of bearing condition monitoring using various non-destructive testing technologies such as Current signature Analysis, thermographs, wear debris analysis[3] etc. One of the most preferable and reliable method is vibration analysis technique which is widely used in bearing diagnostics [1-3]. In general the defect diagnosis methods comprise of three stages in vibration based methods. First stage is vibration data acquisition using appropriate sensors , Next step continued by signal processing statistical feature extraction from time domain[4] , frequency domain[5] and time frequency domain to get the useful information .Signal processing methods such as STFT , Laplace transform , wavelet transform using different wavelets[7] were used in more effective way and further as a final step, the condition of rotating component is predictable by maintenance personnel manually or by means of automated computational intelligence methods such as soft computing data mining algorithms such as ANN[9 ],SVM[4],C4.5[ ], etc.. Some of the researchers used feature selection techniques such as Genetic algorithm, [6] Principle component analysis (PCA)[6] and Particle swam optimization technique (PSO)[10] to reduce the dimensionality of the features and saves the computation time.

## 2 . Experimental setup and experimental procedure

The tests for validate the proposed methodology were performed on designed experimental setup. Experimental set-up shown in Fig-1 consists of variable frequency drive (VFD), three phase 0.5 hp AC motor, bearing ,belt drive, gearbox and brake drum dynamometer with scale. A standard deep groove ball bearing (No. 6005) is used in this experiment. Tri axial type accelerometer (Vibration sensor) is fixed over the bearing block to measure the vibration signals. 24 Bit, ATA0824DAQ51 data acquisition system was used and the signals were collected with the sampling frequency of 12800 Hz. Bearing was driven by an motor at a constant rotating speed of 1700 r/min. Constant load was applied by brake drum dynamometer and the speed was monitor by tachometer.




1.Accelerometer
2.Bearing block
3.A.C Motor
4.Variable frequency drive
5.Central processing unit
6.Brake drum dynamometer
7.Data acquisition system(DAS)
8.Computer display

a. Normal bearing
b. Outer race fault bearing
c. Inner race fault bearing
d. Ball fault bearing

Fig.1. Bearing test rig (experimental set up) & various fault conditions of bearing

The Fig-2 indicates normal, outer race fault, inner race fault and ball fault conditions were formed using the wire cut EDM process by the depth of cut 0.4 mm throughout the width of the component and the induced fault size is maintained for all three components. The experiment of acquiring vibration signal is carried out for all bearing conditions (Shown in fig-2).The number of the sample data for each bearing condition is depicted in Table 2. Totally 40 samples (containing 5000 data points for each sample) taken for each state of bearing condition and totally 160 samples for all state of bearing condition. One example of the four conditions is in time domain and frequency domain illustrated in Fig. 3(a) &3(b). Commonly used transformation technique in health monitoring is Fast Fourier transform (FFT) , which is used to transform the time series data to frequency domain, where the signal is used to deduction of sine and cosine waves from the sample. FFT was also executing on sample signals for all the states of bearing. The bearing can be diagnosed by analysing the abnormal frequency-domain amplitude. The frequency of the abnormal vibration is called fault frequency which is decided by the fault location. The following equations will give the detail of fault characteristic frequencies for different parts of bearing .The characteristic bearing frequencies are BPFO- Ball Pass Frequency Outer Race, BPFI- Ball Pass Frequency Inner Race, FTF- Fundamental Train Frequency and BSF- Ball Spin Frequency. These characteristic frequencies illustrated in Table-2 were useful to find the bearing component defect from the concern component frequencies and its harmonics. In tradition, Frequency analysis may be the most fundamental approach for bearing condition monitoring and fault detection, finding those frequencies and measuring the amplitude variations in the particular characteristic frequency and its side bands as well the harmonics of those frequencies will give the information of the health condition of bearing components. Even though the bearing conditions are difficult to be differentiate by their FFT spectral accurately. The experimental bearing fault diagnosis methodology is illustrated in Fig. 4.

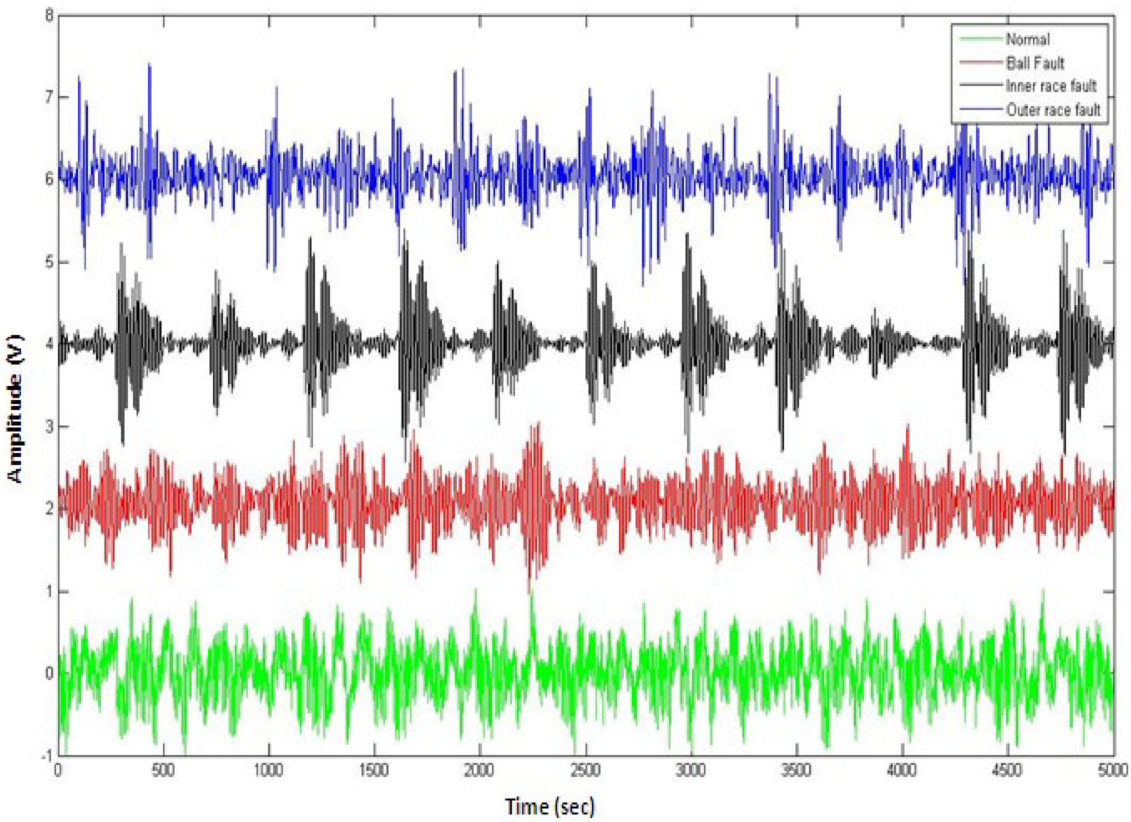



Fig.2. Time domain signals for various fault conditions of bearing

### 3. Signal processing and feature extraction using wavelet packet transform.

Wavelet transform (WT) is a time-frequency decomposition of a sample signal into "wavelet" basic function. Wavelet analysis is widely used for decomposing, de-noising and signal analysis over a non-stationary signal. At high frequencies WT gives good time and poor frequency resolution, and at the same time at low frequencies it gives good frequency and poor time resolution. Recently wavelet transform is getting an good popularity in mechanical vibration signature analysis applications [11]. Investigation with wavelets proceed with breaking up a signal into shifted and scaled versions of its mother (or original) wavelet, that is obtaining one high frequency term from each level and one low frequency residual from the last level of decomposition. Unlike conventional techniques, wavelet decomposition produces a family of hierarchically organized decompositions. The selection of a suitable level for the hierarchy will depend on the signal and experience. Often the level is chosen based on a desired low-pass cut off frequency. The wavelet packet method is a generalization of wavelet decomposition that offers a richer range of possibilities for signal analysis. In wavelet analysis, a signal is split into an approximation and a detail. The approximation is then itself split into a second-level approximation and detail, and the process is repeated. For n-level decomposition (Fig 4), there are n+1 possible ways to decompose or encode the signal. General practice of wavelet packet decomposition will stop until bottom detail reach the single data .But in the practical application, the decomposition layer is always decided by the characteristics of the signal and actual needs [8].Wavelet packet transform(WPT) is used for signal multi-band filtering , and also used for de-noising [12]. Consideration of the best decomposition level of wavelet packet plays an important role in bearing fault diagnosis and criteria for maximum decomposition level explained in [13,14].

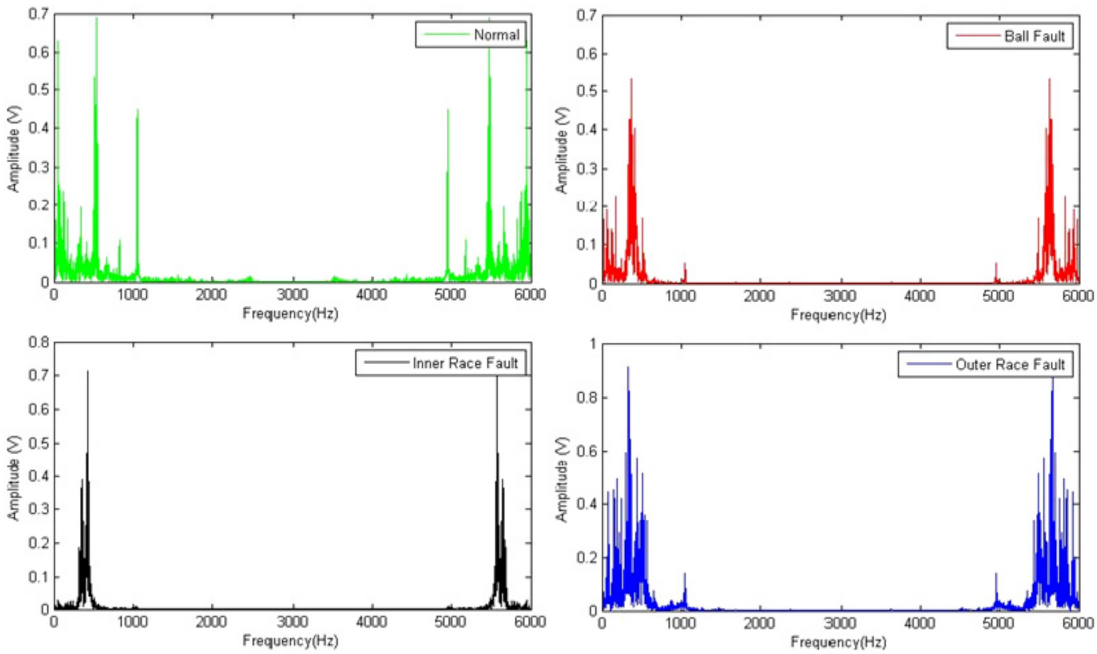



Fig.3. Frequency domain signals for various fault conditions of bearing

Table 1. Discretized sample data set

| Elements | p | q | r | s | T |
|---|---|---|---|---|---|
| 1 | -1 | -1 | 1 | 0 | 1 |
| 2 | 1 | 0 | -1 | -1 | 0 |
| 3 | 0 | 1 | 1 | 1 | -1 |
| 4 | -1 | 0 | 0 | 1 | 1 |
| 5 | 1 | 1 | 0 | -1 | -1 |
| 6 | 1 | 0 | -1 | 0 | 1 |
| 7 | -1 | -1 | 0 | 1 | 1 |
| 8 | -1 | 1 | 1 | 1 | -1 |

The WPT after decomposition gives approximate as well as the details about the given signal. The wavelet packet transformation is given as

$$\left\{\begin{array}{l} D(i,2n)=\sum_j A_{j-2l}Dj(i-1,n) \\ D(i,2n+1)=\sum_j B_{j-2l}Dj(i-1,n) \end{array}\right\} \tag{1}$$

Where $A_j$ and $B_j$ are the wavelet packet coefficients. As per the Parseval equality shown in Eqn (2), the summation of WPT coefficients $D(i,j)$ is equal to the original time domain vibration signal. For this reason feature extraction from WPT coefficients considered as reliable.

$$\int_{-\infty}^{\infty}\left|f(x)\right|^2 dx=\sum\left|D(i,j)\right|^2 \tag{2}$$

As per this direction, in this present work the wavelet packet transform method is used to obtain subset of the features from their coefficients .These features form a transformed space and it is used as the input of next process called classification. The signal is decomposed with 3-layer (i=3) and wavelet packet coefficient were obtained from eight node (shown in Fig.5).

Table 2. Characteristic frequency

| Characteristic frequency | Equation |
|---|---|
| Shaft rotational frequency, (Fs ) | $Shaftspeed/60$ |
| Ball passing frequency outer race , (BPFO) | $Fs\left(\frac{N_b}{2}\right)\left(1-\frac{B_d}{P_d}\cos\phi\right)$ |
| Ball passing frequency inner race (BPFI) | $Fs\left(\frac{N_b}{2}\right)\left(1-\frac{B_d}{P_d}\cos\phi\right)$ |
| Fundamental train frequency (FTF) | $Fs\left(\frac{1}{2}\right)\left(1-\frac{B_d}{P_d}\cos\phi\right)$ |
| Ball spin frequency (BSF) | $Fs\left(\frac{P_d}{2B_d}\right)\left(1-\frac{B_d^{\,2}}{P_d^{\,2}}\cos^2\phi\right)$ |

Table 3.Statistical features

| Statistical feature | Explanation | Equation |
|---|---|---|
| Kurtosis | It denotes the data are peaked or flat relative to a normal distribution. | $k_{kur}=\frac{1}{n}\sum_{i=1}^{n}\left(k(t)-\bar{k}\right)^4$ |
| Skewness | It denotes the symmetry of the data, direction and extent of the data skew. | $k_{skew}=\frac{1}{n}\sum_{i=1}^{n}\left(k(t)-\bar{k}\right)^3$ |
| Variance | It is a measure of dispersion of a waveform about its mean. | $k_{\text{var}}=\frac{1}{n}\sum_{i=1}^{n}\left(k(t)-\bar{k}\right)^2$ |
| Standard deviation | It indicates the energy contenting in the vibration signal. | $k_{Sd}=\sqrt{\left(\frac{1}{n-1}\sum_{i=1}^{n}(k_i-\mu)^2\right)}$ |
| Root mean square | It indicates the severity of a bearing defect by the signal intensity. | $k_{rms}=\sqrt{\left(\frac{1}{n}\sum_{i=1}^{n}k^2\right)}$ |

After getting the eight node components (shown in Fig.5), i.e., node W3,0, W3,1, W3,2.......W3,7 , statistical features used by many of the researchers in the field of fault diagnosis [15]such as Standard deviation (STD),Root mean square(RMS), Peak to peak (PP) and Kurtosis(KUR) were calculated(illustrated in Table-3) for each node, thus resulting in a total of 40 features (8x5) were calculated for a single sample for each state of bearing condition.

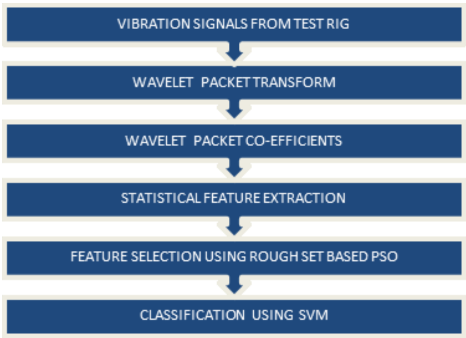



Fig.4. Methodology of bearing condition diagnosis

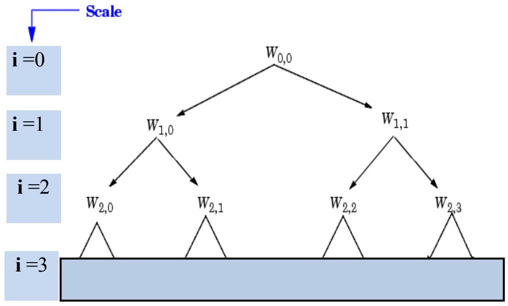



Fig.5. Three layer Wavelet packet decomposition

## 2. Introduction to PSO

Particle Swarm Optimization (PSO) is a progressing computation method developed by Kennedy and Eberhart in 1995[20]. The basic idea was to capture the random actions performed collectively by a group of birds or fishes. Early model were modified by establishing weights to fabricate the regular PSO. Moreover PSO can be used on optimization problems that are partially asymmetrical, noisy, transform over time, etc. PSO is initialized with a population of arbitrary solutions, called ‘particles’. Each particle is considered as a position in the space S. The $x$ th particle is signifies as $A_x = (a_{x1}, a_{x2}, ...a_{xs})$ .The best previous position *Pbest*, (the point giving the best fitness value) of any particle is recorded and correspond to as $P_x = (p_{x1}, p_{x2}, ...p_{xs})$ .The directory of the best particle between every particles in the population is symbolized by ‘*gbest*’. The pace of the position transform (velocity) for particle $x$ is symbolized as $V_x = (v_{x1}, v_{x2}, ...v_{xs})$ . The particles are influenced by the following equation:

$$V_{xn} = w \times v_{xn} + C_1 \times rand() \times (p_{xn} - a_{xn}) + C_2 \times Rand() \times (p_{gn} - a_{xn}) \quad (3)$$

$$a_{xn} = a_{xn} + v_{xn} \quad (4)$$

Where $n$ = 1, 2. . . . S, *w* is the inertia weight; it is a temporal positive linear function varying according to the generation iteration. Choosing the appropriate inertia weight offers equilibrium among global and local searching, and consequently the number of iteration to locate the optimal solution will be less. $C_1$ and $C2$ are the two acceleration constants present in Eq. (3) corresponds to the weighting of the acceleration terms that has potentials to draw each particle toward *pbest* and *gbest* positions. If values are smaller it permits the particles to travel far away from objective regions ahead of being pulled backside, whereas if the values are higher it effects in unexpected progress towards the objective regions. The domain of *rand* ( ) and *Rand* ( ) are defined between [0, 1]. *Vmax* is the particles’ maximum velocity that is restricted on all dimensions. If the value of *Vmax* is excessively small, particles may possibly not travel around adequately away from locally good regions. On the other hand, if the value of *Vmax* is excessively high, particles might fly past good solutions. Eq. (3) provides information about three components; the memory capability, the “flying particles” have on the search space; “cognition”, which represents the confidential opinion of the particle itself; “social”, which represents the teamwork among the particles. Eq. (4) is used to calculate the particle’s new velocity according to its previous velocity and the distances of its current position from its own best experience (position) and the group’s best experience. Then the particle flies toward a new position according to Eq. (4). The performance of each particle is calculated according to a pre-defined fitness function.

## 3. Theoretical background of Rough Set theory

Rough Set (RS) theory is one of the statistical tool deals with feature reduction that is likely to uncover a nominal subset. The fundamental idea of RS has three basic components: an upper and a lower approximation of a data set, the approximation space and models of data sets. The feature reduction is realized by weighing against equivalence relations produced by feature sets. Employ the dependency degree as an significant measure, features are detached and reduced feature set gives the equivalent dependency degree as the original. For further detailed description about the roughest theoretical background refer [21]. For better understanding feature reduction using Rough set is explained with example in the following section. Features have to be discretized before proceeding with data mining algorithms .The process of converting the numerical variables to categorical variables is called as discretization.. The most common, binning discretization method is used in this proposed work. One of the rough set based Relative Reduct (RR) algorithm is used as it keep away from the computation of discernability task or positive regions, which can be computationally expensive without optimizations. According to the algorithm, the dependency of each feature is calculated and the finest candidate is selected [21, 22].

Table 4- Normalized training sample feature values extracted from 3 layers WPT for 4 different states of bearing.

| Sample Number. | $W_{30}$ | $W_{31}$ | $W_{32}$ | $W_{33}$ | $W_{34}$ | $W_{35}$ | $W_{36}$ | $W_{37}$ | STATES | TARGET |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.720021 | 0.078644 | 1 | 0.996521 | 1 | 0.93865 | 0.991471 | 0.981329 | NORMAL | 1000 |
| 2 | 0.85042 | 0 | 0.90546 | 0.950556 | 0.831674 | 0.953284 | 0.879267 | 0.980961 | NORMAL | 1000 |
| . | . | . | . | . | . | . | . | . | . | |
| 39 | 0.480691 | 0.642513 | 0.104047 | 0.048895 | 0.07459 | 0.01784 | 0.001339 | 0.000136 | BALL FAULT | 0001 |
| 40 | 0 | 1 | 0.112192 | 0.047918 | 0.108179 | 0.013748 | 0.003284 | 0.000994 | BALL FAULT | 0001 |

A knowledge representation system can be formulated as follows; knowledge representation is a pair $K=(S,F)$, where $K-$ knowledge; $S-$non empty, finite set called Space; $F-$nonempty, finite set of primitive features. Every primitive feature $f \in F$ is a complete function, $f: S \rightarrow I_f$ where $I_f$ holds the domain of $f$. With every subset of features $E \subseteq F$, we correlate a binary relation $IND(E)$, called an indiscernibility relation and expressed as;

$$IND(E)=\{(x,y)\in S^2 \mid \forall f \in E, f(x)=f(y)\} \tag{5}$$

The partition of $S$ generated by $IND(E)$ is denoted as $S/E$. If $(x,y)\in IND(E)$, then x and y are indiscernible by features from E. The equivalence classes of the E-Indiscernibility relation are represented by $[x]_E$. Let $A \subseteq S$ the E-lower approximation $\underline{E}A$ and E-upper approximation $\overline{E}A$ of set A can be stated as:

$$\underline{E}A=\{x\in S \mid [x]_E \subseteq A\} \tag{6}$$

$$\overline{E}A=\{x\in S \mid [x]_E \cap A \neq 0\} \tag{7}$$

Let $E,G \subseteq X$ be equivalence relation over S, then the positive, negative and boundary region can be stated as:

$$POS_E(G)=\bigcup_{A\in S/G}\underline{E}A \tag{8}$$

$$NEG_E(G)=U-\bigcup_{A\in S/G}\overline{E}A \tag{9}$$

$$BND_E(G)=\bigcup_{A\in S/G}\overline{E}A-\bigcup_{A\in S/G}\underline{E}A \tag{10}$$

The E-positive region of G is the set of all item of the space S which can be definitely classified to classes of $S|G$ Employing knowledge stated by the classification $S|E$. G relies on E in a degree, $d(0\le d\le 1)$ denoted by $E\rightarrow_d G$

$$d=\gamma_E(G)=\frac{|POS_E(G)|}{|S|} \tag{11}$$

Where, E is the set of conditional features, G is the set of decision feature, $\gamma_E(G)$ is the quality of classification. If d=1, then G depends totally on E; if 0<d<1, then G depends partially on E; if d=0 then G does not depend on E. The objective of feature reduction is to remove the repeating features so that the reduced set provides original classification results. The set of all reduct is defined as, $Reduced(Y) = X \subseteq Y \mid \gamma_X(D) = \gamma_Y(D)$, is a reduct of $Y$ if $X$ is independent and $IND(X) = IND(Y)$; $Y$ can have many feature reducts and set of all indispensible relations in $Y$ is :

$$Reduced(Y)_{optimal} = \{X \in Reduced \mid \forall X' \in Reduced, |X| \le |X'|\} \quad (12)$$

**Example for feature reduction using relative reduct.**

Step 1: From the Table-1, Dependencies of individual attributes as calculated with respect to decision attribute and relative dependency measure and fitness value for a particle is defined as $\gamma_X(T) = \dfrac{|U/IND(X)|}{|U/IND(X \cup T)|}$

(13) $U/p = \{1,4,7,8\}\{2,5,6\}\{3\}$ ; $U/q = \{1,7\}\{2,4,6\}\{3,5,8\}$ ;

$U/r = \{1,3,8\}\{2,6\}\{4,5,7\}$ ; $U/s = \{1,6\}\{2,5\}\{3,4,7,8\}$ ; $U/T = \{1,4,6,7\}\{2\}\{3,5,8\}$ ;

$U/IND(p,T) = \{1,4,7\}\{2\}\{5\}\{6\}\{3\}$ ; $\Rightarrow K_p(T) = \dfrac{|U/IND(p)|}{|U/IND(p,T)|} = \dfrac{3}{5} = 0.6 < 1; similarly,$

$U/IND(q,T) = \{1,7\}\{2\}\{4,6\}\{3,5,8\}$ $\Rightarrow K_q(T) = 0.7 < 1$ ;

$U/IND(r,T) = \{1\}\{2\}\{3,8\}\{4,7\}\{5\}\{6\}$ $\Rightarrow K_r(T) = 0.3 < 1$ ;

$U/IND(s,T) = \{1,6\}\{2\}\{3,8\}\{4,7\}$ $\Rightarrow K_s(T) = 0.6 < 1$ ;

Step 2: Dependency values of combined attributes are calculated with respect to decision attribute.

$U/IND(p,q) = \{1,7\}\{2,6\}\{3\}\{4\}\{5\}\{8\}$ $\Rightarrow K_{(p,q)}(T) = \dfrac{|U/IND(p,q)|}{|U/IND(p,q)T|} = \dfrac{6}{6} = 1$

As the stopping criteria for dependency estimation is equal to 1, therefore the dependency estimation for other combination of attribute is not necessary. Hence, {p, q} is the reduced set for the given problem domain.

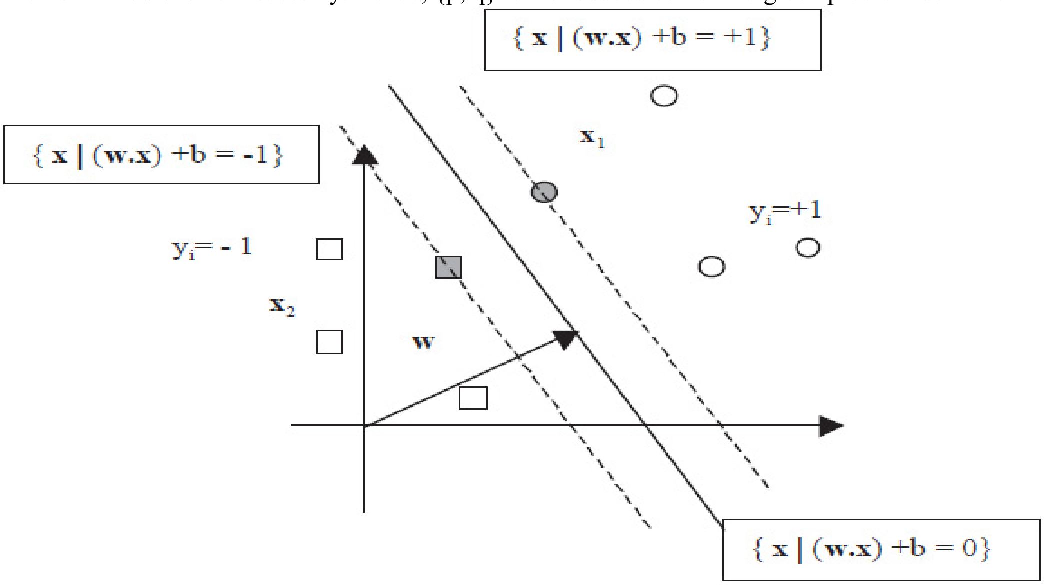



Fig .6 Two class Classification by SVM

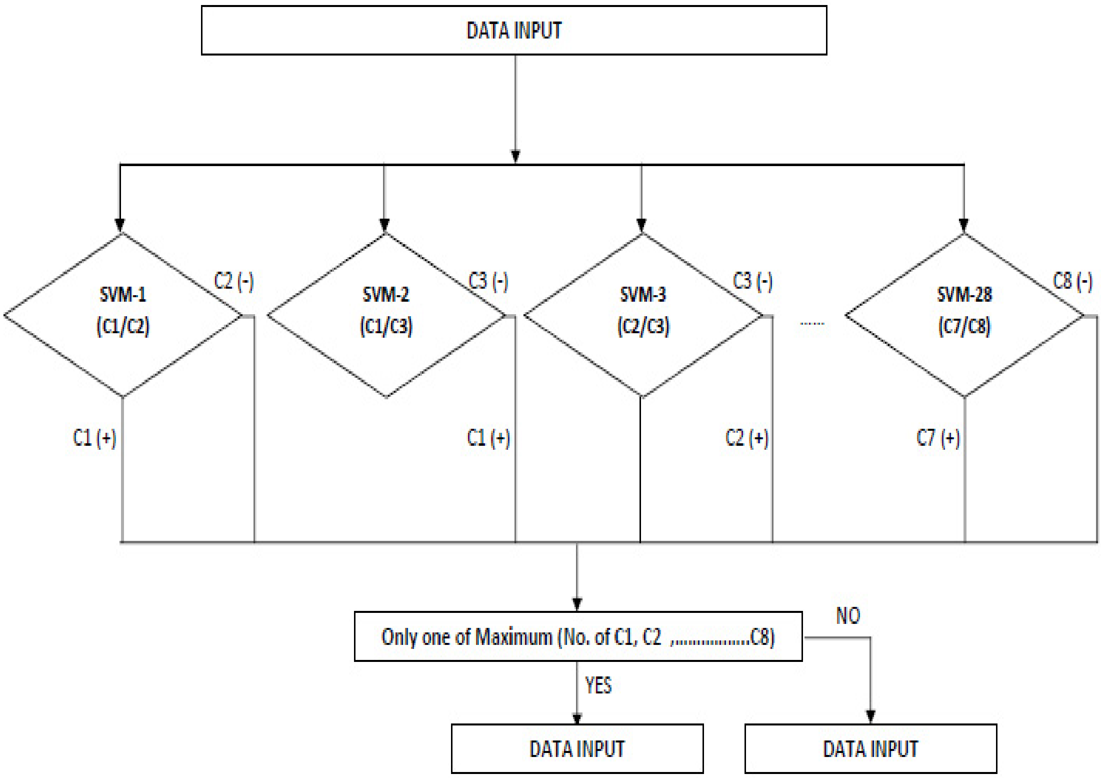



Fig .7 One against One approach of MSVM

Feature selection based on PSO-relative reduct is given in Algorithm part. This method calculates a reduct set than generating all possible subset. As mentioned in the PSO section initial process of random selection of particle and velocity continues except the optimized position and velocity findings are different and based on dependency measure and fitness (Eqn.13) of decision features. As like binary Genetic algorithm each particle 1's are taken as the selected feature and 0's are considered as removed feature. The position and velocity are updated and the new population is generated and the fitness values are computed for each particle until fitness value of the selected feature subset becomes 1[22].

## 4. Theoretical background of Support Vector Machines (SVMs)

The hyper plane separates the data in the binary problem, which is the idea of SVM. The hyper plane defines the use of support vectors and defines the boundary between the two classes. Fig.6 shows how classification done for continues points into two classes of data, class A (circles) and class B (squares) and margin denoted by dotted lines is maximized .SVM finds the distance between the boundary and nearest data in each class which is maximal . The boundary is adjusted in the middle of the margin. Margin is defined by nearest data called support vector (SV). Once SV was selected other points can be omitted [19]. Mathematical description of SVM described as follows.

With the use of support vectors the boundaries are given as,

$$w \cdot x + b = 0, w \in R^N \ , b \in R \qquad \text{(i)}$$

Where w defines the boundary, the input vector of dimension N is represented as x and B is the scalar threshold, support vectors are located at the margin, and equation for A and B are,

$$(w \cdot x) + b = 1 \text{ and } (w \cdot x) + b = -1 \qquad \text{(ii)}$$

Decision function is created to specify the given data point of A or B, for given class,

$$f(x) = sign((w \cdot x) + b), \qquad \text{(iii)}$$

The solution of optimal hyper plane is,
Minimize

$$\tau(w) = \frac{1}{2} \|w\|^2 \qquad \text{(iv)}$$

Subjected to

$$y_i \ ((w \cdot x_i) + b) \geq 1, \ i = 1,....,l, \qquad \text{(v)}$$

Where number of training set is l and quadratic programming (QP) optimization problem is obtained as,

$$w = \sum v_i x_i \ , \qquad \text{(vi)}$$

$X_i$ are Support Vectors, classification function is obtained,

$$f(x) = sign\left(\sum_{i=1}^{l} v_i (x \cdot x_i) + b\right), \qquad \text{(vii)}$$

In case linear boundary could not separate the classes accurately, the hyper plane is created which permits the separation in higher dimensions. This is accomplished through the transformation φ(x) by transforming N dimensional input space to Q dimensional feature space.

$$s = \phi(x), \qquad \text{(viii)}$$

By substituting transformation (8) in (7),

$$f(x) = sign\left(\sum_{i=1}^{l} v_i (\phi(x) \cdot \phi(x_i)) + b\right), \qquad \text{(ix)}$$

The transformation is computationally intensive so kernel function is used. Thus the function must be positive definite and continuous. The kernel function is defined as,

$$K(x, y) = \phi(x) \cdot \phi(y), \qquad \text{(x)}$$

Decision function is modified as,

$$f(x) = sign\left(\sum_{i=1}^{l} v_i K(x, y) + b\right), \qquad \text{(xi)}$$

$V_i$ is used as weighting factors for determining the input vectors that are the actual support vectors, some of the kernel functions are:
The linear kernel

$$K(x, y) = x \cdot y, \qquad \text{(xii)}$$

The polynomial kernel

$$K(x, y) = (\langle x \cdot y \rangle + 1)^p$$ ,Where p is the power number of polynomial. (xiii)

The radial basis kernel(RBF)

$$K(x, y) = \exp(-|x - y|^2 / 2\sigma^2), \qquad \text{(xiv)}$$

The sigmoid kernel

$$K(x, y) = \tanh(v \cdot < x \cdot y > + c), \qquad \text{(xv)}$$

$V_i$ parameter is limited to reduce the effect in outliers in boundary by SV where there is a overlap between classes with non separable data.

$$0 < v_i < C. \qquad \text{(xvi)}$$

Parameter $C$ is a penalty constant for sample points, which are not separated by optimal separation plane. Depending on the number of errors $C$ is infinity for separable case; it may vary for non separable case. The parameters like $C$ and σ is used to control the generalization capability of SVM.

### *4.1 One-to-many SVM classification (Multi class SVM) Implementation*

The above mentioned discussion deals with two class(binary) classification consist only two values ± 1 , but in real time problems, more than two classes has to be solved or classified. Especially in machine faults diagnosis Multiclass classification problems can be solved by one of the voting schemes such as one against all and one against one approach were used [16-17]. These methods decompose a multi-class problem into a number of binary classification problems. Among these two methods the later given the best accuracy even it takes more computation time [18] .Thus in this work one-against-one method is applied to detect the faults of bearings. The flow chart of the working process of one against one is shown in Fig 8. In this method, for *k* classes, will results in *k(k-1)/2* = number of binary SVM classifiers . For example, if k=8, one needs to train 28 binary classifiers and classify them according to the classes which get the highest number of votes. Input feature for training (sample is illustrated in Table .4) consist of all five feature data and fed into SVM. Matlab R2012a platform is used for execution. For SVMs, the RBF kernel function is used and width was selected in the range of 0.2–2.0, a step size of 0.2. Among all extracted features 60% has taken for training the SVM and rest of the 40% for testing the SVM.

**Algorithm-1** : PSO and roughset based relative reduct

**Algorithm**: PSO-RSRR(S,T)
**Input**: S set of all independent features; T, set of decision features
**Output**: X, reduced set

**Step 1**: Initialize P with random position,
$\forall : P_r := randposition()$ ; $V_r := randvelocity()$
fitness:=0; Global best:=fitness; Gbest:=P1; Pbest:=P1;
For r=1,2...C
  Pbest(r) = $P_r$
  Fit(r)=0
End For
**Step 2**: While Fit! =1
For r=1,2...C
  $\forall :$ $P_r$
  X:= {}
  Evaluate fitness function for subset of $P_r$
  X: = feature subset of $P_r$
  $\forall$ x $\in$ S

$$\gamma_X(T) = \frac{|U/IND(X)|}{|U/IND(X \cup T)|}$$

Fitness= $\gamma_{X-\{x\}}(T), \quad \forall P \subset X, \quad \gamma_P(T) \neq \gamma_S(T)$
If Fit(r) > fitness
  Fit(r) = fitness
  Pbest(r) = $P_r$
End
If fitness==1
  Return X
End If
End For
**Step 3**: Evaluate best fit
For r=1:C
If (Fit(r)>Global best
  Gbest:= $P_r$ ; Global best:=Fit(r); Gbest:= $P_r$
End If
End For
  Update velocity()
  Update position()
End (While)
**Output**: reduced set, X

## 5. Conclusions

In this proposed work, the real time experimental vibration data sets are collected and used to evaluate the performance in fault diagnosis of rolling element bearings. As well, WPT used for feature extraction and procedure for relative reduct rough set method as well relative reduct hybrid with PSO algorithm were used for feature selection and is given in detail. By this approach optimal input features were selected and fed to SVM. Improved classification accuracy had obtained due to reasonable choice of sensitive features selection methods (illustrated in Table.5 and Fig.8 (a) & (b)). The final experimental results indicate that, relative reduct rough set hybrid with PSO algorithm is feasible to optimize the input features for SVM and effective with high classification accuracy and consumed less computing time.

Table. 5. Performance comparison of classifiers with and without feature selection

| Classifier | No. of input features | Average Accuracy on testing data (%) | Average computing time for training (sec) |
|---|---|---|---|
| SVM(no feature reduction) | 8(All) | 97.5 | 0.789 |
| SVM( feature reduction with RR) | 5(1,2,5,6,8) | 97 | 0.568 |
| SVM( feature reduction with RR based PSO) | 4(1,3,5,8) | 98.75 | 0.397 |

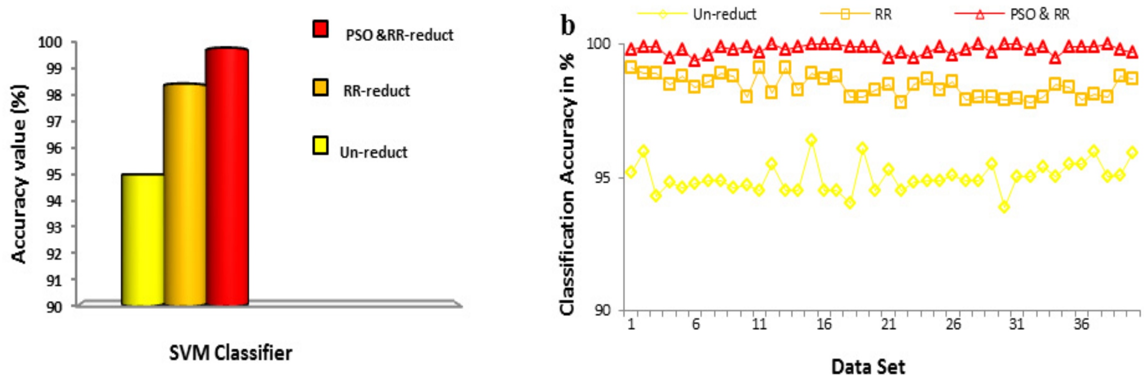



Fig .8 (a) Average classification accuracy; (b) Classification accuracy value (%) of various data sets.

## References

[1] K. Mori, N. Kasashima, T. Yoshioka, Y. Ueno - Prediction of spalling on a ball bearing by applying the discrete wavelet transform to vibration signals –J. Wear 195 (1996) 162-168.
[2] P. D. Mcfadden and J. D. Smith, Model for the vibration produced by a single point defect in a rolling element bearing, Journal of Sound and Vibration, 96 (1984) 69-82.
[3] J.K. Halme - Condition Monitoring of Oil Lubricated Ball Bearing Using Wear Debris and Vibration Analysis - 6th International Tribology Conference, Austrib 02. 2-5( 2002)
[4]O. R. Seryasat,M. Aliyari Shoorehdeli, F.Honarvar, and A. Rahmani,"Multi-fault diagnosis of ball bearing based on features extracted from time-domain and multi-class support vector machine (MSVM)," in Proc. of the IEEE International , The Scientific World Journal Conference on Systems, Man and Cybernetics (SMC '10), pp.4300–4303, October 2010.
[5] Z. L. Liu, Q. Z. Lu,Y. L.Wang, andC.Y.Wei, "Adirect selection method of feature frequency," in Emerging Research in Artificial Intelligence and Computational Intelligence—International Conference,vol. 315, pp. 479–486, 2012.
[6 ]Yang, J., Zhang, Y., Zhu, Y., 2007. Intelligent fault diagnosis of rolling element bearing based on SVMs and fractal dimension. Mech. Syst. Signal Process. 21 (5), 2012–2024.
[7] P.W. Tse,W.-X. Yang, and H. Y. Tam, "Machine fault diagnosis through an effective exact wavelet analysis," Journal of Sound and Vibration, vol. 277, no. 4-5, pp. 1005–1024, 2004.
[8 ] [Z. K. Peng and F. L. Chu, Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography, Mech. Systems And Signal Processing, 11 (5) (2004) 751-765.
[9 ] B. Samanta and K. R. Al-Balushi, "Artificial neural network based fault diagnostics of rolling element bearings using time domain features,"Mechanical Systems and Signal Processing, vol.17, no. 2, pp. 317–328, 2003.
[10] Zhiwen Liu,HongruiCao n, XuefengChen, ZhengjiaHe,ZhongjieShen, Multi-fault classification based on wavelet SVM with PSO algorithm to analyze vibration signals from rolling element bearings,J. Neurocomputing 99 (2013) 399–410.
[11] Z. K. Peng and F. L. Chu, Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography, Mechanical Systems And Signal Processing, 11 (5) (2004) 751-765.
[12] J. Lin, M. J. Zuo, K. R. Fyfe, Mechanical Fault Detection Based on the Wavelet De-noising Technique, J. of Vib and Acou.26(1): 9-16, 2004.
[13] B. Liu, Selection of wavelet packet basis for rotating machinery fault diagnosis, Journal of Sound and Vibration, 284(3-5): 567-582, 2005.
[14] Wang pan-pan, Shi li-ping, Hu yong-jun, Miao chang-xin, Feature extraction of induction motor stator fault based on particle Swarm optimization and wavelet packet,, J. of coal Sci & Eng(china) doi 10.1007/s12404-012-0418-z,pp 432–437 vol.18 no.4 dec. 2012.
[15] Xiaoran Zhu*, Youyun Zhang and Yongsheng Zhu, Intelligent fault diagnosis of rolling bearing based on kernel neighborhood rough sets and statistical features, Journal of Mechanical Science and Technology 26 (9) (2012) 2649~2657
[16] Bottou, L , Cortes, C , Denker, J , Drucker, H ,Guyon, I, Jackel, L, LeCun, Y , Muller, U ,Sackmger, E., Simard, P and Vapnik, V, 1994, "Comparison of Classifier Methods A Case Study in Handwriting Digit Recognition," Proc. international Conference on Pattern Recognition, pp 77-87.
[ 17] Krepel, U, 1999, "Pairwise Classification and Support Vector Machines," in Advances in Kernel Methods-Support Vector Learning, B Scholkopf, C, J, C, Burges, A J Smola, Eds MIT,Press, Cambridge, pp 255 — 268.
[ 18] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines, " IEEE Transactions on Neural Networks, vol. 13, no. 2, pp. 415–425, 2002.
[19] Samanta, B , 2004, "Gear Fault Detection using Artificial Neural Networks and Support Vector Machines with Genetic Algorithms," *Mechanical Systems and Signal Processing,* Vol 18, No 3,pp 625-644.
[20]Y. Shi, R.C. Eberhart, A modified particle swarm optimizer,in: Proceedings of the IEEE International Conference on Evolutionary Computation Anchorage, 4–9, May, AK, USA,1998, pp. 69–73.
[ 21 ]J. Han, X. Hu, T.Y. Lin, Feature subset selection based on relative dependency between attributes, rough sets and current trends in computing, Lecture Notes in ComputerScience 3066 (2004) 176–185.
[22] H. Hannah Inbarani, Ahmad Taher Azarb,, G. Jothi, Supervised hybrid feature selection based on PSO and rough sets for medical diagnosis, J. of .comp methods and programs in bio medicine 1 1 3 ( 2 0 1 4 ) 175–185