International Journal of Fluid Power 7 (2006) No 2 pp. 39-52

FEATURE EXTRACTION, OPTIMIZATION AND CLASSIFICATION BY SECOND GENERATION WAVELET AND SUPPORT VECTOR MACHINE FOR FAULT DIAGNOSIS OF WATER HYDRAULIC POWER SYSTEM

Han Xin Chen, Patrick S. K. Chua and Geok Hian Lim

School of Mechanical and Aerospace Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798
PG01074075@ntu.edu.sg, MSKCHUA@ntu.edu.sg

Abstract

The work described in this paper investigates the fault diagnosis of water hydraulic motor by the optimization and automatic classification of the feature values. The second generation wavelet for the vibration signals analysis of the water hydraulic motor was proposed to extract the feature values. The new optimization method by bi-classification support vector machine (SVM) was proposed to select the optimal feature values based on a rank criterion and the algorithm was developed here. In order to classify the conditions of the pistons used in the hydraulic motor, a two-level structure based on the multi-classification was developed in this work. The multi-classification method of SVM for the fault diagnosis of a piston crack was investigated. The winner-takes-all scheme was studied. The results of the classification were found to be successful.

Keywords: Fault diagnosis, Support vector machine, Water hydraulic system, Feature extraction, Wavelet transform, Neural network

1 Introduction

Water hydraulics was overtaken by oil hydraulics in the early part of the 19th century in terms of research effort and industrial applications (Krutz and Chua, 2004; Lim et al., 2002). In recent years, due to the growing concern about environmental, safety and health issues, there was renewed interest in water hydraulics. As water is nontoxic, environmentally friendly, and readily available, many industries are turning to water hydraulic systems to replace the oil hydraulic counterparts. Water hydraulic systems have been increasingly used in the farming, forestry, food, pharmaceutical and paper industries (Sorensen, 1999; Trostmann, 1995; Simon, 1996). Thus the increasing use of water hydraulic systems will result in a greater need to maintain the systems. The water hydraulic motor is commonly used to provide high torque. Pistons are the principal operating elements of piston-type machines and their performance invariably depends on the smooth and efficient motion of pistons in the cylinder bore. Any crack in a piston will affect the leakage and friction between the sliding surfaces in the cylinder bore. This crack will propagate and gradually cause the motor's performance to deteriorate. It will eventually lead to dangerous situations such as severe vibration or sudden shutdown during operation.

Water hydraulic systems have gained momentum in recent years. There is an increasing amount of research published on axial piston machines (Ivantysynova et al., 2005, Ivantysynova and Lasaar, 2004; Chen, 2006). Research work has been done on fault diagnosis of oil hydraulics and other mechanical systems and this provides a useful reference for condition monitoring and fault diagnosis of water hydraulic systems (Lim, 2003; Crowther, 1998). The approach using rule-based concepts was presented for a number of faults in the typical hydraulic circuits (Watton, 1994). The algorithm for fault detection based on mathematical models is a powerful method for many fluid power systems. The nonlinear model based on neural network was presented and applied to a simulated electro-hydraulic rotary drive system (Daley et al., 1996). Rule-based systems were designed for the on-line fault diagnosis of fluid power system (Stewart, 1994).

Artificial Neural Networks (ANN) have demonstrated great potential for application in machine learning, computer vision and pattern recognition, and the fault diagnosis and classification of fluid power systems (Precetti and Krutz, 1993; Haney et al., 1994; Le et al., 1998; Samanta et al., 2001; Lott and Griffiths, 1997). However, the traditional neural network ap-

This manuscript was received on 27 December 2005 and was accepted after revision for publication on 30 June 2006.

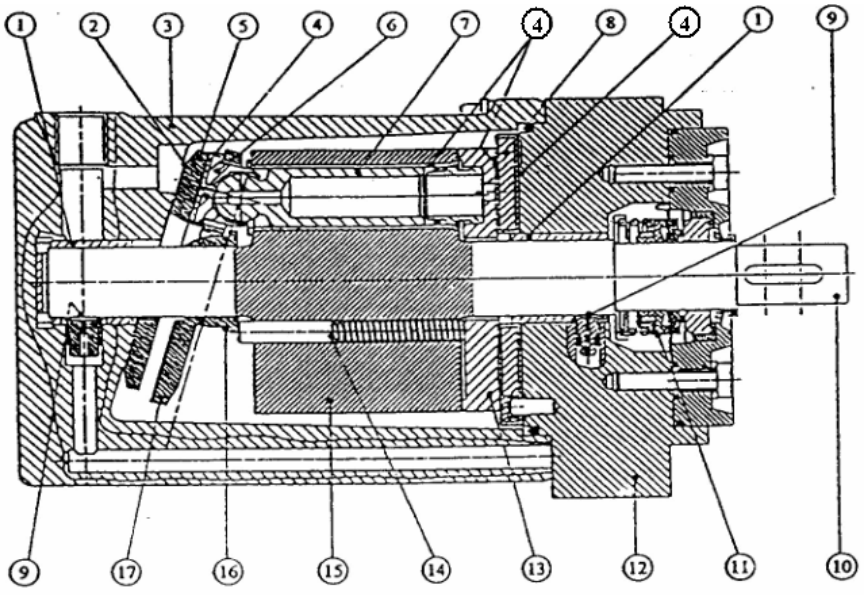
proaches have limitations. Support vector machines (SVMs) are a comparatively recent development. The principle of SVMs is risk minimization that is different from ANN whose principle is traditional empirical risk minimization. Empirical risk minimization is used to minimize the error on the training data. SVMs (Vapnik, 1995) are based on the statistical learning theory and are appropriate for dealing with classification and regression problems. Recently, there is application of SVM in machine condition monitoring (Nandi, 2000).

2 Water Hydraulic Motor and Experimental Set-Up

The actuator studied here was a five-piston axial piston motor used in a water hydraulic system. An accelerometer mounted on the casing of the Danfoss MAH 12.5 water hydraulic motor was used to obtain the vibration signal from the motor. The complete water hydraulic system was provided by Danfoss. Figure 1 shows the general structure of the water hydraulic motor. An axial piston motor consists mainly of a valve port plate with inlet and outlet ports, a swash-plate, an outer shell, a cylinder block, pistons with shoes, a bias spring, a port flange and a shaft. Pistons fit within the bores of the cylinder barrel and are parallel to the output shaft. The swash-plate is positioned at an angle and acts as a surface on which the piston shoes travel. The shoes are held in contact with the swash-plate by the retaining rings and the bias spring. The port plate separates the incoming fluid from the discharging fluid. The output shaft is connected to the cylinder barrel.

As the water enters the inlet and exits at the outlet of the hydraulic motor, the pressure in the cylinder chamber alternates from high pressure to low pressure (Ivantysyn et al., 2002). This causes pressure pulsation to occur. The total cylinder area inside a supply port is variable as a result of the cyclic variations of the piston passing through the supply port. It generates the variations of the axial output moment. The variations of the forces are applied from the piston to the swash plate and the valve cover. The force between the support for the swash plate and the valve cover is opposite in direction. The hydraulic motor body vibrates as a result.

The water hydraulic motor test rig is shown in Fig. 2. It consists of a MAH 12.5 Nessie water hydraulic motor, a AB 34000 brushless servo motor, a digital torque meter and a water hydraulic supply system. The hydraulic circuit of the motor test rig was shown in Appendix A. The water hydraulic system allows axial piston motor operation in the range of 300-3000 rev/min and 0-6 Nm. The operational display processes the signals and shows the results as the digital value of torques and rotational speeds. The brushless servo motor provides the precise torque in the range of 0-10.2 Nm as resistive torque for the Nessie motor. A piezoelectric accelerometer (Bruel and Kjær type 4393) was mounted on the shell of a Nessie motor near the inlet of the water hydraulic motor. The sampling rate uses the Nyquist formula and was chosen to be 2 kHz which is sufficient for the type of signals considered in this work. 2000 data points were taken in each experimental run. The rotational speed of Nessie motor was 630 rev/min. The output axis torque of Nessie motor was 5 Nm.



1-bearings
2-hydrostatic bearing in piston shoe
3-outer housing
4-reinforced plastic
5-swash plate
6-shoe
7-piston
8-valve port plate
9- hydrostatic bearings
10-motor shaft
11-shaft seal
12-port flange
13-thrust plate
14-spring
15-cylinder barrel
16-bushing with a spherical outer surface
17-retaining ring

Fig. 1: *Swash plate water hydraulic motor with five pistons (Danfoss)*

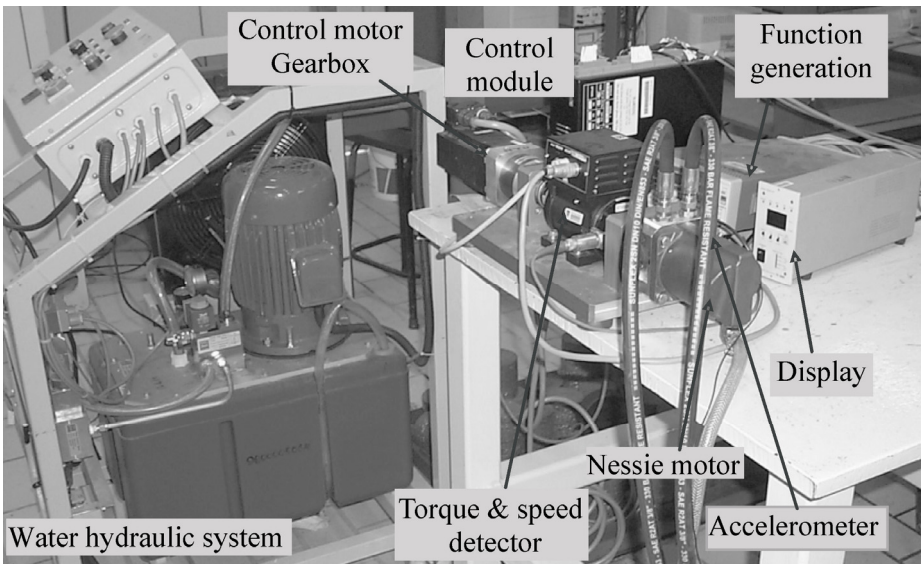


Fig. 2: *The water hydraulic motor test rig*

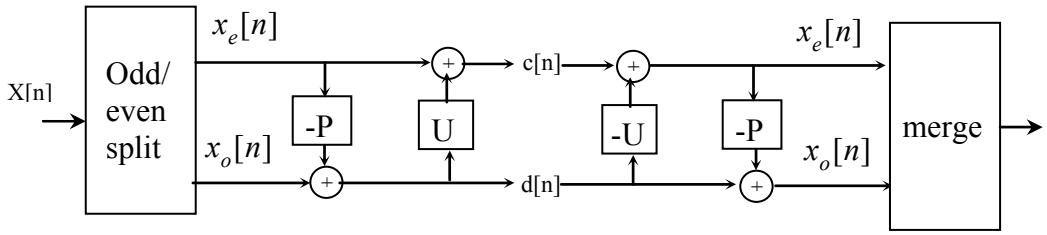


Fig. 3: *Lifting step and inverse lifting step*

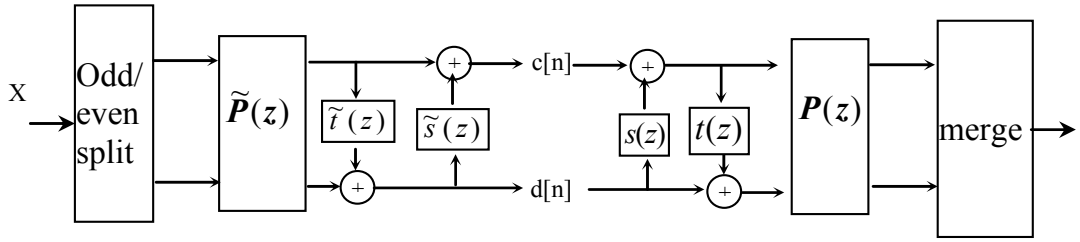


Fig. 4: *The low-pass subband is lifted with the help of the high-pass subband (predict) and the high-pass subband is lifted with the help of the low-pass subband (update)*

Piston crack was created in the axial piston of the Nessie motor by using electro-discharge machine. The introduced crack is 0.1 mm wide. There were five operational conditions in this research work: one normal condition and four faulty conditions with four different kinds of piston cracks. In order to detect the different piston crack locations and piston crack lengths in the Nessie motor, four kinds of piston cracks were introduced as follows: (1) Back long crack (BLC): the crack is 10 mm long and located in the back section of the piston. (2) Back short crack (BSC): the crack is 5 mm long and located in back section of the piston. (3) Front long crack (FLC): the crack is 10 mm long and located in the front section of the piston. (4) Front short crack (FSC): the crack is 5 mm long and located in the front section of the piston. There were 49 vibration signals under NC and 50 vibration signals for BLC, BSC, FLC and FSC operating condition respectively. The number of the raw vibration signals was 249.

3 Feature Extraction

3.1 Second Generation Wavelet

Wavelet represents the general functions as data building blocks, which can obtain the important data sets with a small number of coefficients. The second-generation wavelet transform is constructed by the lifting scheme. The lifting scheme is a spatial or time domain construction of biorthogonal wavelets (Sweldens, 1996; Daubechies and Sweldens, 1998). The basic idea behind lifting is that it provides a simple relationship between all multi-resolution analyses that share the same low-pass filter or high-pass filter. The low-pass filter gives the coefficients of the refinement relation, which entirely determines the scaling functions where the coefficients are given by the high-pass filter.

The lifting scheme consists of iterations of the following three basic operations as shown in Fig. 3.

Firstly, the split divides the original data into two subsets. The original signal is split to the even indexed points and the odd indexed points, that is $\boldsymbol{x}_e[n] = \boldsymbol{x}[2n]$ and $\boldsymbol{x}_o[n] = \boldsymbol{x}[2n+1]$. This generates the wavelet coefficients $\boldsymbol{d}[n]$ as the error by the predicting operator P:

$$\boldsymbol{d}[n] = \boldsymbol{x}_o[n] - \mathrm{P}(\boldsymbol{x}_e[n]) \tag{1}$$

The process of computing a prediction and recording it is called a lifting step. The original signal is transformed from $(\boldsymbol{x}_o,\boldsymbol{x}_e)$ to $(\boldsymbol{x}_o,\boldsymbol{d})$. There is aliasing in the even samples due to the subsample. The second lifting step is introduced to solve this problem. The update combines $\boldsymbol{x}_e[n]$ and $\boldsymbol{d}[n]$ to obtain the scaling coefficients $\boldsymbol{c}[n]$ that represents the coarse approximation in the original signal. The update operator U is used with the wavelet coefficients for $\boldsymbol{x}_e[n]$.

$$\boldsymbol{c}[n] = \boldsymbol{x}_e[n] + \mathrm{U}(\boldsymbol{d}[n]) \tag{2}$$

These three steps form the lifting stage. The iterations of the lifting stage on the output samples create the complete set of discrete wavelet transform scaling and wavelet coefficients $\boldsymbol{c}_j[n]$ and $\boldsymbol{d}_j[n]$ at each scale j.

Figure 4 shows that the low-pass subband is lifted with the help of the high-pass subband (predict) and the high-pass subband is lifted with the help of the low-pass subband (update). The lifting steps are easily inverted. The following equations for invertible lifting is derived:

$$\boldsymbol{x}_e[n] = \boldsymbol{c}[n] - \mathrm{U}(\boldsymbol{d}[n]) \qquad \boldsymbol{x}_o[n] = \boldsymbol{d}[n] + \mathrm{P}(\boldsymbol{x}_e[n]) \tag{3}$$

The lifting theory shows that any other new finite filter $\boldsymbol{g}^1$complementary to $\boldsymbol{h}(z)$ is of the form that is called the update step.

$$\boldsymbol{g}^1(z) = \boldsymbol{g}(z) + \boldsymbol{h}(z)\boldsymbol{s}(z^2) \tag{4}$$

where $\boldsymbol{P}(z)$ and $\widetilde{\boldsymbol{P}}(z)$ are the polyphase matrix. $\boldsymbol{s}(z^2)$ is a Laurent polynomial, $\boldsymbol{g}(z)$ is high-pass filter, and $\boldsymbol{h}(z)$ is low-pass filter.

In update step, the low-pass subband is lifted with the high-pass subband. The new filter $\widetilde{\boldsymbol{h}}^{\mathrm{new}}(z)$ complementary to $\widetilde{\boldsymbol{g}}(z)$ is indicated as:

$$\widetilde{\boldsymbol{h}}^1(z) = \widetilde{\boldsymbol{h}}(z) + \widetilde{\boldsymbol{g}}(z)\widetilde{\boldsymbol{s}}(z^2) \tag{5}$$

where $\widetilde{\boldsymbol{s}}(z^2)$ is a Laurent polynomial, $\widetilde{\boldsymbol{g}}(z)$ is high-pass filter, and $\widetilde{\boldsymbol{h}}(z)$ is low-pass filter.

The high-pass subband is lifted with the help of the low-pass subband. The dual lifting equations (the predict step) are given as

$$\widetilde{\boldsymbol{g}}^1(z) = \widetilde{\boldsymbol{g}}(z) + \widetilde{\boldsymbol{h}}(z)\widetilde{\boldsymbol{t}}(z^2) \tag{6}$$

$$\boldsymbol{h}^1(z) = \boldsymbol{h}(z) + \boldsymbol{g}(z)\boldsymbol{t}(z^2) \tag{7}$$

Some wavelet filters such as Haar, Daubechies, [9-7] and Cubic B-splines wavelet function can be used to build the second-generation wavelet transform into lifting steps. Haar wavelet was used in this research to generate the lifting step. Haar wavelet has the following filters:

$$\begin{aligned} h(z) &= 1 + z^{-1} & g(z) &= -1/2 + 1/2\, z^{-1} \\ \tilde{h}(z) &= 1/2 + 1/2\, z^{-1} & \tilde{g}(z) &= -1 + 1z^{-1} \end{aligned} \tag{8}$$

The multi-decomposition consists of the following implementation:

$$\begin{aligned} \boldsymbol{c}_1^{(0)} &= \boldsymbol{x}_e & \boldsymbol{d}_1^{(0)} &= \boldsymbol{x}_o \\ \boldsymbol{d}_1 &= \boldsymbol{d}_1^{(0)} - \boldsymbol{c}_1^{(0)} & \boldsymbol{c}_1 &= \boldsymbol{c}_1^{(0)} + 1/2\,\boldsymbol{d}_1 \end{aligned} \tag{9}$$

The wavelet transform can be used to represent efficiently the localized features of the signals. It was an ideal tool for the extraction of features. Especially, the vibration was induced by the fluid, which was characteristic of the impulse signal. The impulse component in the vibration signals was an important feature to diagnose the mechanical condition. Figure 5 shows the multi-decomposition based on the lifting scheme, which was shown in Fig. 4. The original signal was decomposed into the scaling coefficients $\boldsymbol{c}_1[n]$ and wavelet coefficients $\boldsymbol{d}_1[n]$. The multi-decomposition based on the lifting scheme split the scaling coefficients into two parts using the same scheme to produce $\boldsymbol{c}_2[n]$ and $\boldsymbol{d}_2[n]$ in the next step. Hence the wavelet decomposition of the signal at scale j had the structure $[\boldsymbol{c}_j[n],\boldsymbol{d}_1[n],\cdots,\boldsymbol{d}_j[n]]$.

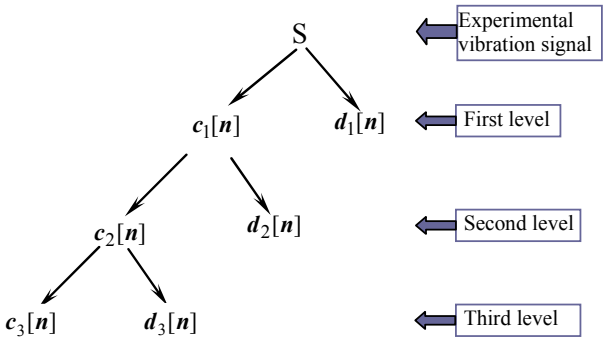


Fig. 5: *Wavelet multi-decomposition tree based on lifting scheme*

Figure 6 shows the typical wavelet multi-decomposition of the experimental vibration signals, which was obtained from water hydraulic motor. Figure 6 (a) shows the original signals, which was collected from the hydraulic motor. Figure 6 (b) shows the scaling coefficient at first scale ($\boldsymbol{c}_1[n]$). Figure 6 (c) shows the wavelet coefficient $\boldsymbol{d}_j[n]$ that was characteristic of the impulse signal. The impulse pressure signal was produced by the pumping mechanism when the piston rotates from the return line to supply line, which results in the overshoot of the piston pressure. Figures 6 (d) and (e) show the scaling coefficients $\boldsymbol{c}_2[n]$ and wavelet coefficients $\boldsymbol{d}_2[n]$ at second scale.

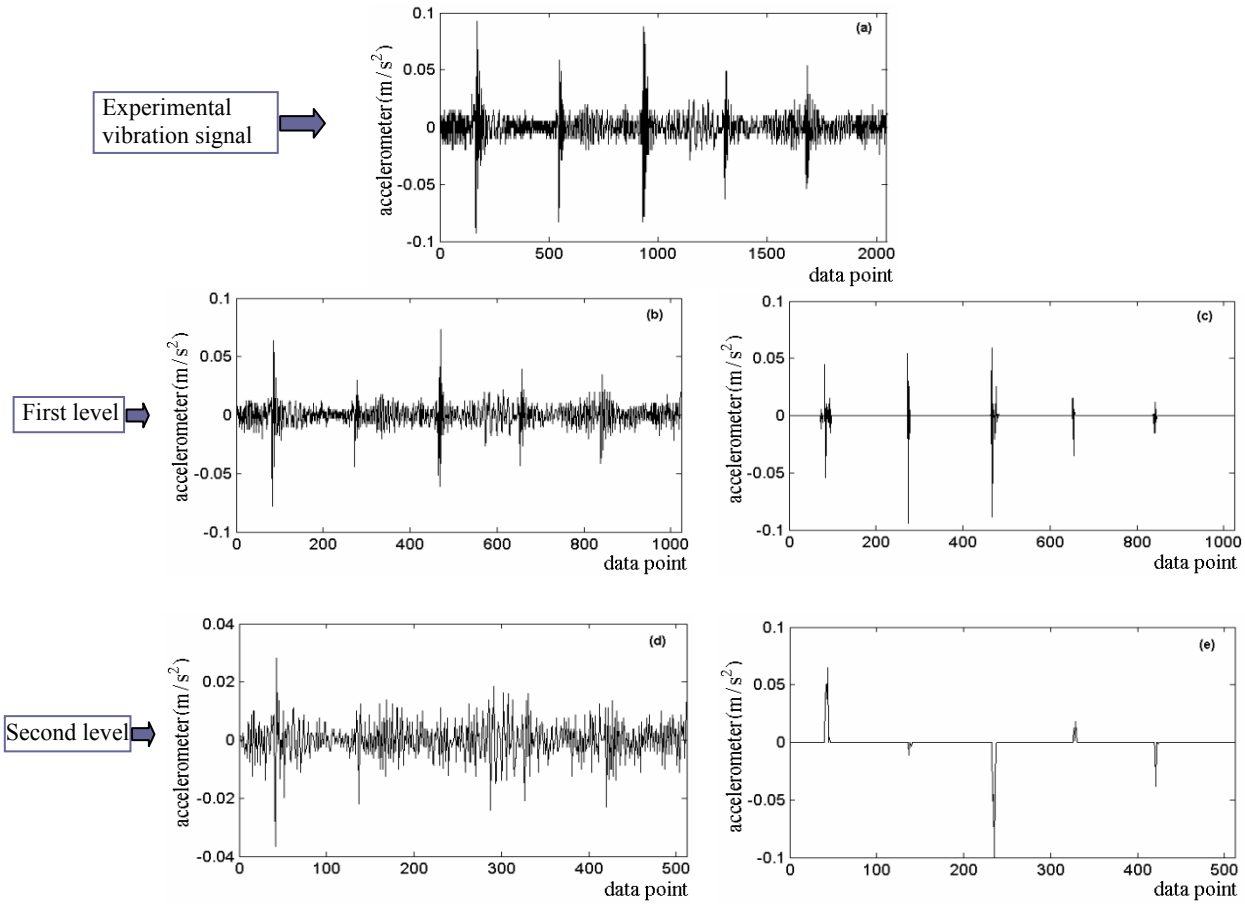


Fig. 6: *The multi-decomposition based on the lifting scheme of the experimental housing vibration signal of motor*

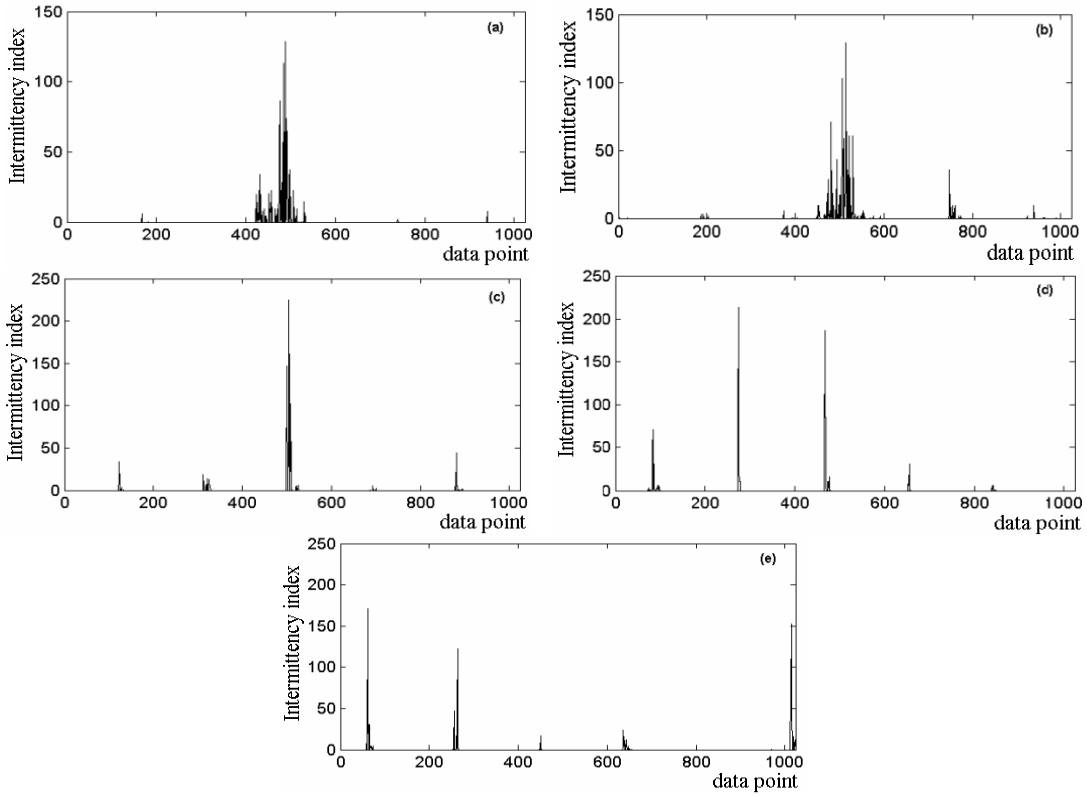


Fig. 7: *Intermittency indices at the first wavelet scale for the vibration signal under five conditions (a) BLC (b) BSC (c) FLC(d) FSC (e) NC*

The wavelet scaling coefficients in the multi-decomposition of the vibration signal were the important information in the impulse vibration signals. In order to diagnose the different conditions, the wavelet scaling coefficients at first scale were used to extract the features.

The intermittency index was used to determine the wavelet scaling coefficients at first scale. The intermittency at each scale can be viewed directly using the intermittency index defined as:

$$\boldsymbol{I}_{\mathrm{m,n}} = \frac{(\boldsymbol{T}_{\mathrm{m,n}})^2}{\left[\boldsymbol{T}^2_{\mathrm{m,n}}\right]_{\mathrm{m}}} \tag{10}$$

Here $\boldsymbol{T}_{\mathrm{m,n}}$ were the wavelet coefficients at scale index *m*. $\left\langle \boldsymbol{T}^2_{\mathrm{m,n}} \right\rangle_{\mathrm{m}}$ was the second order statistical moment of the wavelet coefficients at scale index *m*. For example, a constant value of $\boldsymbol{I}_{\mathrm{m,n}} = 1$ for all *m* and *n* meant that there was no flow variation. A value of 10 at a specific set of indices *m* and *n* meant that there was ten times more energy contained within the coefficients at that location than for the temporal mean for the scale at that location in the signal. Figure 7 showed the intermittency indices at the first wavelet scale for the vibration signals under the five conditions. There were different amplitudes and number of the intermittency indices.

3.2 Statistics of the Coefficients

In order to extract the feature value for condition classification, a number of the different statistical features were used based on the moments of the wavelet coefficients and scale coefficients of the vibration wavelet decomposition. The discrete wavelet transform coefficients $\boldsymbol{T}_{\mathrm{m,n}}$ generated from the full decomposition using the second generation wavelet were considered. The wavelet statistics such as the variance, skewness and flatness were applied as the feature values.

The p^{th} order statistical moment of the wavelet coefficients $\boldsymbol{T}_{\mathrm{m,n}}$ at the scale index *m* is defined as:

$$\left[\boldsymbol{T}^{\mathrm{p}}_{\mathrm{m,n}}\right]_{\mathrm{m}} = \sum_{\mathrm{n}=1}^{2^{\mathrm{M-n}}-1} \left(\boldsymbol{T}_{\mathrm{m,n}}\right)^{\mathrm{p}} \Big/ 2^{\mathrm{M-n}} \tag{11}$$

where the coefficients at scale *m* were used in the summation. The brackets denoted the average on the number of the coefficients at scale *m*. The 2nd order to 7th order statistical moments of the wavelet coefficients and scale coefficients at each level were used as the feature values for the classification. There were twelve feature values at each level. As shown in section 3.1, the multi-decomposition by the second generation wavelet had three levels. These thirty-six feature values were included in the input feature set. That was $\left[\boldsymbol{T}^2_{1,\mathrm{n}}\right]_{1,\mathrm{A}} \cdots \left[\boldsymbol{T}^7_{1,\mathrm{n}}\right]_{1,\mathrm{A}}$ for the scale coefficients $\boldsymbol{C}_1[n]$ of the first level, $\left[\boldsymbol{T}^2_{1,\mathrm{n}}\right]_{1,\mathrm{D}} \cdots \left[\boldsymbol{T}^7_{1,\mathrm{n}}\right]_{1,\mathrm{D}}$ for the wavelet coefficients $\boldsymbol{D}_1[n]$ of the first level, $\left[\boldsymbol{T}^2_{2,\mathrm{n}}\right]_{2,\mathrm{A}} \cdots \left[\boldsymbol{T}^7_{2,\mathrm{n}}\right]_{2,\mathrm{A}}$ for the scale coefficients $\boldsymbol{C}_2[n]$ of the second level. $\left[\boldsymbol{T}^2_{2,\mathrm{n}}\right]_{2,\mathrm{D}} \cdots \left[\boldsymbol{T}^7_{2,\mathrm{n}}\right]_{2,\mathrm{D}}$ for the wavelet coefficients $\boldsymbol{D}_2[n]$ of the second level, $\left[\boldsymbol{T}^2_{3,\mathrm{n}}\right]_{3,\mathrm{A}} \cdots \left[\boldsymbol{T}^7_{3,\mathrm{n}}\right]_{3,\mathrm{A}}$ for the scale coefficients $\boldsymbol{C}_3[n]$ of the third level, and $\left[\boldsymbol{T}^2_{3,\mathrm{n}}\right]_{3,\boldsymbol{D}} \cdots \left[\boldsymbol{T}^7_{3,\mathrm{n}}\right]_{3,\mathrm{D}}$ for $\boldsymbol{D}_3[n]$. Here there were thirty-six features chosen in the input feature set.

The general dimensionless moment function is defined as:

$$\boldsymbol{F}^{\mathrm{p}}_{\mathrm{m}} = \left\langle \boldsymbol{T}^{\mathrm{p}}_{\mathrm{m,n}} \right\rangle \Big/ \left[\left\langle \boldsymbol{T}^2_{\mathrm{m,n}} \right\rangle_{\mathrm{m}} \right]^{\mathrm{p}/2} \tag{12}$$

where the p^{th} order moment was normalized by the re-scaled variance. The scale-dependent coefficient skewness factor was defined as the 3rd order moment $\boldsymbol{F}^3_{\mathrm{m}}$. The scale–dependent coefficient flatness factor was defined as the 4th order moment $\boldsymbol{F}^4_{\mathrm{m}}$. The 3rd order to 6th order statistical moments of the detail coefficients and approximation coefficients at the first and second level were used as the feature values for the classification. They are $\boldsymbol{F}^3_{1,\mathrm{D}} \cdots \boldsymbol{F}^6_{1,\mathrm{D}}$ for the detail coefficients of the first level, $\boldsymbol{F}^3_{1,\mathrm{A}} \cdots \boldsymbol{F}^6_{1,\mathrm{A}}$ for the approximation coefficients of the first level, $\boldsymbol{F}^3_{2,\mathrm{D}} \cdots \boldsymbol{F}^6_{2,\mathrm{D}}$ for the detail coefficients of the second level, and $\boldsymbol{F}^3_{2,\mathrm{A}} \cdots \boldsymbol{F}^6_{2,\mathrm{A}}$ for the approximation coefficients of the second level. At the third level, the 3rd order to 6th order statistical moments of the approximation coefficients were used. They are the $\boldsymbol{F}^3_{3,\mathrm{A}} \cdots \boldsymbol{F}^6_{3,\mathrm{A}}$ for the approximation coefficients of the third level. Here there were twenty features chosen in the input feature set.

In Section 3, the intermittency index was used to analyze the vibration signals. The mean $\boldsymbol{M}_{\mathrm{I}}$, variance $\boldsymbol{V}_{\mathrm{I}}$ and sum $\boldsymbol{S}_{\mathrm{I}}$ of the intermittency index were used as the feature values in the input feature set for the classification.

In the sum, there are 59 feature values to be input for the optimization and classification by the SVMs as discussed in the following section.

4 Support Vector Machine

In the late 1960s, SVMs were introduced (Vapnik, 1995), based on the foundation of statistical learning theory. The algorithm for the numerous practical applications by SVMs emerged following greater availability of computing power from the middle of 1990s (Schlkopf, 1998; Burges, 1998; Gunn, 1998; Scholkopf, 1990). SVM for separating two classes is used to create a line or hyperplane, which is defined by support vectors. The margin between both planes is maximized. The SVMs orient the boundary such that the distance between the boundary and the nearest data point in the class is maximal. The line or boundary is in the middle of the margin between two kinds of data. The nearest data points are known as support vectors.

There are two finite subsets of vectors from a train-

ing set,

$$(\boldsymbol{x}_1, \boldsymbol{y}_1), (\boldsymbol{x}_2, \boldsymbol{y}_2), \cdots (\boldsymbol{x}_\mathrm{L}, \boldsymbol{y}_\mathrm{L}), \tag{13}$$

where $\boldsymbol{x}_\mathrm{i} \in \boldsymbol{R}^\mathrm{n}$ and $\boldsymbol{y}_\mathrm{i} \in \{-1,1\}$, for $\boldsymbol{i} = 1,2,...,\boldsymbol{N}$.

One subset I for which $\boldsymbol{y} = 1$, and another subset Π for which $\boldsymbol{y} = -1$ are separable by the hyperplane $(\boldsymbol{x} * \varphi) = \boldsymbol{c}$, where φ is one vector. A unit vector $\varphi(|\varphi| = 1)$ and a constant $\boldsymbol{c}$ exist. So the following inequalities hold true.

$$(\boldsymbol{x}_\mathrm{i} * \varphi) > \boldsymbol{c} \quad \text{if} \quad \boldsymbol{x}_\mathrm{i} \in \mathrm{I} \tag{14}$$

$$(\boldsymbol{x}_\mathrm{j} * \varphi) < \boldsymbol{c} \quad \text{if} \quad \boldsymbol{x}_\mathrm{j} \in \Pi \tag{15}$$

where $\boldsymbol{x} * \varphi$ is the inner product between vectors $\boldsymbol{x}$ and φ.

The vector φ_0 and the constant $\boldsymbol{c}_0$ exist and determine the hyperplane. This hyperplane is called the maximal margin hyperplane or the optimal hyperplane, as shown in Fig. 8.

The following equations can be obtained:

Maximizing

$$\boldsymbol{L}_\mathrm{d}(\alpha) = -0.5\alpha^\mathrm{T} \boldsymbol{H}\alpha + \boldsymbol{f}^\mathrm{T}\alpha \tag{16}$$

subject to

$$\begin{aligned} &\boldsymbol{y}^\mathrm{T}\alpha = 0 \\ &\alpha \geq 0 \end{aligned} \tag{17}$$

where $(\alpha)_\mathrm{i} = \alpha_\mathrm{i}$, H denotes the Hessian matrix $\left(\boldsymbol{H}_\mathrm{ij} = \boldsymbol{y}_\mathrm{i}\boldsymbol{y}_\mathrm{j}(\boldsymbol{x}_\mathrm{i}\boldsymbol{x}_\mathrm{j}) = \boldsymbol{y}_\mathrm{i}\boldsymbol{y}_\mathrm{j}\boldsymbol{x}_\mathrm{i}^\mathrm{T}\boldsymbol{x}_\mathrm{j}\right)$, and $\boldsymbol{f}$ is a unit vector $(\boldsymbol{f} = [1 \quad 1 \quad 1 \quad 1])$.

The parameters ζ_0 and $\boldsymbol{b}_0$ of the optimal hyperplane are obtained as:

$$\zeta_0 = \sum_{\mathrm{i}=1}^{L} \alpha_{0\mathrm{i}} y_\mathrm{i} x_\mathrm{i}, \qquad \mathrm{i} = 1, L \tag{18}$$

$$\boldsymbol{b}_0 = \frac{1}{\boldsymbol{N}_\mathrm{sv}} \left(\sum_{\mathrm{s}=1}^{\boldsymbol{N}_\mathrm{sv}} \left(\frac{1}{\boldsymbol{y}_\mathrm{s}} - \boldsymbol{x}_\mathrm{s}^\mathrm{T}\zeta_0\right)\right), \qquad \boldsymbol{s} = 1, \boldsymbol{N}_\mathrm{sv} \tag{19}$$

where $\boldsymbol{N}_\mathrm{sv}$ is the number of the support vectors and L is the number of the training data. The optimal weights vector ζ_0 and the bias term $\boldsymbol{b}_0$ are calculated by support vectors.

The decision hyperplane $\boldsymbol{D}(\boldsymbol{x})$ and the indicator function $\boldsymbol{i}_\mathrm{F}$ are defined as:

$$\begin{aligned} &\boldsymbol{D}(\boldsymbol{x}) = \sum_{\mathrm{i}=1}^{L} y_\mathrm{i}\alpha_\mathrm{i}\boldsymbol{x}^\mathrm{T}\boldsymbol{x}_\mathrm{i} + \boldsymbol{b}_0 \\ &i_\mathrm{F} = \mathrm{sign}(\boldsymbol{D}(\boldsymbol{x})) = \mathrm{sign}\left[\sum_{s=1}^{l} y_i\alpha_i \boldsymbol{K}(\boldsymbol{x}, \boldsymbol{x}_s) + \boldsymbol{b}_k\right] \end{aligned} \tag{20}$$

Where $\boldsymbol{k}(\boldsymbol{x}, \boldsymbol{x}_s)$ was Kernel function. Kernel function maps the training data nonlinearly into a higher-dimensional feature space, and constructs a separating hyperplane with a maximum margin. It yielded a nonlinear decision boundary in input space. It was possible to compute the separating hyperplane without explicitly carrying out the map into the feature space. It was used to map the data into some other dot product space (called the feature space) via a nonlinear map. It required the evaluation of dot products, that was

$$k(x, y) := (\Phi(x) \cdot \Phi(y)) \tag{21}$$

In some cases, however, there was a simple *kernel k* that can be evaluated efficiently. The polynomial kernel can be shown to correspond to the space spanned by all products of exactly *d* dimensions of $\Re^N$.

$$k(x, y) = (x \cdot y)^d \tag{22}$$

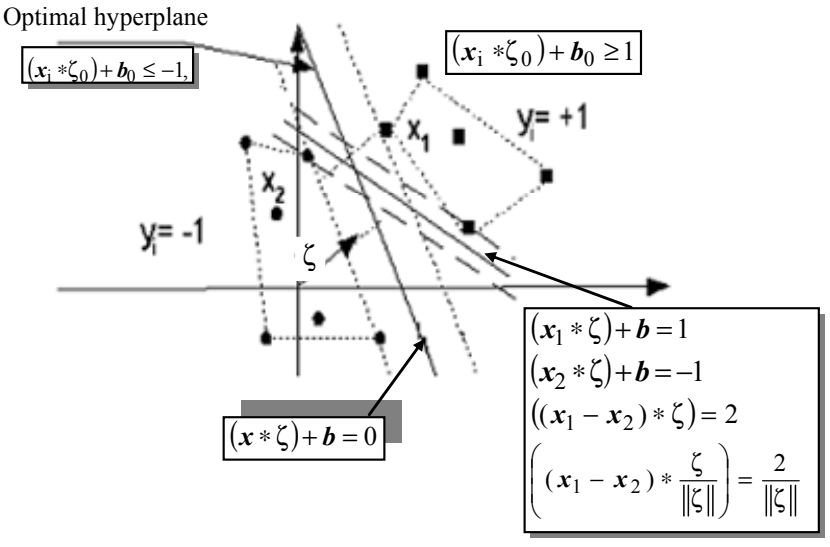


Fig. 8: *The optimal separating hyperplane separates the data with the maximal margin*

When $\boldsymbol{d}=1$, the kernel function $\boldsymbol{k}(\boldsymbol{x},\boldsymbol{y})$ was linear. When $\boldsymbol{d}\geq 2$ and $\boldsymbol{x},\boldsymbol{y}\in\Re^2$, the kernel function $\boldsymbol{k}(\boldsymbol{x},\boldsymbol{y})$ was nonlinear. In the work, the parameter $\boldsymbol{d}$ equalled to 1 and 2 for linear and nonlinear kernel function respectively. The following equation was derived as:

$$(x\cdot y)^2=\left(\begin{pmatrix}x_1\\x_2\end{pmatrix}\cdot\begin{pmatrix}y_1\\y_2\end{pmatrix}\right)^2=\left(\begin{pmatrix}x_1^2\\\sqrt{2}x_1x_2\\x_2^2\end{pmatrix}\cdot\begin{pmatrix}y_1^2\\\sqrt{2}y_1y_2\\y_2^2\end{pmatrix}\right) \quad (23)$$
$$=\left(\Phi(x)\cdot\Phi(y)\right)$$

5 SVM-RFE for Feature Optimization

A known challenge in the classification and machine learning in general is to find ways to reduce the dimensionality of the feature space to overcome the risk of "overfitting". Data overfitting arises when the number of features is large and the number of training patterns is comparatively small. A decision function can be easily found to separate the training data but will perform poorly on the test data. The new pruning method is investigated to eliminate some of the original input features and retain a minimum subset of the features that yield best classification performance. The method using the ranking features with the magnitude of the weights in a linear discriminant classifier is important in optimizing the features. However a good feature ranking criterion is not necessarily a good ranking criterion for subsets. The criteria overcome the effect of removing one feature at a time on the objective function. They become the sub-optimal set when it removes several features at a time. The problem can be solved by using the following iterative procedure, which is Recursive Feature Elimination (RFE).

SVMs can handle the linear or non-linear decision boundaries of the arbitrary complexity. The linear SVMs are particular linear discriminant classifiers. It can extend to non-linear discriminant classifiers. The training data set is linearly separable and the linear SVMs is a maximum margin classifier. The decision boundary is positioned to leave the largest possible margin on either side. The SVMs and SVMs-RFE algorithm was programmed in Matlab. The SVMs algorithm procedure was described as follow:

(a) Input the training examples $\{\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots,\boldsymbol{X}_{\boldsymbol{k}},\cdots,\boldsymbol{X}_{\boldsymbol{l}}\}$ and class labels $\{\boldsymbol{Y}_1,\boldsymbol{Y}_2,\cdots,\boldsymbol{Y}_{\boldsymbol{k}},\cdots\boldsymbol{Y}_{\boldsymbol{l}}\}$.

(b) Maximize Eq. 16 subject to Eq. 17 and output the optimal coefficients α_{k}.

(c) The optimization of SVMs is typical of a quadratic program. Genetic Algorithm (GA) is applied to optimize the SVMs to obtain the optimal parameters. GA is described in the following section.

(d) Calculate the parameters ζ_0 and $\boldsymbol{b}_0$ of the optimal hyperplane according to Eq. 18 and 19, output the decision function by Eq. 20.

The new ranking criterion was proposed in SVM-RFE according to the data characteristics.

The SVMs-RFE algorithm procedure was:

(a) Input the training data $\{\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots,\boldsymbol{X}_{\mathrm{k}},\cdots,\boldsymbol{X}_{\mathrm{L}}\}$ and class labels $\{\boldsymbol{Y}_1,\boldsymbol{Y}_2,\cdots,\boldsymbol{Y}_{\mathrm{k}},\cdots\boldsymbol{Y}_{\mathrm{L}}\}$. Initialize the subset of the surviving features and feature ranked list.

(b) Training the classifier $\alpha=\boldsymbol{SVM}-\mathrm{train}(\boldsymbol{X},\boldsymbol{Y})$. Compute the weight vector ζ

(c) Compute the ranking criteria $\boldsymbol{c}_{\mathrm{i}}=1/(\zeta_{\mathrm{i}})^2$ for all $\boldsymbol{i}$.

(d) Find the feature with the smallest ranking criteria. Update the feature ranked list and eliminate the feature with the smallest ranking criterion.

(e) Output the ranked list and repeat steps (b) to (e).

6 Genetic Algorithm

Genetic algorithms (GAs) are applied in a wide variety of areas with increasing interest (Goldberg, 1999). GA is used to search the optimal solution space through simulated evolution of 'survival of the fittest' that is analogy with natural behavior of species. The GA considers six basic issues: chromosome representation, selection function, genetic operator such as crossover and mutation for reproduction function, creation of initial population, termination criteria and the evaluation function. The real value encoding is applied in the chromosome representation because of the closeness of the representation to the problem space, better average performance and more efficient numerical implementation. The chromosome (C) contains the real value of the features, which represents the coefficients in Eq. 16 and 17. Every chromosome represents a solution to the problem. The fitness of each chromosome can be calculated and determined by the objective function. The reproduction operator is implemented by using a roulette wheel with slots sized according to fitness. The chromosomes with higher fitness are given opportunities to reproduce by mating with other individuals in the population. The chromosomes with smaller fitness are less likely to get selected for reproduction and gradually disappear. In the cross phase, strings are chosen by pairs. For each random pair, two random numbers are generated to decide which pieces of the strings are to be interchanged. The crossover operator is applied by probability p_{c}. In the mutation phase, the bits in the string in the population will be chosen with probability p_{m}.

The real number encoding, fitness scaling, stochastic universal selection, single-point crossover and non-uniform mutation are employed for the optimization of parameter in SVM. The size of the population is 60. GA is used to optimize the parameter α in SVM. So the length of each string equals to the number of the training data. The number of the string in a population equals to the number of testing data, which is 119. The simulation runs are 2000. The crossover rate is set to the typical values of 0.6. It is typical to use the relatively higher mutation value such as 0.14 when the real number encoding is used. The population can converge

more rapidly by increasing the crossover and mutation rate to introduce more variation into the population. The mutation and crossover rates are set to be 0.2633 and 0.6, respectively, to obtain better classification success rate and increase the speed of convergence for this experiment as shown in Fig. 9.

Figure 9 shows the relation between the generation and the average fitness and optimal fitness. Here GA is used to optimize the parameters in SVM, which is simple linear optimization to classify two classes of data. The classification results also deeply depend on the clusters of feature values in two classes. When the clusters of feature values are better, the convergence speed of classification is shorter. Therefore, based on the good feature values and linear SVM with good parameters in GA, the speed of GA convergence and SVM classification are faster, which is especially suitable for on-line machinery fault diagnosis.

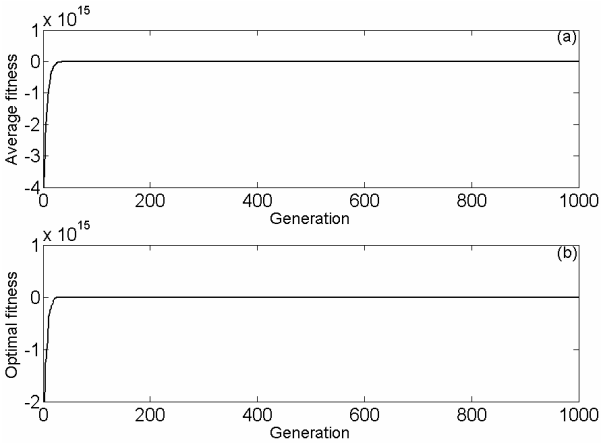


Fig. 9: *The relation between the fitness and generation (a) Average fitness, (b) Optimal fitness*

7 Results and Discussion

The extension from the binary two-class to multi-class is an important question for the support vector machine approach. The winner-takes-all scheme is proposed for multi-class classification. The two-class decision function is extended to multi-class by constructing binary decision functions for all classes as follows:

$$\boldsymbol{f}_{\mathrm{k}} : \boldsymbol{R}^{\mathrm{N}} \rightarrow \{\pm 1\} \qquad \begin{cases} +1 \\ -1 \end{cases} \tag{24}$$

Figure 10 shows the winner-takes-all scheme for the three classes. The three lines can separate one class from the both others correctly. However in the middle of the picture all the decision functions suggest that this region belongs to none of the three classes. Similar question arises for the three outer triangles with the shaded region that in each case where two decision functions are $+1$ simultaneously.

To solve the problem, the tie situation is broken to neglect the sign operation in the decision functions as shown in equation 20 and to use the real input values. The classification results can be viewed as the discriminant vector where the largest component is chosen as the class decision. The tie breaking is shown in Fig. 11.

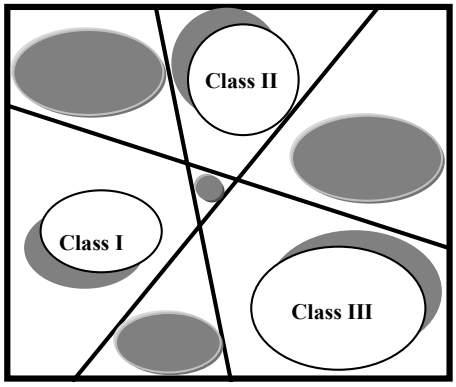


Fig. 10: *Winner-takes-all for three classes*

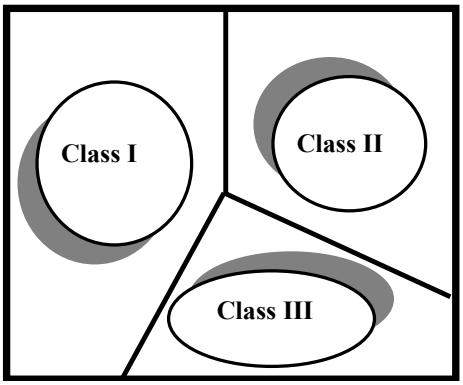


Fig. 11: *Tie breaking by "winner-takes-all" for three classes*

In order to classify the five piston conditions, the two-level classification structure by the multi-classification is proposed. Figure 12 shows the structure of the two-level classification. In the first level, the multi-classification is applied to classify the BLC, BSC and other condition. In the second level, the FLC, FSC and NC are classified.

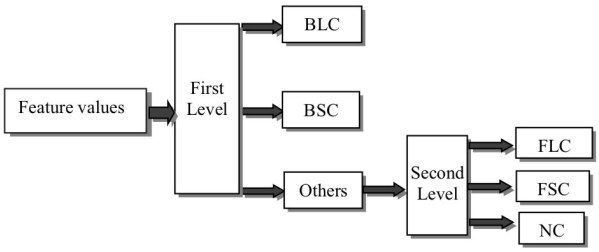


Fig. 12: *The structure of the two-level classification based on "one-against-all" multi-classification*

There are 47 experimental vibration signals of the water hydraulic motor under BSC and 48 vibration signals for BLC, FLC, FSC and NC operating conditions. The training data consists of 23 vibration signals under BSC and 24 vibration signals under BLC, FLC, FSC and NC operating conditions. The testing data consists of another 24 vibration signals under BSC and another 96 vibration signals under BLC, FLC, FSC and NC operating conditions.

As shown in section, there are 59 feature values to be used for classification, which are $\left[\boldsymbol{T}_{1,\mathrm{n}}^{2}\right]_{1,\mathrm{A}} \cdots \left[\boldsymbol{T}_{1,\mathrm{n}}^{7}\right]_{1,\mathrm{A}}$, $\left[\boldsymbol{T}_{1,\mathrm{n}}^{2}\right]_{1,\mathrm{D}} \cdots \left[\boldsymbol{T}_{1,\mathrm{n}}^{7}\right]_{1,\mathrm{D}}$,

$[T^2_{2,n}]_{2,A} \cdots [T^7_{2,n}]_{2,A}$, $[T^2_{2,n}]_{2,D} \cdots [T^7_{2,n}]_{2,D}$, $[T^2_{3,n}]_{3,A}$ $[T^2_{3,n}]_{3,A} \cdots [T^7_{3,n}]_{3,A}$, $[T^2_{3,n}]_{3,D} \cdots [T^7_{3,n}]_{3,D}$, $F^3_{1,D} \cdots F^6_{1,D}$, $F^3_{1,A} \cdots F^6_{1,A}$, $F^3_{2,D} \cdots F^6_{2,D}$, $F^3_{2,A} \cdots F^6_{2,A}$, $F^3_{3,A} \cdots F^6_{3,A}$, M_I, V_I, and S_I. The corresponding input item number of the classifier is from 1 to 59. These feature values are optimized by the SVMs-RFE algorithm and two-level classification based on "one-against-all" multi-classification in Fig. 12.

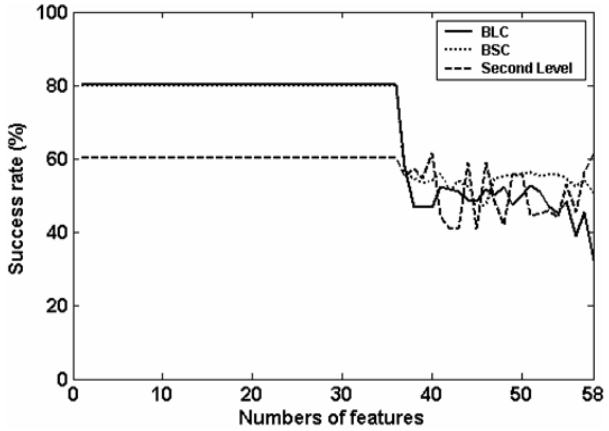


Fig. 13: *The feature value optimization at first level*

Figure 13 shows the optimization results of the feature values for BLC, BSC and others by SVMs-RFE, which were shown at first level in Fig. 12. The *x* axis represents the number of the features with the optimization algorithm. The *y* axis represents the classification success rate. With the process of the SVMs-RFE, the number of the feature values decreased and the success rate of the optimal feature values increased. The first 36 feature values were selected as the optimal features and its success rate was optimal. Table 1 shows the 36 optimal features at first level of Fig. 12.

Figure 14 shows the optimization results of the feature values for FLC, FSC and NC by SVMs-EFE, which were shown at second level in Fig. 12. The first 36 feature values were selected as the optimal feature values. The success rate was optimal and constant. Table 2 shows the 36 optimal features at second level of Fig. 12.

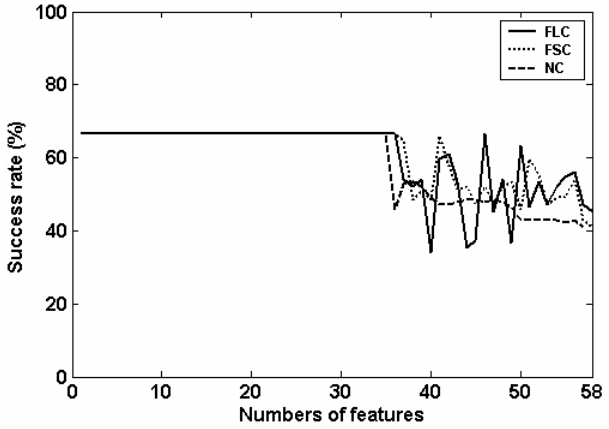


Fig. 14: *The feature value optimization at second level*

Table 3 shows the classification success rate by the multi-classification success rate by two level multi-classification structure in Fig. 12. The linear and nonlinear classifiers as described in section 4 were used to detect the piston conditions. The "winner-takes-all" method by tie breaking was also used for the nonlinear classifier to improve the classification success rate.

At first level, the classification success rates of the training data by the linear and nonlinear kernel of the multi-classification under BLC, BSC and Other (FLC, BSC, NC) piston conditions were 80.7%, 79.8%, 60.5% and 80.67%, 79.8%, 60.5%. The classification success rates of the training data by the nonlinear kernel of Winner-takes-all (Tie-breaking) under BLC, BSC and Other (FLC, BSC, NC) were 80.7%, 79.8%, 60.5%. Compared with the classification success rate of the training data, the classification success rates of the testing data by the linear and nonlinear kernel of the multi-classification under BLC, BSC and Other (FLC, BSC, NC) piston conditions were 80%, 80%, 60% and 80%, 80%, 60%. The classification success rates of the training data by the nonlinear kernel of Winner-takes-all (Tie-breaking) under BLC, BSC and Other (FLC, BSC, NC) were 80%, 80%, 60%.

At second level, the classification success rates of the training data by the linear and nonlinear kernel of the multi-classification under FLC, FSC and NC piston conditions were 66.7%, 66.7%, 66.7% and 66.7%, 66.7%, 66.7%. The classification success rates of the training data by the nonlinear kernel of Winner-takes-all (Tie-breaking) under BLC, BSC and NC were 72.2%, 69.4%, 72.2%.

The classification success rates of the testing data by the linear and nonlinear kernel of the multi-classification under BLC, BSC and NC piston conditions were 66.7%, 66.7%, 66.7% and 66.7%, 66.7%, 66.7%. The classification success rates of the training data by the nonlinear kernel of Winner-takes-all (Tie-breaking) under BLC, BSC and Other (FLC, BSC, NC) were 66.7%, 33.3%, 66.7%.

The classification success rate showed that the classification of the five piston conditions was successful. The classification success rate of the testing data was as good as the training data under the all the piston conditions and different classifiers. The classification success rate of tie-breaking was better than the multi-classification without tie-breaking except the testing data under FSC piston condition. This is because the classification success rate of the piston conditions has a close relation to the input feature values.

Table 1: The 36 optimal features at first level

Piston condition (One-Against-All)	The corresponding item number of the 36 optimal features
BLC	1 2 3 7 8 13 14 15 19 20 25 26 31 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59
BSC	1 2 3 7 8 9 13 14 16 19 20 25 26 31 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58
Second level (FLC, FSC and NC)	1 2 3 7 8 9 13 14 15 16 19 25 26 31 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58

Table 2: The 36 optimal features at second level

Piston condition (One-Against-All)	The corresponding item number of the 36 optimal features
FLC	1 2 3 7 8 9 13 14 15 19 20 25 31 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59
FSC	1 2 3 7 8 9 13 14 15 16 19 20 25 26 31 32 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56
NC	1 2 3 7 8 9 13 14 15 19 20 25 31 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59

Table 3: The results of the two-level classification success rate

		Piston condition	Multi-classification		Winner-takes-all (Tie breaking)
			Linear	Non-linear	Non-linear
First level	Training data	BLC	80.7	80.67	80.7
		BSC	79.8	79.8	79.8
		Others	60.5	60.5	60.5
	Testing data	BLC	80	80	80
		BSC	80	80	80
		Others	60	60	60
Second level	Training data	FLC	66.7	66.7	72.2
		FSC	66.7	66.7	69.4
		NC	66.7	66.7	72.2
	Testing data	FLC	66.7	66.7	66.7
		FSC	66.7	66.7	33.3
		NC	66.7	66.7	66.7

8 Conclusion

The second-generation wavelet was proposed here as a novel method for the feature extraction from the vibration signals of the water hydraulic motor. The statistical feature values of the vibration signals were used as the features for the classification of the mechanical condition. Due to the large number of feature values, the features should be optimized to extract the optimal features. The new optimization method by bi-classification SVM to extract the optimal features was developed in this work. The new rank criterion was proposed to reduce the dimension of the feature values and obtain the optimal feature value for the fault diagnosis of water hydraulic motor. The optimal 36 features were obtained by the proposed SVMs-RFE algorithm. The optimal feature values were selected successfully. In order to classify the five piston conditions, the two-level structure based on the multi-classification was developed. The multi-classification method of SVM for the fault diagnosis of the piston crack was investigated. The winner-takes-all scheme was studied. At the first level, the success rate of the classification for BLC, BSC, and others were 80%, 80% and 60%. At second level, the success rate of the classification for FLC, FSC and NC were 66.7%, 66.7% and 66.7%. The results of the classification showed the applicability of the proposed method.

Nomenclature

BLC	crack is 10 mm long and located in the back section of the piston
BSC	crack is 5 mm long and located in back section of the piston
FLC	crack is 10 mm long and located in the front section of the piston
FSC	crack is 5 mm long and located in the front section of the piston
P	predict operator
U	update operator
$\boldsymbol{P}(z), \widetilde{\boldsymbol{P}}(z)$	polyphase matrix
$\boldsymbol{s}(z^2), \widetilde{\boldsymbol{s}}(z^2)$	Laurent polynomial
$\boldsymbol{g}(z), \widetilde{\boldsymbol{g}}(z)$	high-pass filter
$\boldsymbol{h}(z), \widetilde{\boldsymbol{h}}(z)$	low-pass filter.
$\boldsymbol{T}_{\mathrm{m,n}}$	wavelet coefficients at scale index m.
$\boldsymbol{x}_{\mathrm{e}}[n]$	even indexed point
$\boldsymbol{x}_{\mathrm{o}}[n]$	odd indexed point
$\boldsymbol{d}[n]$	wavelet coefficient
$\boldsymbol{c}_{\mathrm{j}}[n]$	discrete wavelet transform scaling at each scale $\boldsymbol{j}$.
$\boldsymbol{d}_{\mathrm{j}}[n]$	discrete wavelet transform wavelet coefficients at each scale $\boldsymbol{j}$.
$\boldsymbol{T}_{\mathrm{m,n}}$	discrete wavelet transform coefficient
$\left[\boldsymbol{T}_{\mathrm{m,n}}^{\mathrm{p}}\right]_{\boldsymbol{m}}$	p^{th} order statistical moment of coefficients $\boldsymbol{T}_{\mathrm{m,n}}$ at scale index m
$\left[\boldsymbol{T}_{1,\mathrm{n}}^{2}\right]_{1,\mathrm{A}} \cdots \left[\boldsymbol{T}_{1,\mathrm{n}}^{7}\right]_{1,\mathrm{A}}$	2nd to 7th order statistical moment of scale coefficients $\boldsymbol{T}_{\mathrm{m,n}}$ at first level
$\left[\boldsymbol{T}_{1,\mathrm{n}}^{2}\right]_{1,\mathrm{D}} \cdots \left[\boldsymbol{T}_{1,\mathrm{n}}^{7}\right]_{1,\mathrm{D}}$	2nd to 7th order statistical moment of wavelet coefficients $\boldsymbol{T}_{\mathrm{m,n}}$ at first level
$\left[\boldsymbol{T}_{2,\mathrm{n}}^{2}\right]_{2,\mathrm{A}} \cdots \left[\boldsymbol{T}_{2,\mathrm{n}}^{7}\right]_{2,\mathrm{A}}$	2nd to 7th order statistical moment of scale coefficients $\boldsymbol{T}_{\mathrm{m,n}}$ at second level
$\left[\boldsymbol{T}_{2,\mathrm{n}}^{2}\right]_{2,\mathrm{D}} \cdots \left[\boldsymbol{T}_{2,\mathrm{n}}^{7}\right]_{2,\mathrm{D}}$	2nd to 7th order statistical moment of wavelet coefficients $\boldsymbol{T}_{\mathrm{m,n}}$ at second level
$\left[\boldsymbol{T}_{3,\mathrm{n}}^{2}\right]_{3,\mathrm{A}} \cdots \left[\boldsymbol{T}_{3,\mathrm{n}}^{7}\right]_{3,\mathrm{A}}$	2nd to 7th order statistical moment of scale coefficients $\boldsymbol{T}_{\mathrm{m,n}}$ at third level
$\left[\boldsymbol{T}_{3,\mathrm{n}}^{2}\right]_{3,\boldsymbol{D}} \cdots \left[\boldsymbol{T}_{3,\mathrm{n}}^{7}\right]_{3,\mathrm{D}}$	2nd to 7th order statistical moment of wavelet coefficients $\boldsymbol{T}_{\mathrm{m,n}}$ at third level
$\boldsymbol{F}_{\mathrm{m}}^{\mathrm{p}}$	general dimensionless moment function
$\boldsymbol{F}_{1,\mathrm{D}}^{3} \cdots \boldsymbol{F}_{1,\mathrm{D}}^{6}$	3rd to 6th order general dimensionless moments of detail coefficients at first level
$\boldsymbol{F}_{1,\mathrm{A}}^{3} \cdots \boldsymbol{F}_{1,\mathrm{A}}^{6}$	3rd to 6th order general dimensionless moments of approximation coefficients at first level
$\boldsymbol{F}_{2,\mathrm{D}}^{3} \cdots \boldsymbol{F}_{2,\mathrm{D}}^{6}$	3rd to 6th order general dimensionless moments of detail coefficients at second level
$\boldsymbol{F}_{2,\mathrm{A}}^{3} \cdots \boldsymbol{F}_{2,\mathrm{A}}^{6}$	3rd to 6th order general dimensionless moments of approximation coefficients at second level
$\boldsymbol{F}_{3,\mathrm{A}}^{3} \cdots \boldsymbol{F}_{3,\mathrm{A}}^{6}$	3rd to 6th order general dimensionless moments of approximation coefficients at third level
$\boldsymbol{M}_{\mathrm{I}}, \boldsymbol{V}_{\mathrm{I}}, \boldsymbol{S}_{\mathrm{I}}$	mean, variance and sum of the intermittency index
$\boldsymbol{F}_{\mathrm{m}}^{3}$	skewness factor
$\boldsymbol{F}_{\mathrm{m}}^{4}$	flatness factor
$\boldsymbol{I}_{\mathrm{m,n}}$	intermittency index
φ	unit vector
$\boldsymbol{c}$	constant
$\boldsymbol{H}$	Hessian matrix
$\boldsymbol{N}_{\mathrm{SV}}$	number of the support vectors
$\boldsymbol{l}$	number of the training data in support vectors
ζ_0	optimal weights vector in support vectors
$\boldsymbol{b}_0$	bias term in support vectors
$\boldsymbol{D}(\boldsymbol{x})$	decision hyperplane
$\boldsymbol{i}_{\mathrm{F}}$	indicator function
$\boldsymbol{k}(\boldsymbol{x}, \boldsymbol{x}_s)$	Kernel function
RFE	Recursive Feature Elimination
α	classifier in SVMs
ζ	weight vector in SVMs
$\boldsymbol{c}_{\mathrm{i}}$	ranking criteria in SVMs-RFE
$\boldsymbol{p}_c$	crossover probability in GA
$\boldsymbol{p}_m$	mutation probability in GA

Acknowledgements

The authors wish to thank the School of Mechanical and Aerospace Engineering at Nanyang Technological University, Singapore, for providing the research funding and technical support for this research.

References

Burges, C. J. C., 1998. A tutorial on support vector machines for pattern recognition, *Data Mining and Knowledge Discovery 2,* pp. 955-974.

Chen, H. X., **Chua, Patrick. S. K.**, and **Lim, G. H.**, 2006. Dynamic Vibration Analysis of A Swash-plate Type Water Hydraulic Motor, *Mechanism and Machine Theory*, V41, n5, pp. 487-504.

Crowther, W. J., **Edge, K. A.**, **Burrows, C. R.**, **Atkinson, R. M.**, and **Woollons, D. J.**, 1998. Fault diagnosis of a hydraulic actuator circuit using neural networks, an output vector space classification approach, *Proceedings of the I MECH E Part I Journal of Systems & Control Engineering 212*, pp.57 - 68.

Daley, S. and **Wang, H.** 1996. Fault diagnosis in fluid power systems, *Engineering Simulation*, 13(6), pp.993-1008.

Danfoss, Inc. http://www.danfoss.com/Nessie/Introduction/index.asp, Denmark.

Daubechies, I. and **Sweldens, W.** 1998. Factoring wavelet transforms into lifting steps, *Journal of Fourier Analysis and Application*, pp. 41-44.

Goldberg, G. E. 1999. *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, NY, USA.

Gunn, S. R. 1998. Support vector machines for classification and regression, *Technical report, University of Southampton*, Department of Electrical and Computer Science.

Haney, L., **Precetti, C.**, **Gibson, H.**, **Krutz, G.** and **Cassens, D.** 1994. Color Matching of Wood With a Real-time Machine Vision System, *American Society of Agricultural Engineers*, Atlanta, Georgia, December, pp. 13-16.

Lim G. H., **He, Y. B.**, and **Chua, P. S. K.** 2002. Modern Water Hydraulics-The New Energy Transmission Technology in Fluid Power, *Proceedings of the 9th International Energy Conference and Exhibition (Energex' 02)*, Crakow, Poland.

Ivantysyn, J. and **Ivantysynova, M.** 2002. Hydrostatic Pumps and Motors, *Technip Books International*, New Delhi-110 019, India, ISBN 81-8830508-1.

Ivantysynova, M., **Huang, C.** and **Behr, R.** 2005. Measusrement of elastohydrodynamic pressure field in the gap between piston and cylinder, *Power Transmission and Motion Control*, PTMC 2005, pp. 451-465.

Ivantysnova, M., and **Lasaar, R.** 2004. An investigation into micro-and macrogeometric design of piston/cylinder assembly of swash plate machine, *International Journal of Fluid Power*, V5, n1, March, pp. 23-36.

Krutz, G. W., and **Chua, Patrick S. K.** 2004. Water Hydraulics-Theory and Applications, *Workshop on Water Hydraulics*, Agricultural Equipment Technology Conference (AETC'04).

Le, T. T., **Watton, J.**, and **Pham, D. T.** 1998. Fault classification of fluid power systems using a dynamics feature extraction technique and neural networks, *Proceedings of the I MECH E Part I Journal of Systems & Control Engineering 212*, pp. 87 – 97.

Lim, G. H., **Chen, H. X.** and **Chua, P. S. K.** 2003. Fault Diagnosis of Piston Crack in a Water Hydraulic Motor, *Proceedings of the International Conference and Exhibition on Instrumentation and Control*, PICS-ICEIC 2003, ISBN: 1656-2364, November, Manila, Philippines, pp. 27-29.

Lott, P. W. and **Griffiths, A. J.** 1997. Fault diagnosis of pumping machinery using artificial neural networks, *Proceedings of the I MECH E Part E Journal of Process Mechanical Engineering 211*, pp. 185-194.

Nandi, A. K., 2000. Advanced digital vibration signal processing for condition monitoring, *Proceedings of COMADEM*, Houston, TX, USA, pp.129-143.

Precetti, C. J. and **Krutz, G. W.** 1993. SPR: *Statistical pattern recognition software,* Copyright No. TX 3 653 808, PRF.

Samanta, B. and **Al-Balushi, K. R.** 2001. Use of time domain features for the neural network based fault diagnosis of a machine tool coolant system, *Proceedings of the I MECH E Part I Journal of Systems & Control Engineering 215*, pp. 199-207.

Schlkopf, B. 1998. SVMs-a proactical consequence of learning theory, *IEEE Intelligent Systems 13*, pp. 18-19.

Scholkopf, B. and **Smola, A. J.** 1990. *Learning with kernels, support vector machines, regularization, optimization, and beyond*, New York, MIT Press.

Simon, U. 1996. Water Hydraulic power- The technology matures (putting water to work), National Fluid Power Association Technical Paper Series, pp. 7-13.

Sorensen, P. 1999. News and trends by the industrial application of water hydraulics, *The Six Scandinavian International Conference on Fluid Power*, pp. 651-674.

Stewart, C. J. and **Watton, J.** 1994. Some aspects of rule-based expert systems design for on-line fault diagnosis of fluid power systems, American Society of Mechanical Engineers, Petroleum Division (Publication) PD, *Design: Analysis, Synthesis, and Application* 4(8-2), pp.457-461.

Sweldens, W. 1996. The lifting scheme: A custom-design construction of biorthogonal wavelets, *Applied and Computational Harmonic Analysis* 3(2), pp. 186-200.

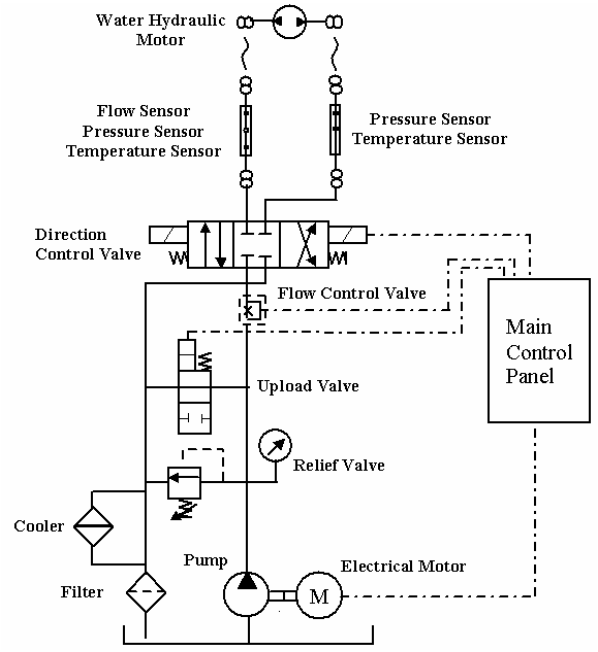
Trostmann, E. and **Clausen, P. M.** 1995. Hydraulic components using tap water as pressure medium, *The Fourth Scandinavian International Conference on Fluid Power*, pp. 942-954.

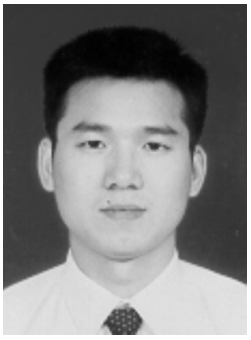
Vapnik, V. 1995. *The nature of statistical learning theory*, Springer, Berlin.

Watton, J., **Lucca-Negro, O.** and **Stewart, J. C.** 1994. On-line approach to fault diagnosis of fluid power cylinder drive systems, *Proceedings of the Institution of Mechanical Engineers. Part I, Journal of System & Control Engineering*, 208(14), pp.249-262.

Appendix A

Schematic hydraulic circuit of the modern water hydraulic motor test rig





H. X. Chen
He received the M. Eng. degree in the School of Mechanical Science and Technology, Huazhong University of Science & Technology, China. He currently finish oral defence of Ph.D. student and will be conferred PHD degree in Feb, 2006 at Nanyang Technological University, Singapore. His main research areas are in condition monitoring and fault diagnosis, signal processing, pattern recognition and its application to fluid power systems.

Patrick S.K. Chua (Dr.)
He worked as an assistant technical manager in a woodworking and machine tool company before joining National Semiconductor as lead R & D engineer and subsequently as R & D manager before joining Nanyang Technological University (NTU)as a lecturer. Presently as an Associate Professor at NTU, Singapore, his research interests are in the areas of fluid power technology, automated assembly, condition monitoring and biomedical engineering.

Lim Geok Hian (Dr.)
He is currently an Associate Professor at Nanyang Technological University, Singapore. His research interests are in the areas of mechanical vibration, condition monitoring and fluid power analysis. Prior to joining NTU, he had research experience at Sperry Vickers, IMI Norgren and British Gas, U.K. He received his Ph.D. from Aston University, U.K. and the B.Sc (Eng) and M. Eng. degrees from Imperial College of Science Technology and Medicine, London University.