Fault Classification of Low Speed Bearings based on Support Vector Machine for Regression and Genetic Algorithms using Acoustic Emission

Henry Ogbemudia Omoregbee[1] and P. Stephan Heyns[2]

[1,2]Centre for Asset Integrity Management
Department of Mechanical and Aeronautical Engineering
University of Pretoria, Pretoria, South Africa.
[1]omoregbeeho@gmail.com, [2]Stephan.Heyns@up.ac.za

Abstract

This paper explores the use of support vector machines (SVM) for regression and genetic algorithm (GA) which may be referred to as SVMGA, to classify faults in low speed bearings over a specified speed range, with sinusoidal loads applied to the bearing along the radial and axial directions. GA is used as a heuristic tool to find an approximate solution to the difficult problem of solving the highly non-linear situation through the application of the principles of evolution by optimizing the statistical features selected for the SVM for regression training solution. It is used to determine the training parameters of SVM for regression which can optimize the model and thereby generate new features from the original dataset without prior knowledge of the probabilistic distribution. The fault recognition and the nonlinear regression is achieved by using SVM for regression. Classification is performed for three classes. In this work the GA is used to first optimize the statistical features for best performance before they are used to train the SVM for regression. Experimental studies using acoustic emission caused by bearing faults showed that SVMGA with a Gaussian kernel function better achieves classification on the bearings operated at low speeds, regardless the load type and, under different fault conditions, compared to the exponential kernel function and the other many kernel functions which also can be used for the same conditions.

Keywords: Acoustic emission; Artificial neural networks (ANN); exponential kernel function; Gaussian kernel function; Rolling element bearing; Support vector machines (SVM); Genetic Algorithm (GA).

1 Introduction

The operational state and health of a machine is often determined through condition monitoring. Several factors may affect the operation of a machine and these may include machine speed, load, lubrication, alignment, etc. and often significantly affect the machine life. Researchers often distinguish different stages during the bearing life cycle, which are traditionally referred to as "infant mortality, useful life, and wear out", "good, damaged, failure imminent" or "pre-failure, failure, near catastrophic" [1]. Different vibration techniques have been developed to detect the faults in rolling element bearings (REB). This has been done with two main purposes: firstly, to separate the bearing related signal from other components and to minimize the noise that may mask the bearing signal. The widely used techniques for this purpose are self-adaptive noise cancellation (SANC), adaptive filtering, synchronous averaging and discrete random separation. The second purpose is to identify the status of the bearing, thereby distinguishing the good and the faulty bearings to indicate the defective component [1].

In an experiment conducted by [2], it was reported that parameters such as amplitude and energy provided information on the condition of a particular low speed rotating bearing. [2,3] reported the exploration of acoustic emission (AE) for the monitoring of rolling element bearings at speeds from 10 to 1850 rpm and it was concluded that at low speeds and steady loads, the base bending/strain of the bearing housing enables the AE transducer to detect the AE signature from very small defect.

An area that has gained much recognition in handling AE and vibration data in fault classification and recognition is Artificial Neural Networks, which has the capability of mimicking human experience gained in pattern recognition. Neural Networks may be designed to classify input patterns in predefined classes or to categorize the patterns by grouping them according to their similarities. They may also be designed to respond in real time to the changing system state descriptions provided by continuous sensor inputs. Artificial-intelligence-based fault diagnosis methods have the potential to tackle problems without human intervention, with a method which aims at recognizing different machine health conditions via the features extracted from the AE signals, with the accuracy of the identification of these conditions been further enhanced through classifiers that exhibit good performance [4, 13, 14].

Support vector machines (SVMs) on the other hand are based on statistical learning theory that is special for solving learning problems of smaller sample numbers that provide better generalization than ANN and guarantees the local and global optimal solutions to be exactly the same. SVMs are hence introduced into rotating machinery for fault diagnosis with high accuracy and good generalization for smaller sample numbers [5]. SVMs represent a machine learning approach that is widely used for data analysis and pattern recognition. The algorithm was developed by Vapnik and the current standard incarnation was proposed by Cortes and Vapnik. SVMs have well defined formulations which are consistent with mathematical theory.

As pointed out by [5,6], there are several challenges associated with the use of SVM, which is known for its time series predictions that span many practical application areas, among which is the free parameter selection. To date there is no universal method for hyper-parameter selection which contains some parameters such as the kernel parameter, the regularization parameter that controls the generalization performance of SVM and the ε-insensitive zone which determines the number of support vectors. Hence if one is using arbitrary SVM parameters, the performance of SVM could differ over a wide range. The original input space in SVM is mapped into high dimensional dot product space called feature space in which the optimal hyper-plane is determined to maximize the generalization ability of the classifier.

This paper introduces a support vector machine for regression [17] and genetic algorithm (SVMGA) based failure diagnostic method. The main contribution of the work lies in the utilization of a support vector machine for regression and genetic algorithm optimization that uses interactive procedures to simulate the process of evolution of possible solutions populations to the particular non-linear, stochastic problem nature of the slow speed bearing loaded at varying conditions. SVMGA is a technique of mono-class classification, which tries to solve the problem of best features selection by applying the principles of evolution by optimizing the statistical components selected for the SVM for regression training solution. By optimizing, it solves the problem of outlier detection and reduces the dimensionality of the data.

The choice of features can affect the performance of classification as features generated are often refined to try to achieve the desired level of performance. However, developing features manually can be time consuming; generated features should have the ability to identify complex relationship within large data dataset where the mapping from data to class labels is often obscure [22].

Many research works dealing with support vector machine and regression for fault diagnosis have been reported in the recent literature. Examples of these include fault diagnosis of a rolling bearing based on feature extraction and neural network algorithm [11], bearing fault diagnosis based on Kullback-Leibler transform and support vector machine [14], and an investigation for fault diagnosis based on a hybrid approach using wavelet packet and support vector classification [19]. However, none of these works dealt with the preferred selection process of features for best performance. SVMGA, which aims at enhancing bearing fault diagnostic is developed by the fusion of multiple statistical features through SVMGA, where a selection of best features from experimental data is done at the classifier training stage.

Unlike the target value of a SVM which can only be used to handle a binary problem, the target value of a SVMGA is continuous, having shown great potential in time series prediction and can also establish a stable nonlinear relationship between inputs and outputs. Outputs of small deviations from their target values are formed when the feature vectors extracted from the samples belonging to the same class are fed to the trained SVR [3, 15]. In an article by [8], a complex dynamic model for aligning roller bearings was established, thereby studying the problem of surface damage, preload and radial clearance. [8] worked on a physical model and one-class support vector machine for rolling bearing fault diagnosis.

SVM was originally designed for binary classification [9]. Some binary classification problems do not have a simple hyper-plane as a useful separating criterion. For SVM problems, there are variant of mathematical approaches that retain nearly all the simple separating hyper-plane such as “one-against one”, “one-against all” like in the classification of health states.

In this work, features were extracted from AE signals and used to train the SVM for fault classification of low speed bearings which were loaded sinusoidally along axial and radial directions, and grouped into ‘good bearing’ data signals, faulty ‘debris induced bearing’ signals and an ‘outer race crack bearing’ signal. The essence of this work is to show that regardless the load applied and speed especially at varying conditions like in rolling mills plants, fault classifications can still be achieved. SVMGA have distinct advantages over SVM and other AI algorithms like ANN. Because it is used for classifying faults which is over two classes that the SVM supports. Also, SVMGA uses high dimensional input space that does not necessarily depend on the number of features, because they have the potential to handle large feature spaces [11, 17]. This has been shown in applications such as in text categorization, biological studies, etc. This has however not been applied to low speed bearings which are sinusoidally loaded along axial and radial directions. This emulates real life scenarios where bearings in functional equipment may be loaded with varying cyclic loads at frequencies that might be unrelated to the equipment rotational frequency and its harmonics, due to auxiliary equipment (such as pump) mounted on the machine and running at its own rotational frequency. Application of SVMGA, to low speed bearings potentially holds distinct advantages for the following reasons:

- High training speeds.
- Good classification accuracy.
- Unlike the target value of SVM which is only used to handle binary problem, the target value of SVMGA is continuous.
- Feature vectors extracted from the samples belonging to the same class when fed to the trained SVMGA produces outputs of small deviation from the target values. [10]

Therefore these properties are more attractive to be used for solving multi-class problem to when compared with SVM based classifier or other AI algorithms. [19] used a hybrid method that combines wavelet packet transform (WPT) and support vector classification (SVC) to deal with the difficulty to obtain a large number of fault samples under practical conditions for mechanical fault diagnosis. [20] developed a novel personalized diagnosis methodology and used it to investigate shaft unbalance, misalignment, rub-impact and their combinations. The method looks promising and the probable tools needed for it are the numerical simulation (including finite element method), the big data technique or a combination of the two. The associated draw-back with this method is its cumbersomeness and the large amounts of data involved which is not appropriate for online solution as is the case found in this paper. [24] used a Redundant Second-Generation Wavelet Packet Transform (RSGWPT) and Local Characteristic-Scale Decomposition (LCD) to detect and extract fault feature from vibration signals. This method is based on the energy ratio in obtaining a set of desired intrinsic Scale Components (ISCs) while [25] did a fault classification and detection in a rotor bearing rig by using SVM and a compensation distance evaluation technique for selecting two sensitive features from the twelve-time domain features of the measured vibration signals from each of the accelerometer used.

The structure of this paper is organized as follows. Section 2 gives a brief preview of the theoretical background of SVM for regression and GA. Section 3 gives the layout of the experimental setup. Feature extraction methods are presented in section 4 and section 5 covers the results, discussion and conclusion.

2.1 Theoretical Background of SVM For Regression

Support vector machine genetic algorithm SVMGA theory is formed based on the principle of a support vector machine (SVM) which is used for time series prediction and genetic algorithm (GA) which is a probabilistic search technique for attaining an optimum solution to combinatorial problems that work on the principles of genetics. In a genetic algorithm, populations of candidate solutions (called individuals, creatures, or phenotypes) to an optimization problem are evolved towards better solutions. It is a method for solving optimization problems that are both constrained and unconstrained and is based on natural selection. With a GA-based solution, the basic form of the solution is predefined. The main difference between ANN and SVM is in the principle of risk minimization (RM). In SVM, the structural risk minimization (SRM) principle is used, thereby minimizing an upper bound on the expected risk while in ANN, traditional empirical risk minimization (ERM) is used minimizing the error on the training data and the difference in RM leads to better generalization performance in SVM than ANN [10, 12, 16, 21]. The aim of SVM is to obtain a function $f(x)$ which can predict the output y_i within the error limit of ε with the estimation function $f(x)$ being as flat as possible to ensure good generalization and variance.

The function is given as follows

$$f(s) = w \cdot x + b \qquad (1)$$

where w is the weight vector and b is a constant. This function is obtained by solving the following optimization problem

SVM for regression uses the same principles as SVM, for classification. With regression, a margin of tolerance ϵ is set in approximation to the SVM. Once trained, the SVMGA will generate predictions using the formula:

$$f_{(x)} \equiv \sum_{i=1}^{m} \vartheta_i \emptyset(x, x_i) + b \qquad (2)$$

For us to minimize the error, we will be individualizing the hyper-plane that maximizes the margin knowing fully well that of the error is been tolerated.

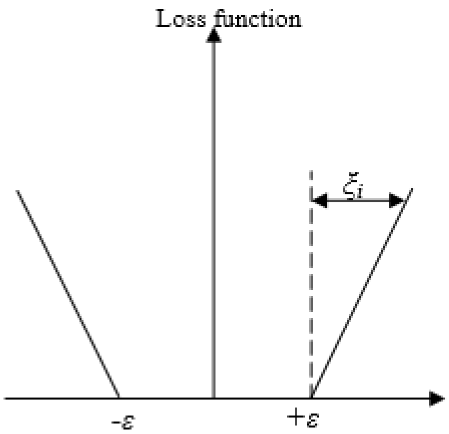

Fig. 1 The loss function of SVM

The quadratic optimization problem becomes

$$\min_{w} \frac{1}{2} w^T \cdot w \qquad (3)$$

$$s.t. \begin{cases} y_i - (w^T \cdot \emptyset(x) + b) \le \epsilon \\ (w^T \cdot \emptyset(x) + b) - y_i \le \epsilon \end{cases}$$

where $\emptyset(x)$ is the kernel function and w is the margin.

If the bound is added to the set, the tolerance on error that can be committed will be giving as;

$$\min_{w,b} \frac{1}{2} w^T \cdot w + C \sum_{i=1}^{m} (\xi_i + \xi_i^*) \qquad (4)$$

$$s.t. \begin{cases} y_i - (w^T \cdot \emptyset(x) + b) \le \epsilon + \xi_i \\ (w^T \cdot \emptyset(x) + b) - y_i \le \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0, \quad i = 1 \ldots \ldots \ldots \ldots \ldots m \end{cases}$$

where ξ_i and ξ_i^* are the slack variables and C is a positive constant which penalizes the errors larger than $\pm\varepsilon$ using ε- insensitive loss function given. Now been a minimization problem, we can set all constraints ≥ 0 by multiplying through by negative sign

$$\mathrm{R} = \min_{w,b} \frac{1}{2} w^T \cdot w + C \sum_{i=1}^{m} (\xi_i + \xi_i^*) \tag{5}$$

$$s.t. \begin{cases} -y_i + (w^T \cdot \emptyset(x_i) + b) + \epsilon + \xi_i \geq 0 \\ y_i - (w^T \cdot \emptyset(x_i) + b) + \epsilon + \xi_i^* \geq 0 \\ \xi_i, \xi_i^* \geq 0, \quad i = 1 \ldots \ldots \ldots \ldots \ldots m \end{cases}$$

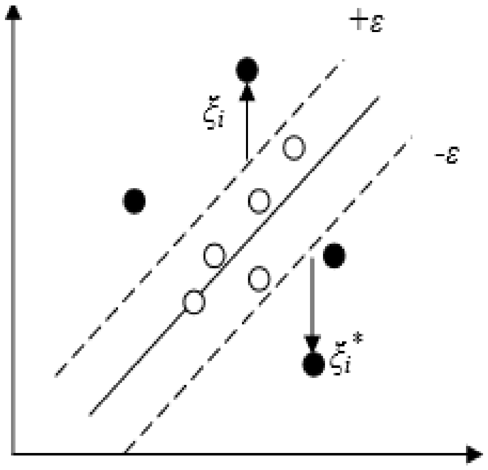

Fig. 2 The regression line of SVM

Fig. 2 show the regression line of SVM, the upper and lower boundary lines while Fig. 1 showing the loss function for SVR. For solving the optimization problem given in eq. (3) the following Lagrangian is needed

$$L = \frac{1}{2} w^T \cdot w + \sum_{i=1}^{m} C(\xi_i + \xi_i^*) - \sum_{i=1}^{m} (\eta_i \xi_i + \eta_i^* \xi_i^*) - \sum_{i=1}^{m} \alpha_i (\varepsilon + \xi_i - y_i + w \cdot x_i + b) - \sum_{i=1}^{m} \alpha_i^* (\varepsilon + \xi_i^* + y_i - w \cdot x_i - b) \tag{6}$$

Subject to $\alpha_i, \alpha_i^*, \eta_i, \eta_i^* \geq 0,\ i \ldots \ldots \ldots m$ are Lagrange multipliers that must be satisfied with the partial derivatives of the Lagrange equation L w.r.t. the primal variables w, b, ξ_i, ξ_i^* having to vanish for optimality

$$\begin{cases} \frac{\partial L}{\partial b} = \sum_{i=1}^{m} (\alpha_i^* - \alpha_i) = 0 \\ \frac{\partial L}{\partial w} = w - \sum_{i=1}^{m} (\alpha_i - \alpha_i^*) \emptyset(x_i) = 0 \qquad \rightarrow w = \sum_{i=1}^{m} \emptyset(x_i)(\alpha_i - \alpha_i^*) \\ \frac{\partial L}{\partial \xi_i} = C - \alpha_i - \eta_i = 0 \qquad \rightarrow \eta_i = C - \alpha_i, \quad \alpha_i \in [0, C] \\ \frac{\partial L}{\partial \xi_i^*} = C - \alpha_i^* - \eta_i^* = 0 \qquad \rightarrow \eta_i^* = C - \alpha_i^*, \quad \alpha_i^* \in [0, C] \\ \frac{\partial L}{\partial \eta_i} = \sum_{i=1}^{m} \xi_i = 0 \\ \frac{\partial L}{\partial \eta_i^*} = \sum_{i=1}^{m} \xi_i^* = 0 \end{cases} \tag{7}$$

Substituting eq. (7) into eq. (6), the dual optimization problem is given, for the sake of space a lot of substitution has been omitted to arrive at the following

$$max \frac{-1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \emptyset(x_i)(x_j) (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) - \varepsilon \sum_{i=1}^{m} (\alpha_i + \alpha_i^*) + \sum y_i (\alpha_i - \alpha_i^*) \tag{8}$$

$$s.t. \sum_{i=1}^{m} (\alpha_i - \alpha_i^*) = 0 \quad and\ \alpha_i, \alpha_i^* \in [0, C]$$

By solving the optimization problem, a linear regression function is presented as follows

$$f(x) = \sum_{i=1}^{m} (\alpha_i - \alpha_i^*)(x_i \cdot x) + b \tag{9}$$

To compute for b we exploit the Karush-Kuhn-Tucker (KKT) conditions which state that at the point of the solution the product between dual variables and constraints has to vanish,

$$\alpha_i(\varepsilon + \xi_i - y_i + \langle w, x_i \rangle + b) = 0 \tag{10}$$
$$\alpha_i^*(\varepsilon + \xi_i^* + y_i - \langle w, x_i \rangle - b) = 0$$

And

$$(C - \alpha_i)\xi_i = 0$$
$$(C - \alpha_i^*)\xi_i^* = 0 \tag{11}$$

which allows us to make relevant conclusions that are useful. Firstly only samples (x_i, y_i) with corresponding $\alpha_i^{\langle * \rangle} = C$ lie outside the ε-insensitive tube and secondly $\alpha_i, \alpha_i^* = 0$, meaning that there can never be a set of dual variables α_i, α_i^* which are both simultaneously nonzero [18]. Thus allowing us to conclude that

$$\varepsilon - y_i + \langle w, x_i \rangle + b \geq 0 \quad and \quad \xi_i = 0 \quad if \quad \alpha_i < C \tag{12}$$

$$\varepsilon - y_i + \langle w, x_i \rangle + b \leq 0 \quad if \quad \alpha_i > 0 \tag{13}$$

The kernel function is applied here to map the input vector into a high dimensional space because the linear regression function is not sufficient enough to process the non-linear problem. Hence the regression function is derived as follows

$$f(x) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) K(x_i, x) + b \tag{14}$$

where $K(x_i, x) = \varphi(x_i) \cdot \varphi(x)$ is a symmetric positive defined kernel function given by the Mercer's theorem [7,9]. In this paper the Gaussian and the exponential kernel functions are adopted which are respectively examples of the radial function kernel and the cubic polynomial function

$$K(x, y) = exp\left[\frac{-\|x - y\|^2}{2\sigma^2}\right] \tag{15}$$

where σ is a positive real number, alternatively it could be implemented using

$$K(x, y) = exp(-\gamma \|x - y\|^2) \tag{16}$$

The adjustable parameter σ plays an important role in the performance of the kernel and hence should be carefully tuned (by either using the classical technique which employs some method of determining a subset of centres or by first clustering to select a subset of centres) to the problem being solved. If overestimated, the exponential will behave almost linearly and the higher dimensional projection will start to lose its non-linear power, and if underestimated the function will lack regularization making the decision boundary highly sensitive to noise in the training data.

2.2 Genetic Algorithm to Configure SVR For Regression

A genetic algorithm is employed here to configure the SVR for regression. In these genetic algorithms consecutive populations of feasible solutions are created. It then evolves a population of chromosomes as potential solutions to an optimization problem for which the optimal solution is often obtained after a series of iterative operations.

To evaluate the fitness function of the chromosomes and genetic operators, selection and reproduction are employed to create new populations. A typical genetic algorithm requires: (1) A genetic representation of the solution domain and (2) A fitness function to evaluate the solution domain.

Once the genetic representation and the fitness function are both defined, the GA proceeds to initialize a population of solutions and then to improve it through repetitive application of the mutation, crossover, inversion and selection operators [23].

The three most important aspects of using GA are: (1) definition of the objective function, (2) definition and implementation of the genetic representation, and (3) definition and implementation of the genetic operators. Once these three have been defined, the generic GA should work fairly well.

Since a solution must represent a SVM for regression, the corresponding chromosomes are composed by two genes i.e. one for each SVM for regression parameter and the values of the genes which are obtained in ranges [0.01, 60000] and [1.0E-6, 8] for the genes representing C and γ respectively. The values of C and γ are limited into certain ranges to assure the generalization capability of the SVM for regression. These chromosomes are constructed by using a binary coding system (see table 2). To determine the optimal values of the regularization parameter C and γ which determines the tradeoff between the fitting error of the SVR for regression model and the model complexity and which assures the optimal accuracy and generalization stability simultaneously, the GA is used. It should be noted that the values of parameters C and γ are limited to certain ranges so as to assure the generalization capability of the SVR for regression. The values of these parameters in this work were set for [0.01, 60000] and [1.0E-6, 8]. The crossover and mutation rates were set to 0.5 and 0.1 respectively. The evolutionary process was terminated using two stopping criteria which are after 600 generations or if the fitness value of the best solution does not change after 60 generations. The formula for the fitness function used is $f = sum(x.^* z)/sum(z)$ and the constraints is $0 < x < 1, \quad sum(z.^* x) = class$

3. EXPERIMENTAL SETUP

The test rig setup comprises of a brushless AC motor (Rockwell Automation MPL-3680B) mounted on a NSK 6309 single row bearing used to drive the system as shown in Fig. 3. The angular velocity of the motor was retrieved from one of the angular outputs made available in the motor drive, which is a Rockwell Automation Kinetix 6000 series BM-01 and allows a continuous speed variation from 0 to 3600 rpm. A Soundwel AE sensor with model number SR 150M with a frequency range of 25-530kHz was used for measurement.

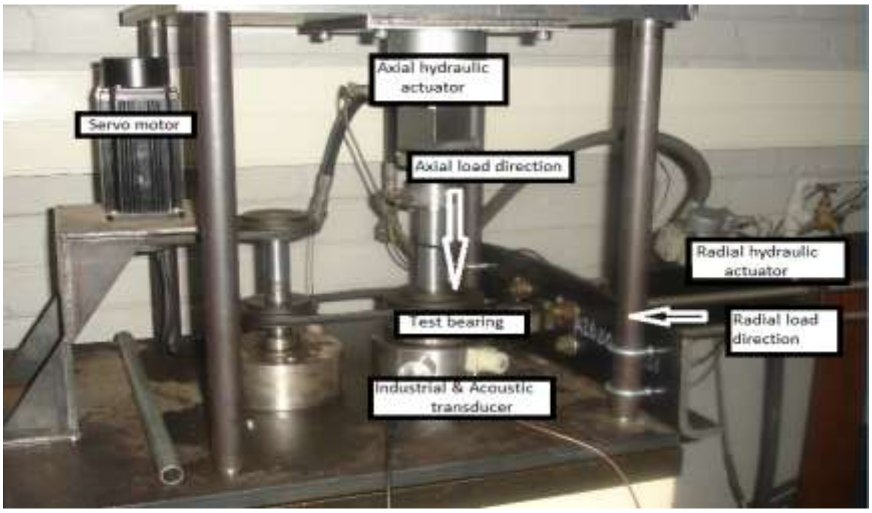

Fig. 3 Test rig setup

Two servo-hydraulic actuators were used for loading sinusoidally, axial and radial loads on test bearings whose purpose is to allow simulating a real life scenario of the bearings of a functional machine experiencing varying cyclic loads at different rotational speeds and frequencies. The test bearings used were three taper roller bearings (Timken HR 30307 J) which were inserted into the test rig one after the other with two having been introduced with defects and the third bearing been left undamaged. The test rotating speeds for the slow rotating bearings ranged from 70 to 100 rpm. Fig. 4 shows a schematic view of the test rig setup.

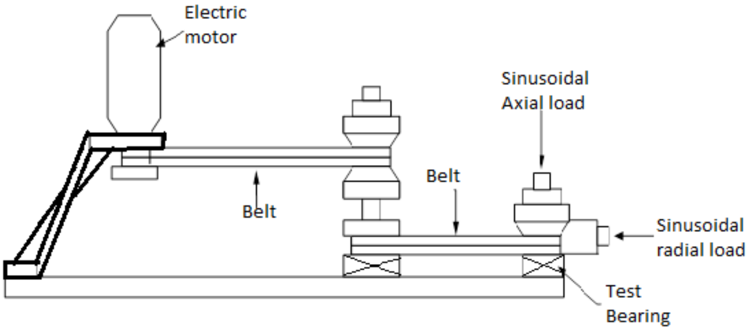

Fig. 4 Schematic diagram of test rig setup

The first bearing (good) was loaded sinusoidally with forces of amplitude 500N at a frequency of 2Hz on the axial load and amplitude of 900N at a frequency of 1Hz on the radial load. Bearing two, in which debris was introduced, was also loaded sinusoidally with forces of amplitude of 500N at a frequency of 2Hz in the axial direction and amplitude of 900N at a frequency of 1Hz in the radial direction. Bearing three, with a crack on the outer race was loaded sinusoidally with forces of amplitude of 500N on the axial at a frequency of 2Hz and 900N at a frequency of 1Hz in the radial direction as the other two bearings. The three roller bearing AE signatures were collected for four speeds, set at 70 rpm, 80 rpm, 90 rpm and 100 rpm using an FFT analyzer, a National Instrument data acquisition card (BNC-2110) with a shielded BNC connector block.

The induced crack was seeded on the outer raceway of the bearing (as shown in Fig. 5) with the use of a small hand drilling machine to which a small disk was mounted, which was then used to introduce a groove on the outer raceway of the taper bearing.

Fig. 5 Seeded damage on outer race of a bearing

4 Feature Extraction

The procedure used to develop the classifier in order to achieve the fault classification for the three conditions of consideration starts from the acoustic emission acquisition from two bearings which have faults induced and a good bearing. GA is used to determine the optimal values of the regularization parameter C and γ which determines the tradeoff between the fitting error of the SVR for regression model and the model complexity and assures the optimal accuracy and generalization stability simultaneously. It evolves a population of chromosomes as potential solutions to an optimization problem which is obtained after a series of iterative operations. The fitness function in GA is used to evaluate the goodness (i.e. the fitness) of the chromosomes and the genetic operators based on selection to create new populations (i.e. generations). The individual which gives the best solution in the final population was taken to define the best approximation to the optimum problem of investigation. After the signals created by the machine with normal and faulty conditions were measured, the optimized developed classifier was then used to classify the operational conditions of the machines. The GA is used to find the optimal values of C and γ that assures the optimal predication accuracy and generalization ability of the SVM for regression simultaneously. To determine the individuals that are included in the next generation (i.e. survivals) we employed tournament selection where only the best n solutions are copied straight into the next generation.

The three groups of bearing problems and their classification are presented in table 1.

Table 1: Bearing grouping and their classification

Fault Grouping	Classification	Bearing speed (rpm)	Class index	Binary code used
Good bearing	Good	70	1	0 0 1
Good bearing	Good	80	1	0 0 1
Good bearing	Good	90	1	0 0 1
Good bearing	Good	100	1	0 0 1
Debris induced bearing	Debris induced	70	2	0 1 0
Debris induced bearing	Debris induced	80	2	0 1 0
Debris induced bearing	Debris induced	90	2	0 1 0
Debris induced bearing	Debris induced	100	2	0 1 0
Outer race crack bearing	Outer race	70	3	1 0 0
Outer race crack bearing	Outer race	80	3	1 0 0
Outer race crack bearing	Outer race	90	3	1 0 0
Outer race crack bearing	Outer race	100	3	1 0 0

5. Results and Discussion

Fig. 6 shows a typical acoustic emission signal obtained from the test rig of the three bearing conditions at 100 rpm speed level. The skewness and kurtosis indicator used as input to the system often provide good detection at high speeds and decreases sharply in their detection abilities as the speed decrease as is reported in the literature.

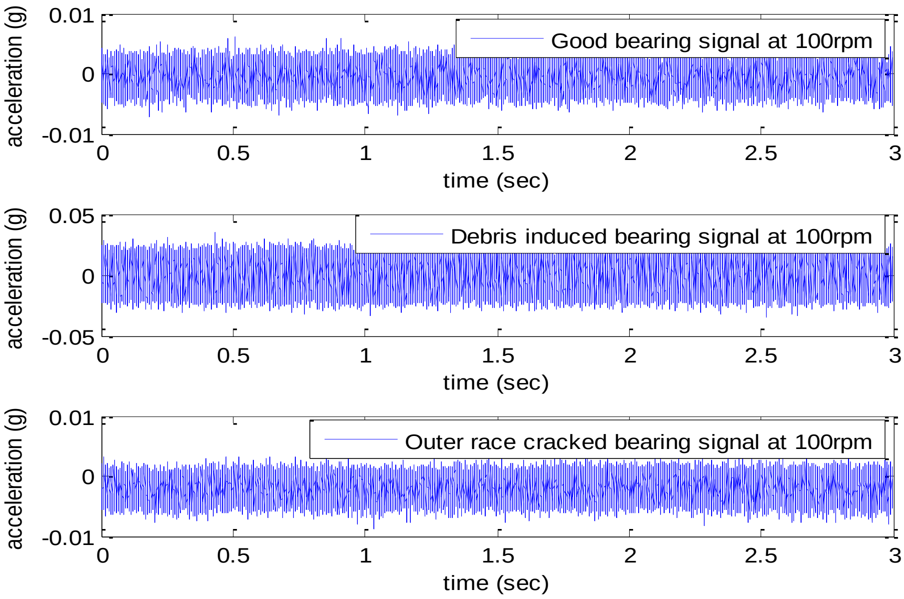

Fig. 6 Acoustic emission obtained from test rig

A test was done on neural network pattern recognition and table 2 shows the result obtained after the training was done for the neural network (NN).

Table 2: NN output classification result

	Result at speed 70 rpm, took 27 iterations		
Target class	**2**	**1**	**3**
Simulated class	1	1	3
	Result at speed 80 rpm, took 21 iterations		
Target class	2	1	3
Simulated class	2	1	3
	Result at speed 90 rpm, took 42 iterations		
Target class	2	1	3
Simulated class	2	1	3
	Result at speed 100 rpm, took 26 iterations		
Target class	2	1	3
Simulated class	2	1	3

The network has three layers (the input layer, the hidden layer and the output layer) with the hidden layer having 10 states. It was observed from the training run that at low speed under the varying load and low speed conditions which mimicked the rolling mills plants, the NN was not able to classify correctly at a low speed of 70 rpm but shows correct classification at higher speed. Figure 7 shows the confusion matrix for the neural network classification method.

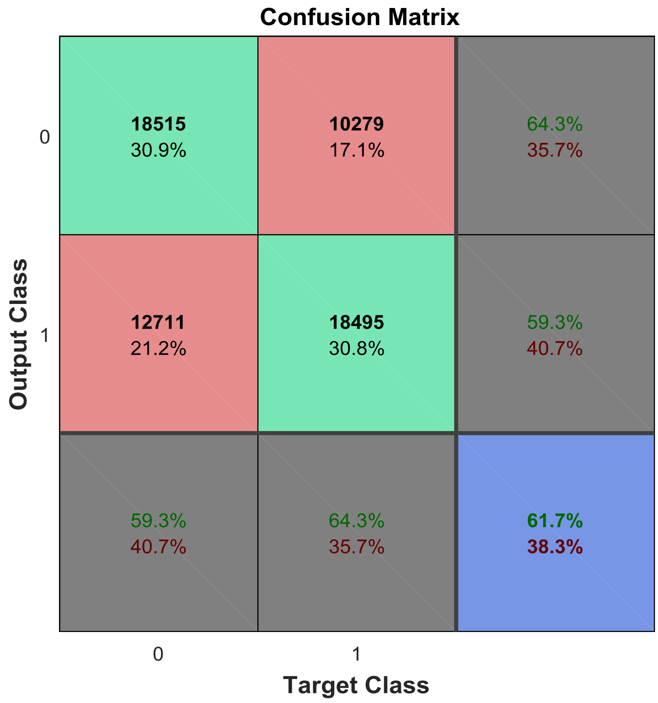

Fig. 7 The confusion matrix for the NN classification method

The total percentage of correctly classified cases with the NN as specified in Fig. 7 is 61.7 % while the misclassified cases is 38.3 %. This result from the confusion matrix shows that NN is not a very good classifier for the varying load and low speed condition of rolling element bearing.

As indicated in the previous section, the number of inputs to the network for the SVM for regression in this paper is thirteen. The SVMGA can be used for several applications in the field of engineering of which it is an innovation of SVM. Here it is used to develop the condition monitoring of the bearing i.e. to identify whether the bearing is defective or normal and to achieve this the Gaussian kernel and exponential kernel, spline, radial basis function (rbf), polynomial, periodic and sigmod function were used to perform various test of classification, so as to classify the simulated output to the target class that was set for the various conditions of the bearings as indicated in the experimental setup to check for which that can be used under this scenario. Only the Gaussian kernel functions proved effective for this purpose under this test at varying load and at low speed condition. However due to space limitations results for only few samples are shown here, for only the Gaussian and exponential kernel functions.

Table 3: Genetic Algorithm best fitness values for the bearings.

Bearing types	Applying classifier class	Generated GA possible class value matching to a measurement			Speed (rpm)
		Good bearing	Debris induced bearing	Outer race cracked bearing	
Good bearing	1.000000e+00	1.0000	-0.3968	-2.9452	70
Debris induced bearing	2.000000e+00	-0.6540	2.0000	1.2329	
Outer race cracked bearing	3.000001e+00	-1.0547	0.7102	3.0000	
Good bearing	9.999999e-01	1.0000	0.1936	-2.9858	80
Debris induced bearing	2.000000e+00	-0.7467	2.0000	0.3255	
Outer race cracked bearing	3.000000e+00	-1.1085	0.1383	3.0000	
Good bearing	1.898637e+00	1.0000	0.8860	1.8986	90
Debris induced bearing	1.142809e+00	1.1428	-0.5207	-0.7566	
Outer race cracked bearing	3.000000e+00	0.0005	1.4212	3.0000	
Good bearing	1.000000e+00	1.0000	0.4193	-0.9147	100
Debris induced bearing	2.000000e+00	-1.2400	2.0000	-0.2920	
Outer race cracked bearing	1.274203e+00	1.2742	0.0401	-0.8171	

The most suitable method was chosen according to the application constraints and the number of training samples, for easy recognition of classes. Target data for recognition was set to consist of vectors of all zero values except for a 1 in element I, where I is the class they represent (see table

1). Table 3 presents the values of the best possible classification values generated by GA by looking for weight values that produces the best fitness result based on the set of constrains provided in section 2.2. The debris induced bearing was classed 2, while the good bearing was classed 1 and the outer race cracked bearing was classed 3.

The results obtained for the classification of the bearings at different low speeds using the kernel functions of Gaussian and exponential is presented in table 4 and 5 below.

Table 4. Result of bearing classification using Gaussian kernel function.

Class	Kernel function	Debris bearing	Good bearing	Outer race crack bearing	Speed (rpm)	Classifying result
Target	**Gaussian**	**2**	**1**	**3**	**70**	**Correctly classified**
Simulated		**2**	**1**	**3**		
Target	**Gaussian**	**2**	**1**	**3**	**80**	**Correctly classified**
Simulated		**2**	**1**	**3**		
Target	**Gaussian**	**2**	**1**	**3**	**90**	**Correctly classified**
Simulated		**2**	**1**	**3**		
Target	**Gaussian**	**2**	**1**	**3**	**100**	**Correctly classified**
Simulated		**2**	**1**	**3**		

Table 5. Result of bearing classification using exponential kernel function.

Class	Kernel function	Debris bearing	Good bearing	Outer race crack bearing	Speed (rpm)	Classifying result
Target	**Exponential**	2	1	3	70	**Incorrectly classified**
Simulated		0	2	-2		
Target	**Exponential**	**2**	**1**	**3**	**80**	**Correctly classified**
Simulated		**2**	**1**	**3**		
Target	**Exponential**	**2**	**1**	**3**	**90**	**Correctly classified**
Simulated		**2**	**1**	**3**		
Target	**Exponential**	**2**	**1**	**3**	**100**	**Correctly classified**
Simulated		**2**	**1**	**3**		

In table 5 it is observed that at the lowest bearing speed of 70 rpm the exponential kernel function could not classify the three bearing conditions according to the index created to them. There was wrong classification of the good bearing, outer-race defect and the exponential function could not classify the debris bearing as shown in figure 8, Index 2 was repeated twice but in opposite direction. At every other speed it rightly classified the bearing. The precision obtained from the classification made by the Gaussian kernel function was perfect at the various speed category of the bearing condition. The simulated classes rightfully met their respective target class.

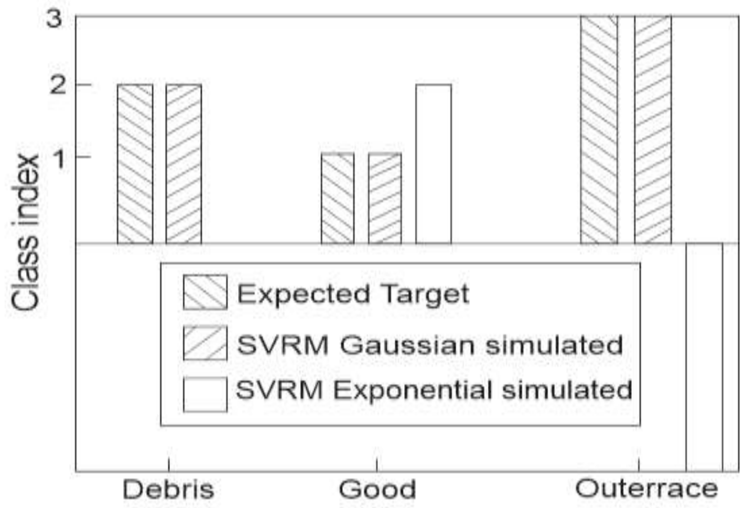

Fig. 8 Bearing classification as simulated at 70 rpm

The change in speed as indicated in the simulation done in table 5 did not affect the expected result for the Gaussian kernel. We were restricted to just three classes here in this work; the class of the good bearing indicated from table 1 as index 1, the debris induced bearing which is indicated as index 2 and the outer race cracked bearing indicated as index 3. To reduce space consumption bearing classification simulated at 90 rpm is given in figure 9.

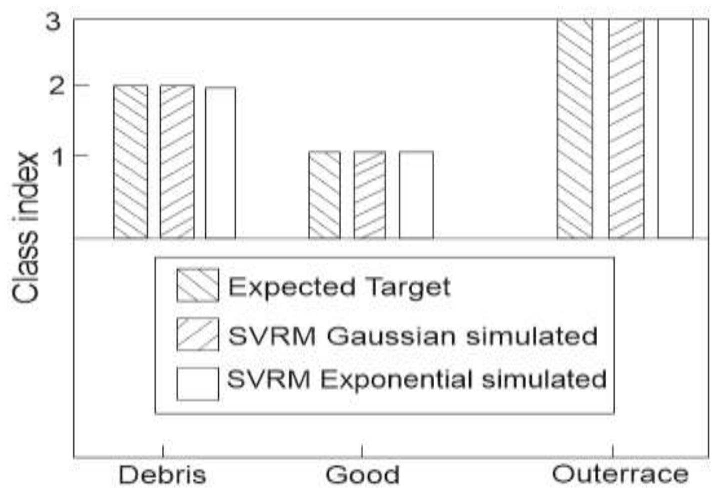

Fig. 9 Bearing classification as simulated at 90 rpm

Conclusion:

This work involves extracting representative statistical parameters by using a GA-based feature extractor from a raw acoustic emission dataset and using it to classify the inputs for SVM for regression. The GA was not only able to enhance the classification performance, but also reduced the dimensionality that describes the problem. Thus, the features extracted proved to be good indicators of defect intensity and it also showed that with the Gaussian kernel function, effective classification can be achieved over exponential kernel function with SVMGA as it classified correctly in all damage conditional cases of the bearing faults with all the data set generated in this experimental work.

The SVMGA algorithm has high accuracy in classification performance and wider generalization ability in small group of samples using the learning and testing pattern as revealed by the simulation results. Moreover a combination of GA and SVM for regression has been used for intelligent fault diagnostic, showing that the proposed method can reduce the dimensionality in dataset, solve the problem of outliers and be used to optimize features parameter selection for training purpose. SVMGA proves to be a better classifier that maximizes the fault classification accuracy as it was able to identify faults and classify better with the Gaussian kernel function than any other kernel function like the polynomial function.

There are several challenges associated with the use of only SVM which is known for its time series prediction, among which is the free parameter selection. To date there is no universal method for hyper-parameter selection which contains some parameters such as the kernel parameter, the regularization parameter that control the generalization performance of SVM and the ε-insensitive zone which determines the number of support vectors. Hence if one is using arbitrary SVM parameters, the performance of SVM could differ over a wide range. This makes SVMGA (an innovation to SVM) a better option for solving classification problem especially at low speed conditions as it gives a decision that is much unique than SVM while retaining its classification accuracy.

Hence the SVMGA with the use of Gaussian kernel function proves to be superior to the exponential kernel function in the classification of faults. Compared with the traditional classification methods reported in many other studies, this study accomplished effective classifying of different bearing fault patterns especially at low speeds and at varying load condition for which the results showed whether the bearing is normal or defective.

References:

[1] Hamadache M. and Lee D., Improving signal-to-noise ratio (SNR) for inchoate fault detection based on principal component analysis (PCA), *14th International Conference on Control, Automation and Systems (ICCAS 2014)*, 2014; Oct. 22-25, in KINTEX, Gyeonggi-do, Korea.

[2] Jamaludin N., Mba D. and Bannister R. H., Condition monitoring of slow-speed rolling element bearing using stress waves, *Proceedings of the institute of Mechanical Engineering, Part E: Journal of process Mechanical Engineering* Vol 215 No 4 2001; pp 245 - 271

[3] Achmad W., Eric Y. K., Jong-Duk S., Bo-Suk Y., Andy C.C.T., Dong-Sik G., Byeong-Keun C. and Joseph M., Fault diagnosis of low speed bearing based on relevance vector machine and support vector machine, *Elsevier: Expert Systems with Applications* 36 2009; pp. 7252-7261.

[4] Changqing S., Dong W., Yongbin L., Fanrang K. and Peter W.T., Recognition of rolling bearing fault patterns and sizes based on two-layer support vector regression machines, *Smart Structures and Systems*, Vol. 13, 2014; No. 3 000-000.

[5] Yu Y., Dejie Y. and Junsheng C., A fault diagnosis approach for roller bearing based on imf envelope spectrum and svm, *Elsevier (ScienceDirect) Measurement* 40. 2006; pp. 943-950.

[6] Yang B., Han T. and Hwang W., Fault diagnosis of rotating machinery based on multi-class support vector machine, *Journal of Mechanical Science and Technology (KSME Int. J.)*, Vol. 19, No. 3, 2005; pp 846 - 859

[7] Kan X. *ECE 480 Team 4*, Support vector machine, concept and matlab build 2011.

[8] Li X. and Chen W., Rolling bearing fault diagnosis based on physical model and one-class support vector machine, *Hindawi Publishing Corporation, ISRN Mechanical Engineering* Vol. 2014, Article ID 160281, 4 pages.

[9] Sloukia F. E, Bouarfa R., Medromi H. and Wahbi M., Bearings prognostic using mixture of gaussian hidden markov model and support vector machine, *International Journal of Network Security & Its Applications (IJNSA)*, Vol. 5, No. 3 2013; pp 85-97.

[10] Shen C., Wang D., Liu Y., Kong F. and Tse P.W., Recognition of rolling bearing fault pattern and sizes based on two-layer support vector regression machines, *Smart Structures and Systems*, Vol. 13, No. 3 2014.

[11] Muhammet U., Mustafa D., Mustafa O. and Haluk K., Fault diagnosis of rolling bearing based on feature extraction and neural network algorithm, *Recent Advances in Telecommunications, Signals and Systems*. ISBN: 978-1-61804-169-2. 2013; www.wseas.us/e-library/confrences/2013/lemesos

[12] Hariharan V. and Srinivasan P.S.S., New approach of classification of rolling element bearing fault using artificial neural network, *Journal of Mechanical Engineering*, Vol. ME 40, No. 2 2009; pp. 119 – 130.

[13] Hagan M.T., Demuth H.B. and Beale M., Neural network design, *Boston: Thomson*, 1997.

[14] Shuang L., Fujin Y. and Jing L., Bearing fault diagnosis based on k-l transform and support vector machine, *Third International Conference on Natural Computation* (ICNC 2007) 0-7695-2875-9/07 2007.

[15] Yu Y., Song M. and Song J., A novel hyper-parameters selection approach for support vector machines to predict time series, *Journal of Computers*, Vol. 7, No. 12. 2012; pp. 2921-2930.

[16] Rao S. S, Mechanical vibrations, *Fifth Edition in SI Units, University of Miami. Prentice Hall. Published in 2011 by Pearson Education South Asia Pte Ltd*. 23/25 First Lok Yang Road, Jurong Singapore 629733. 2011.

[17] Gun S. R., Support vector machines for classification and regression, *Technical Report, Faculty of Engineering, Science and Mathematics, School of Electronics and Computer Science, University of Southampton*. 1998.

[18] SmolHamadache A.J, Lee M D.,. Improving signal-to-noise ratio (SNR) for inchoate fault detection based on principal component analysis (PCA). *In 14th International Conference on Control, Automation and Systems (ICCAS 2014). KINTEX, Gyeonggi-do, Korea*, 2014 pp. 561–566.

[19] Li, P., Jiang, Y. & Xiang, J., Experimental Investigation for Fault Diagnosis Based on a Hybrid Approach Using Wavelet Packet and Support Vector Classification. *The Scientific World Journal*, 2014. Available at: http://dx.doi.org/10.1155/2014/145807.

[20] Xiang, J. & Zhong, Y., 2016. A Novel Personalized Diagnosis Methodology Using Numerical Simulation and an Intelligent Method to Detect Faults in a Shaft. *applied sciences*, 414(6).

[21] Y. Kim, A. C. C Tan, J. Mathew and B. Yang, Condition monitoring of low speed bearings: a comparative study of the ultrasound technique versus vibration measurements, *WCEAM Paper* 029, 2006; pp 1-10

[22] Guo H., Jack L. B., and Nandi A., Feature Generation using Genetic Programming with Application to Fault Classification, *IEEE Transactions on Systems, Man, And Cybernetics-Part B: Cybernetics*, 2005 Vol. 35, No. 1 pp. 89-99

[23] Lu D and Qiao W., A GA-SVM Hybrid Classifier for Multiclass Fault Identification of Drivetrain Greaboxes,*Energy Conversion Congress and Exposition (ECCE)*, 2014 IEEE pp. 3894-3900

[24] Tong Q., Han B., Lin Y., Zhang W., Cao J. and Zhang X. A Fault Feature Detection Approach for Fault Diagnosis of Rolling Element Bearings Based on Redundant Second Generation Wavelet Packet Transform and Local Characteristic-Scale Decomposition, *Journal of Vibration Engineering & Technologies, Vol. 5, No.1,* 2017 pp. 101-110.

[25] Fatima S., Mohanty A.R. and Kazmi H.F. Fault Classification and Detection in a Rotor Bearing Rig, *Journal of Vibration Engineering & Technologies, Vol. 4. No. 6*,2016