International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 4 (2015) pp. 8725-8740

http://www.ripublication.com

# An Intelligent Gear Fault Diagnosis Based On Wavelet Packet Transform, Information Gain And Multiclass Least Squares Support Vector Machines

**C.Rajeswari[1], B.Sathiyabhama[2], S.Devendiran[3]& K.Manivannan[4]**

[1] *Research Scholar, Faculty: Information and Communication Engineering, Anna University, Chennai, India*
*raji.chen@gmail.com*
[2]*Department of Computer Science and Engineering, Sona College of Technology, Salem, Tamil Nadu, India*
*sathiya674@yahoo.co.in*
[3,4]*School of Mechanical and Building Sciences, VIT University, Vellore, Tamil Nadu, India*
*devendiran@vit.ac.in*
*manivannan.k@vit.ac.in*

## Abstract

This paper presents annew approach to intelligent gear fault diagnosis in the field of rotating machinery condition monitoring.The prediction process is focused on the vibration signals collected from the accelerometer sensor mounted on the test rig for critical components monitoring. The pre-processed signals were decomposed into several signals containing one approximation and some details using Wavelet Packet Decomposition and thestatistical features were extracted from 4th level WP decomposition. The J48 algorithm which uses gain ratio to determine the splits of decision tree and to select the most important features were used for selecting the predominant features and selected features were fed as input for training and testing of Multi class Least Squares Support Vector Machines (MCLS-SVM) together with different kernel functions such as radial basis function kernel, linear-kernel, Polynomial-kernel functions for classification and predict the fault condition of the components and machines.Analysis is done for various LS-SVM models such as One against One and One against All. The proposed approach is applied to fault diagnosis of gear, and testing resultsshows the proposed approach can reliably recognise different fault categories and severities.

**Keywords:**Wavelet packet decomposition , Gain ratio,Least Squares Support Vector Machines,radial basis function .

## 1. Introduction

Vibration based diagnosis method for machine component fault identification have been in existence over a period of approximately 35 years. The developments in electronic data acquisition equipment, sensors, computers and software made an automated supervision system in condition monitoring of machines. Particularly, gears play an important role in a many of industrial rotating and transport machinery applications. Early fault diagnosis of the gear may prevent unnecessary failures of whole rotating machinery system .Therefore many researchers strives towards the extensive investigations on this research using vibration signals based on time domain, frequency domain, and time–frequency domain of techniques [1-3].The traditional signal processing techniques such as Fourier transform and fast Fourier transform can convert the signal into frequency domain, but they cannot provide both time information and frequency information simultaneously [4]. In addition, the gear vibration signals frequently show non-stationary characteristics, whereas the Fourier transform is developed on a global concept and stationary assumption [5]. Therefore, the Fourier transform can’t fulfil the gear fault diagnosis task well.The vibration measurement of the gear can be made using some accelerate sensors that are placed on the gearbox. When vibration features of gears are obtained, its health condition can be determinedby comparing these patterns with those corresponding to its normal and failure conditions.There are many vibration-based diagnosis techniques currently available for the detection of gear faults such as frequency/cepstrum analysis, time/statistical analysis, and time–frequency analysis,among which time–frequency analysis has become the well-accepted techniques. It has beenapplied to the gear fault diagnosis widely because the time–frequency method can provide both the time and frequency information of the vibration signal. However, the common time–frequency methods such as windowed Fourier transform (WFT) and wavelet transform have its limitations, respectively. WFT can display a time signal on a joint time–frequency plane, but once the window function is chosen, its size of the time–frequency window would be fixed, therefore the time and frequency resolution are same for all components that include different time scales [6,7]. Hilbert–Huang transform includes empirical mode decomposition (EMD). EMDmethod is based on the local characteristic time scale of signal and could decompose the complicated signal into a number of intrinsic mode functions (IMFs). By analysing each resulting IMF component that involves the local characteristic of the signal, the characteristic information of the original signal could be extracted more accurately and effectively. In addition, EMD can be regard as a special filter whose bandwidth and central frequency change with the signal itself, therefore, EMDis a self-adaptive signal processing method that can be applied to non-linear and non-stationary process perfectly [8].After signal processing stage, statistical features of acquired vibration signal were calculated for different running conditions of rotating machinery. After extracting statistical features these features are input to sophisticated artificial intelligence algorithms for classification through machine learning tools (such as the artificial neural network (ANN), support vector machines (SVMs), and fuzzy logic systems (FLS)) for its meaningful diagnostics of conditions of faults [9].Many of the times vast number of features within varietyof domains poses challenges to classification task. Main objective of feature selection is reducing computation time by reduce the dimensionality of input features for the purpose of efficient diagnosis results. Principal component analysis (PCA) [10] and Genetic algorithm (GA) [11] were widely used in this field to decrease the relativity between features and decrease dimensions of features.

In Section 1 we briefly reviewed the works which are strictly connected to the subject of this paper. In Section 2 contains theoretical background of wavelet packet transform, statistical features and feature reduction methods were discussed. In section 3 Gain ratio based feature selection procedure is discussed, In Section 4 Theoretical background of MCLS-SVM is discussed .In Section 5 Experimental setup and experimental procedure is presented. In Section 6 Result analysis of simulated data according to the presented procedure is presented. In Section 7 contains conclusions. The methodology of proposed work is shown in the Fig.1.

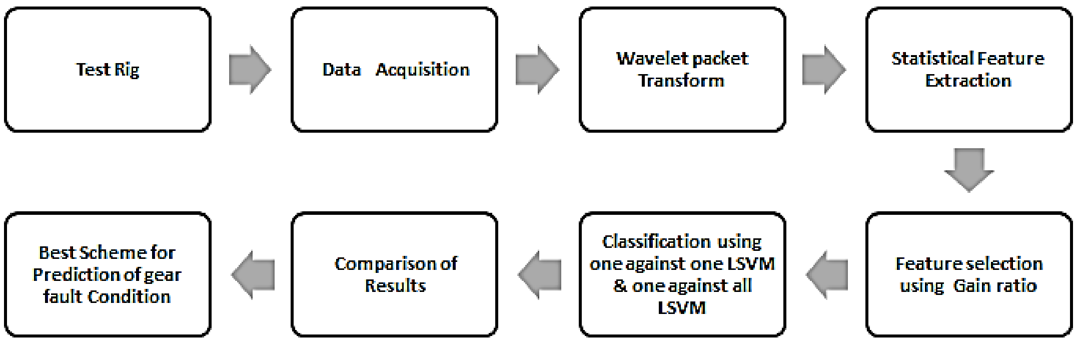


**Fig.1 Methodology**

## 2. Experimental setup and experimental procedure

Experiment process of gear fault diagnosis is made on the well designed and fabricated test rig (Fig. 1(a)). Experimental set comprise of AC motor, variable frequency drive (VFD) as a motor speed regulator, tri axial accelerometer as a vibration measurement sensor, data acquisition system (DAS-ATA0824DAQ51) integrated withcomputer and the brake drum dynamometer used to apply a load. The frequency range of gear fault is mostly identified at middle-frequency band [5$f_0$, 1000

Hz] were $f_0$ is rotating frequency[7]. Therefore the sample frequency is set to 6400 Hz. The number of sampling rate is 12400 /sec. For each gear condition 25 groups of sample data were acquired.

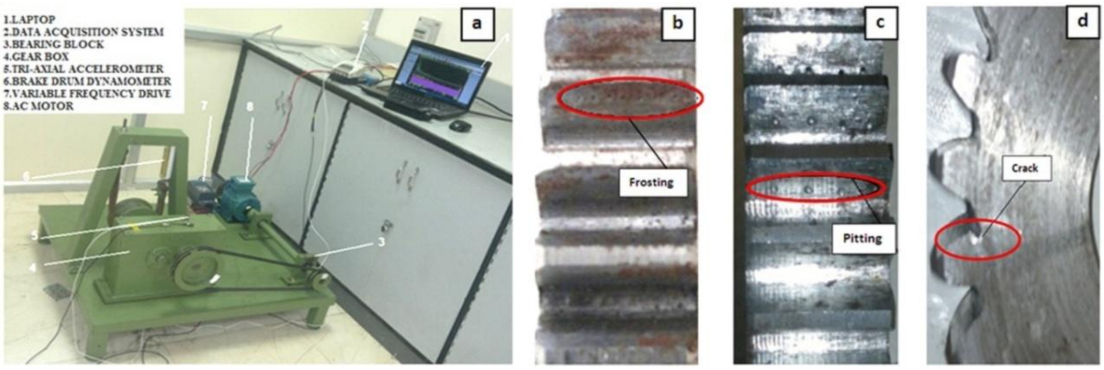


**Fig.1. (a) Gear test rig (experimental set up) (b) frosting induced gear (c) pitting induced gear (d) crack induced gear**

Using accelerometer sensor,vibration signals were collected for 4 gear conditions. Four gear conditions such as Normal, frosting( Fig 1(b)), pitting (Fig.1(c)) and crack (Fig.1(d)) were artificially made using electrical discharge machining (EDM) process and used for the experimental process separately.The experiment was carried out in constant speed of 1200 rpm and a constant load conditions. The time domain waveforms and spectrums of vibration signal for four conditions of gear are shown in Fig.2.

**Table:1 Statistical feature description**

| Statistical feature | Equation | Definition |
|---|---|---|
| Standard deviation | $k_{Sd} = \sqrt{\left(\frac{1}{n-1}\sum_{i=1}^{n}(k_i-\mu)^2\right)}$ | Square root of an unbiased estimator of the variance of the population |
| Kurtosis | $k_{kur} = \frac{1}{n}\sum_{i=1}^{n}\left(k(t)-\bar{k}\right)^4$ | Fourth central moment of 'X', divided by fourth power of its standard deviation. |
| Root mean square | $k_{rms} = \sqrt{\left(\frac{1}{n}\sum_{i=1}^{n}k^2\right)}$ | Root of sum of squared values |

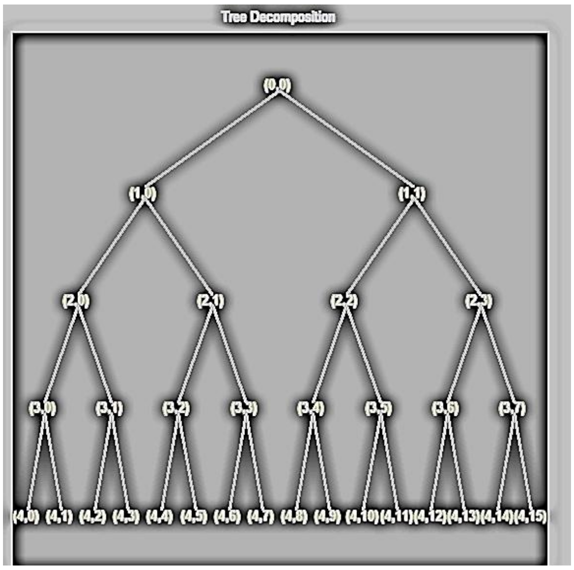


**Fig.2 Four level WPD**

## 3. Statistical Feature Extraction using WPT

In WPT wavelet packets contain a complete set of decompositions and details at every level which provides high resolution in high frequency region .The WPD is broke up to 4 resolution levels is shown in Fig.2. In the figure, for example node (4,1) 4 represents the 4th resolution and 0th subspace. Hence, the orthogonal wavelet package transform is able to produce multi-layer division of the frequency band, and analyse the high frequency part which is not fractionized in wavelet transform. A three level WPT produces a total of 16 sub bands with each sub band covering one-sixteen of the signal frequency spectrum. Top level of the wavelet packet transform is the time representation of the signal whereas bottom level has better frequency resolution which also provides more features about a signal as compared to wavelet transforms [12]. Hence, the orthogonal wavelet package transform is able to produce multi-layer division of the frequency band, and analyse the high frequency part which is not fractionized in wavelet transform. A three level WPT produces a total of 16 sub bands with each sub band covering one-sixteen of the signal frequency spectrum. Top level of the wavelet packet transform is the time representation of the signal whereas bottom level has better frequency resolution which also provides more features about a signal as compared to wavelet transform. In general many wavelet functions are available but smooth wavelet and compact wavelet such as Daubechies 4 wavelet is arbitrarily chosen [13] for this work to analyse the non-stationary data. There is no logic behind the selection of Daubechies order [14].

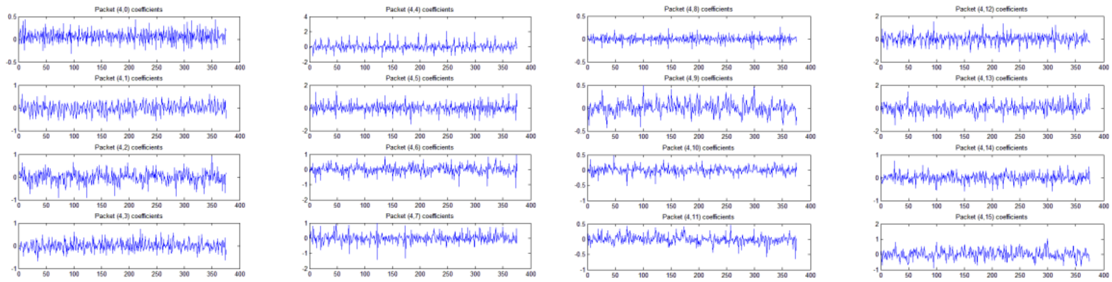

**Fig.3(a).4$^{th}$ level WP decomposition of crack gear signal**

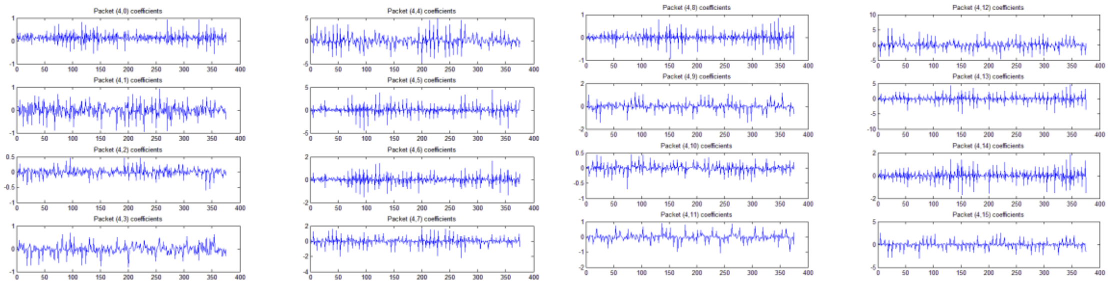

**Fig.3(b).4$^{th}$ level WP decomposition of pitting gear signal**

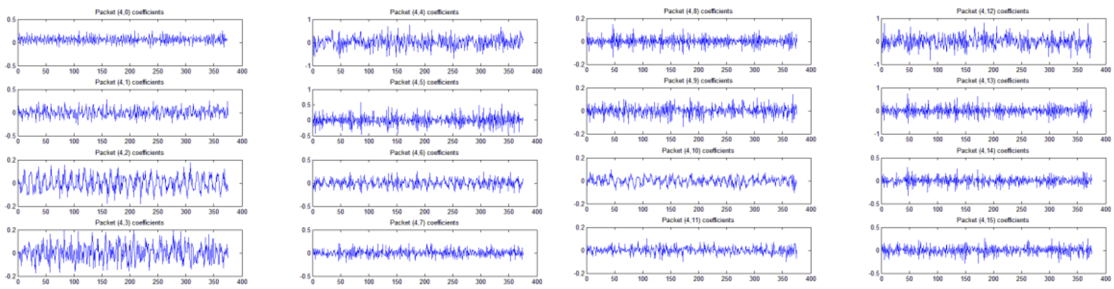

**Fig.3(c). 4$^{th}$ level WP decomposition of frosting gear signal**

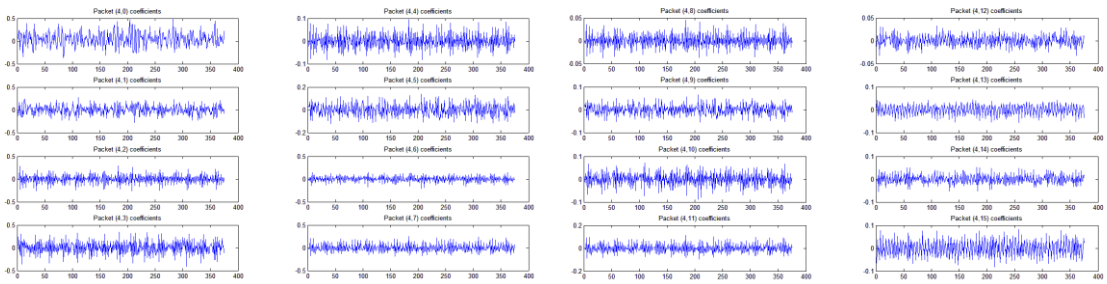

**Fig.3(d).4$^{th}$ level WP decomposition of normal gear signal**

For each state of signal WPT is applied for 4$^{th}$ level and 16 WP coefficients were extracted. The most able statistical features were calculated using the formulas listed in Table.1 for each WP coefficient. This feature vectors are extracted and listed separately for training and testing of proposed scheme in 60% and 40 % respectively for the acquired samples. The decomposition using WPT for 4 states of gear conditions are depicted in Fig.3(a)to Fig.3(d).

## 4. Gain Ratio based feature selection

The most prominent drawback of information gain measure is that this methodology works towards tests with many outcomes. That is, it prefers to select attributes having a large number of possible values over features with fewer values even though the latter is more informative. For example consider an attribute that acts as a unique identifier, such as an employee Number in the employee database. A split on this feature would result in a large number of partitions; as each record in the database has a unique value for employee number. So the information required to classify database with this partitioning would be 0. Clearly, such a partition is useless for classification.

ID3 is the basic decision tree algorithm [15]. C4.5 [16], a successor of ID3, uses an extension to information gain known as gain ratio (GR) [17], which attempts to overcome the bias. Let D be a set consisting of d data samples with n distinct classes. The expected information needed to classify a given sample is given by

$$I(D) = -\sum_{i=1}^{n} p_i \log_2(p_i) \quad (1)$$

Where$p_i$ is the probability that an arbitrary sample belongs to class $C_i$. Let attribute A have v distinct values. Let $d_{ij}$be number of samples of class $C_i$ in a subset$D_j$. $D_j$Contains those samples in D that have value $a_j$ofA. The entropy based on partitioning into subsets by A A, is given by

$$E(A) = -\sum_{i=1}^{n} I(D) \frac{(d_{1i} + d_{2i} + \cdots d_{mi})}{d} \qquad (2)$$

The encoding information that would be gained by branching on A is

$$\text{Gain } (A) = I(D) - E(A) \qquad (3)$$

C4.5 applies a kind of normalization to information gain using a "split information" value defined analogously with Info (D) as

$$\text{Split Info}_A(D) - \sum_{j=i}^{v} \frac{D_j}{|D|} \log_2 \left(\frac{|D_j|}{|D|}\right) \qquad (4)$$

This value represents the information computed by splitting the dataset D, into partitions, corresponding to the v outcomes of a test on feature A. For each possible outcome, it considers the number of tuples having that outcome with respect to the total number of tuples in D. The gain ratio is defined as

$$Gain\ Ratio(A) = \frac{\text{Gain (A)}}{\text{SpiltInfo(A)}} \qquad (5)$$

The attribute with maximum gain ratio is selected as the splitting attribute.

## 5. Theoretical background of Multiclass Least Square Support Vector Machines

Multiclass Least Square Support Vector Machines (LS-SVM) are a group of supervised learning methods that can be used for classification of two class label dataset. For multiclass case the input training data $\{x_i\}_{i=1}^{i=P}$ , $\{y_i^k\}_{i=1,k=1}^{i=P,k=m}$ is the output training data of the $k^{th}$ output unit for pattern $i$. The derivation of the MC-LS-SVM [18]can be expressed asfollows :

$$\text{Minimize } J_{LS}^{(m)}\left(w_{k,} b_k l_{i,k}\right) = \frac{1}{2} \sum_{k=1}^{m} w_k^T\, w_k + \frac{\gamma}{2} \sum_{i=1}^{P} \sum_{k=1}^{m} e_{i,k}^2 \qquad (6)$$

The applied equality constraints are:

$$\begin{cases} y_i^1[w_1^T g_1(x_i) + b_1] = 1 - e_{i,1}, i = 1,2,\dots P \\ \vdots \\ y_i^m[w_m^T g_m(x_i) + b_m] = 1 - e_{i,m}, i = 1,2,\dots P \end{cases} \quad (7)$$

Where P denotes the number of training dataset. The $e_{i,m}$ and $b_k$ denotes the classification error and the bias, respectively. $w_k$and$\gamma$ are the weight vector of $k^{th}$ classification error and the regularization factor, respectively.

The $g_k(.)$is a nonlinear function and it maps the input space into a higher dimensional space. The Lagrangian multipliers $\alpha_{i,k}$ can be defined as:

$$L^{(m)}(w_k, b_k, e_{i,k}; a_{i,k}) = J_{LS}^{(m)} - \sum_{i,k} a_{i,k}\{y_i^k[w_k^T g_k(x_i) + b_k] - 1 + e_{i,k}\} \quad (8)$$

Which provides the following conditions for optimality

$$\begin{cases} \frac{\partial L}{\partial w_k} = 0 \rightarrow w_k = \sum_{i=1}^{P} \alpha_{i,k}\, y_i^{(k)} g_k(x_i) \\ \frac{\partial L}{\partial b_k} = 0 \rightarrow \sum_{i=1}^{P} \alpha_{i,k}\, y_i^{(k)} = 0 \\ \frac{\partial L}{\partial l_{i,k}} = 0 \rightarrow \alpha_{i,k} = \gamma e_{i,k} \\ \frac{\partial L}{\partial \alpha_{i,k}} = 0 \rightarrow y_i^{(k)}[w_k^T g_k(x_i) + b_k] = 1 - e_{i,k} \end{cases} \quad (9)$$

Where $i$=1,2,..,P and $k$=1,2,... $m$. Elimination of $w_i$ and $e_{i,k}$ provides the linear system as:

$$\begin{bmatrix} 0 & Y_M^T \\ Y_M & \Omega_M \end{bmatrix} \begin{bmatrix} b_M \\ \alpha_M \end{bmatrix} = \begin{bmatrix} 0 \\ \bar{1} \end{bmatrix}$$

With the following matrices:

$$Y_M = blockdiag\left\{ \begin{bmatrix} y_1^{(1)} \\ \cdot \\ \cdot \\ \cdot \\ y_P^{(1)} \end{bmatrix}, \dots, \begin{bmatrix} y_1^{(m)} \\ \cdot \\ \cdot \\ \cdot \\ y_P^{(m)} \end{bmatrix} \right\}$$

$$\Omega_M = blockdiag\{\Omega_1, \dots \Omega_M\}$$

$$\Omega_i^k = y_i^k y^k g_k^T(x) g_k(x_i) + \gamma^{-1} I$$

$$\bar{1} = [1, \dots 1]$$

$$b_M = [b_1, .. b_m]$$

$$\alpha_{i,k} = [\alpha_{1,1}, \dots, \alpha_{P,1}; \dots; \alpha_{1,m}, \dots, \alpha_{P,m}]$$

Where $K_k(x, x_i) = g_k^T(x) g_k(x_i)$ is kernel function, which satisfies Mercer conditionthe decision function of MC-LS-SVM is expressed as:

$$f(x) = sign\left[\sum_{i=1}^{P} \alpha_{ik} y_i^{(k)} K_k(x, x_i) + b_k\right] \quad (10)$$

The radial basis function (RBF) kernel for MC-LS-SVM can be expressed as:

$$K_k(x, y) = exp\left[\frac{-\|x-y\|^2}{2\sigma_k^2}\right] \quad (11)$$

Where$\sigma_k$ controls the width of RBF function.

The linear kernel function for MC-LS-SVM can be expressed as:

$$K_k(x, y) = x.y \quad (12)$$

Similarly, the polynomial kernel function for MC-LS-SVM can be expressed as:

$$K_k(x, y) = (x.y + 1)^d \quad (13)$$

Where $d$ is the degree of the polynomial kernel parameter.

## 6. Experimental analysis,results and discussion

The test results of proposed fault diagnosis system on the gearbox health condition identification are demonstrated in this section. In general non stationary fluctuations will be shown in vibration signals emitted from faulty gear and bearing components .The Wavelet transform(WT) offers allowing a fine time resolution in higher frequency domain while a fine frequency resolution in lower frequency range. Consequently, WT has gained an immense popularity in mechanical vibration analysis applications [19]. In the proposed diagnosis system, 16 node components were obtained first by implementing a four-level WPT, as shown in Fig. 1.Once getting the 16 node components, i.e., nodes (4, 0) (4, 1) ….. (4, 15), three statistical features were calculated for each of the 16 nodes, thus resulting in a total of 48 features.The extracted features analysed using different schemes .The schemes are formed based on the combination of different types of MCLS-SVM, Kernel functions and feature selection process.Waikato Environment for Knowledge Analysis (WEKA) is a comprehensive suite of Java class libraries that implement many state-of-the-art machine learning and data mining algorithms. The experimental analysis was carried out with and without feature selection schemes. As per the diagnosis methodology, features are selected and were performed using open source software WEKA [20, 21]. WEKA package which has its own version of C4.5 known as J4.8 to identify the significant features and its process time selected features are depicted in Table.2.Multiclass classification is solved using based on one of the voting schemes, which are based on combining binary classification decision functions Various approaches, such as one-against all (OAA)[22] ,one-against-one(OAO) [23].There are different kernel functions used in MCLS-SVM, such as linear, polynomial, chi-square and RBF, which avoid the computational burden of explicitly representing the feature vectors. The selection of an appropriate kernel function is important, since the kernel function defines the feature space in which the training set data will be classified. The kernel expresses prior knowledge about the phenomenon being modelled and encoded as a similarity measure between the vectors. In this work, linear, polynomial and RBF kernel functions were evaluated. The choice of kernel function is data dependent and there are no definite rules governing its choice that might yield a satisfactory performance. In this present work $d$ is the degree of the polynomial kernel parameter and the width of the RBF kernel parameter is given by $\sigma$ and both parameters can be determined in general by an iterative process selecting an optimum value based on the full feature set [24]. In this present work $\sigma$ =0.9 and $d=2$were taken after many iterations based on higher classification accuracy. These kernels are also well accepted for constructing SVM and provide excellent results for real-world applications [25]. Investigation is done on MCLS-SVM classifier using the one-against-one method and the one against-all method.The training and simulation of MCLS-SVMwere performed by Matlab Platform. Before the tests, the features extracted from the gear vibration signals were normalized into the range [-1, 1].Results are analysed based on the most important criterion for evaluating the performance of these methods is their classification success rate.The results in Table 3 show that the performance of one against- all scheme of classifiers is better than that of one against- one classifier schemes from the view of classification accuracy (depicted in Fig.4 ) and training time (depicted in Fig.5 ).Among these classifiers, polynomial kernel function gives the best with high training and testing accuracy.

**Table :2 Feature selection process results**

| **Sl. No** | **Feature selection scheme** | **No. of selected features** | **Feature description** | **Elapsed time(sec)** |
|---|---|---|---|---|
| 1. | No feature selection | 48 | F1-F48 | 0 |
| 2. | Gain ratio based J48 | 19 | F1,F1,F2,F5,F7,F8,F11,F12,F16,F17,F18,F21,F24,F27,F28,F29, F31,F34,F40,F45. | 0.0198 |

**Table : 3 Classification process results for different Schemes.**

| Classifier | Kernel Function (%) | Scheme | Frosting fault (%) | Pitting fault (%) | Crack fault (%) | Normal (%) | Total Testing Accuracy (%) | Training Time (sec ) |
|---|---|---|---|---|---|---|---|---|
| OAO MCLS-SVM (without feature selection) | RBF | Scheme-1 | 83.45 | 76.88 | 80.25 | 87.67 | 82.06 | 0.174 |
| | linear | Scheme-2 | 72.87 | 78.9 | 70.56 | 76.78 | 74.78 | 0.118 |
| | Polynomial | Scheme-3 | 80.67 | 89.73 | 85.98 | 88.68 | 86.27 | 0.155 |
| OAO MCLS-SVM (Gain ratio based feature selection) | RBF | Scheme-7 | 94.88 | 85.76 | 89.52 | 97.8 | 91.54 | 0.189 |
| | linear | Scheme-8 | 90.55 | 98.33 | 99.72 | 91.11 | 92.42 | 0.085 |
| | Polynomial | Scheme-9 | 95.95 | 99.19 | 92.66 | 90.3 | 94.52 | 0.191 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| OAA MCLS-SVM (without feature selection) | RBF | Scheme-4 | 90.83 | 85.83 | 91.67 | 86.67 | 88.75 | 0.170 |
| | linear | Scheme-5 | 83.24 | 75.23 | 78.52 | 85.79 | 80.30 | 0.126 |
| | Polynomial | Scheme-6 | 92.34 | 99.16 | 99.73 | 96.76 | 94.50 | 0.112 |
| OAA MCLS-SVM (Gain ratio based feature selection) | RBF | Scheme-10 | 89.75 | 97.18 | 86.91 | 94.57 | 92.10 | 0.175 |
| | linear | Scheme-11 | 95 | 95.83 | 93.06 | 97.5 | 95.35 | 0.098 |
| | Polynomial | Scheme-12 | 94.81 | 91.83 | 92.12 | 99.36 | 97.03 | 0.157 |

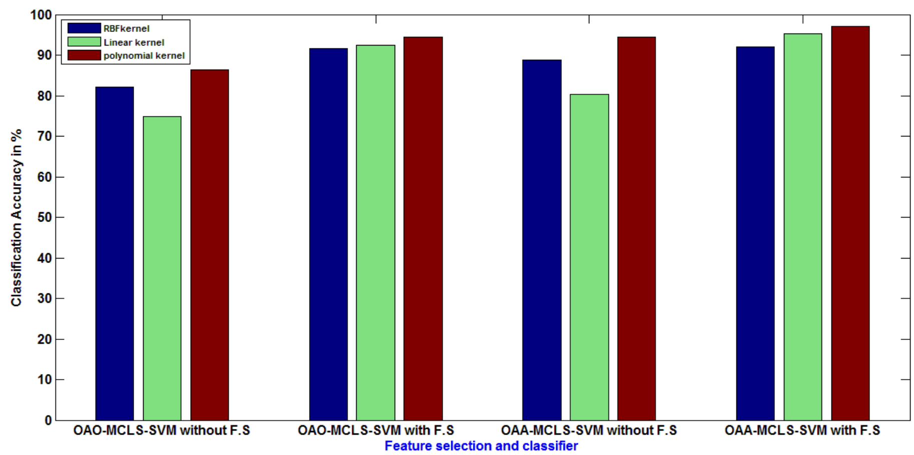


**Fig.4. Classification accuracy Vs different schemes**

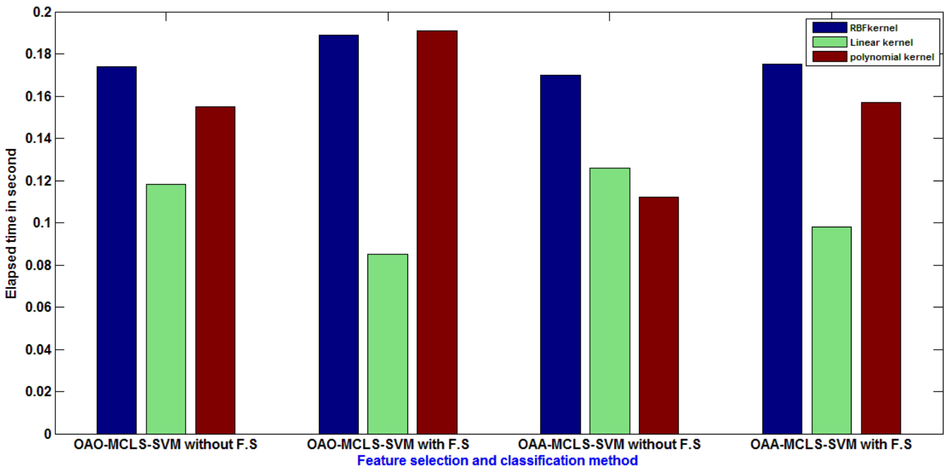


**Fig.5. Elapsed training time Vs different schemes**

## 7. Conclusion

The experiments have demonstrated that the proposed MCLS-SVMs based gear fault diagnosis approach for rotating machinery. Although the mechanical behaviour of the each fault results was complete and non-stationary, the WPT has been demonstrated to be a useful signal processor to extract different time-frequency statistical features of symptoms of various fault conditions. It is shown that using only the predominant features for classification can obtain high success rate and reduces the training time of the MCLS-SVMs classifier. The effect of kernel functions in SVM classifications is also discussed. Finally one-against-all MC MCLS-SVM classifier using a polynomial kernel schemeshows superior performance of 97.03 % comparison with other 11 schemes. MCLS-SVMs hold significant

Promise in the diagnosis of rotating machinery due to their ability to give optimal performance.

## References

[1] J.I. Taylor, (1995), “Back to the basics of rotating machinery vibration analysis”, Sound Vibration 29 (2) 12–16.

[2] H.Endo,R.B.Randall,(2007), “Enhancementofautoregressivemodelbasedgeartoothfaultdetectiontechniqueby theuseofminimumentropy deconvolutionfilter”, Mechanical SystemsandSignalProcessing 21,906–919.

[3] P.D. McFadden,(2000),“Detection of gear faults by decomposition of matched difference of vibration signals”, Mechanical Systems and SignalProcessing 14 (5) ,805–817.

[4] V.K. Rai, A.R. Mohanty, (2007) “Bearing fault diagnosis using FFT of intrinsic mode functions in Hilbert–Huang transform”, Mechanical Systems and Signal Processing 21 (2),2607–2615.

[5] D.J. Yu, J.S. Cheng, Y. Yang, (2005), “Application of EMD method and Hilbert spectrum to the fault diagnosis of roller bearings”, Mechanical Systems and Signal Processing 19 (2) 259–270.

[6] S. Qian, D. Chen, 1996.“Joint Time–frequency Analysis”, Prentice-Hall, Englewood Cliffs, NJ,

[7] H. Oehlmann, D. Brie, M. Tomczak , et al., (1997) “A method for analyzing gearbox faults using time–frequency representations”, Mechanical Systems and Signal Processing 11 (4) 529–545.

[8] N.E. Huang, Z. Shen, S.R. Long, (1999), “A new view of nonlinear water waves:the Hilbert spectrum”, Annual Review of Fluid Mechanics 31 ,417–457.

[9] C.J.C. Burgess, (1998) ,“A tutorial on support vector machines for pattern recognition”, Data Min. Knowl. Disc. 2 955–974.

[10] Ngoc-Tu Nguyen, Hong-Hee Lee and Jeong-Min Kwon, (2008) , “Optimal feature selection using genetic algorithm for mechanical fault detection of induction motor” Journal of Mechanical Science and Technology 22 ,490~496.

[11] V. Sugumaran, V. Muralidharan and K. I. Ramachandran, (2007), “ Feature selection using Decision Tree and classification through Proximal Support Vector Machine for fault diagnostics of roller bearing”, Mech. Systems and Signal Processing 21 ,930-942.

[12] Mallat S.(1999) A wavelet tour of signal processing. 2nd ed.New York: AcademicPress;

[13] StaszewskiWJ,Giacomin J(1997) “Applications of wavelet based FRFs to the analysis of non-stationry vehicle data”, Proc. Int. Modal Analysis Conference 1:425-431

[14] VafeaiS,Raahnejat H(2003) “Indicated repeatable runout with wavelet decomposition (IRR-WD) for effective determination of bearing induced vibration” . Journal of sound and vibration 260(1):67-82.

[15] J.R. Quinlan, Induction of Decision Trees (1986), Machine Learning 1: pp.81-106, Kluwer Academic Publishers, Boston,

[16] J.R. Quinlan, San Mateo, (1993), C4.5 Programs for Machine Learning: Morgan KaufmANN,

[17] J. Han And M. Kamber, “Data Mining: Concepts and Techniques”. San Francisco, Morgan KauffmannPublishers(2001).

[18] J.A.K. Suykens and J. Vandewalle,( 1999), “Multiclass least squares support vector machines,” in Proc. of the International Joint Conference onNeural Networks, Washington DC, USA, pp. 900-903.

[19] Z. K. Peng and F. L. Chu, (2004) “Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography”, Mechanical Systems And Signal Processing, 11 (5) ,751-765.

[20] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann and Ian H. Witten,( 2009) “The WEKA Data Mining Software: An Update”, ACM SIGKDD Explorations Newsletter, Vol.11, No.1, pp. 10-18,

[21] http://www.cs.waikato.ac.nz/~ml/weka/.

[22] Hsu, C W and Lin, C J (2002), "A Comparison of Methods for Multiclass Support Vector Machines," IEEE Trans, on Neural Networks,Vol 13, No 2, pp 415-425

[23] Krepel, U, (1999), "Pairwise Classification and Support Vector Machines," in Advances inKernel Methods-Support Vector Learning, B Scholkopf, C, J, C, Burges, A J Smola, Eds MIT Press, Cambridge, pp 255 — 268.

[24] Scholkopf, B, (1997), Support Vector Learning, Oldenbourg-Verlag, Germany.

[25] Strauss, D, J and Steidel, G, 2002, "Hybrid Wavelet-Support Vector Classification of Waveforms,"Journal of Computational and AppliedMathematics, Vol. 148, pp 375—400