Vibration-based fault diagnosis of slurry pump impellers using neighbourhood rough set models

X M Zhao[1], **Q H Hu**[2], **Y G Lei**[1], and **M J Zuo**[1]*
[1]Department of Mechanical Engineering, University of Alberta, Edmonton, Alberta, Canada
[2]School of Energy Science and Engineering, Harbin Institute of Technology, Harbin, People's Republic of China

The manuscript was received on 11 June 2009 and was accepted after revision for publication on 5 October 2009.

DOI: 10.1243/09544062JMES1777

Abstract: Rough set models have been widely used as a method for feature selection in fault diagnosis. A neighbourhood rough set model can deal with both nominal and numerical features, but selecting the neighbourhood size for its application may be a challenge. In this article, the authors illustrate that using a common neighbourhood size for all features may overestimate or underestimate a feature's dependency degree. The neighbourhood rough set model is then modified by setting different neighbourhood sizes for different features. The modified model is applied to the fault diagnosis of slurry pump impellers. Results show that the chosen feature subsets generated by the modified neighbourhood rough set model can be physically explained by the corresponding flow patterns and can achieve higher classification accuracy than the raw feature subsets and the feature subsets generated by the original neighbourhood rough set model.

Keywords: neighbourhood rough set, fault diagnosis, impeller damage, feature significance, neighbourhood size, noise level

1 INTRODUCTION

Slurry pumps have a wide range of industrial applications. As slurries contain abrasive and erosive solid particles, slurry pump impellers are subjected to severe wear. As a result, the pump performance is reduced or the pump system may even fail. Monitoring the condition of the impellers plays a key role in maintaining safe running pump systems, thereby contributing to an overall improvement in plant efficiency and reliability.

Very limited attention has been paid to the condition monitoring of the slurry pumps. Most reported studies on slurry pumps focus on the improvement of their design and understanding of wear mechanisms. Liu *et al.* [**1**] investigated the erosive wear of the impellers and liners of centrifugal slurry pumps. The eroded material surfaces of impellers and liners were studied by a scanning electron microscope. Khalid and Sapuan [**2**] fabricated a wear testing rig for a water pump impeller and selected a parameter to determine the wear degrees of the slurry pump impeller as a function of operating hours. The erosion was found to be the domain type of wear. The surface topography under the microscope indicated that wear occurred at both the region near the centre (leading edge) of the impeller and the region at the rim (trailing edge) of the impeller. The former encountered less wear compared with the latter.

Vibration analysis has been adopted in fault diagnosis of pump mechanical components such as bearing and sealing [**3**, **4**]. Wang and Hope [**5**] investigated the application of an artificial neural network into bearing fault diagnosis for pumps in waster water industry. Six bearing conditions were successfully classified by cascade-forward back-propagation neural networks with two hidden layers. Wang and Hu [**6**] applied the basic fuzzy logic principle to classify frequency spectra according to the likely fault conditions for a five-plunger pump and realized correct classification and condition recognition of different fault spectra

*Corresponding author: *Department of Mechanical Engineering, University of Alberta, 4-9 Mechanical Engineering Building, Edmonton, AB, T6G 2G8, Canada.*
email: ming.zuo@ualberta.ca

of pumps. Hardly any vibration analysis on impeller damage has been reported. Chudina [**7**] used noise spectra structure to detect the beginning and the development of cavitation which may cause damage to the impeller by pitting and erosion. However, slurry pumps, by design, are subjected to corrosive and erosive wear irrespective of whether or not cavitation occurs. And damages in impellers may occur at different parts of it, e.g. at the the trailing edge and the leading edge [**2**, **8**]. The detection of impeller damage, which is different from cavitation detection, is worthy of further study.

Distinguishing damaged impellers from undamaged impellers can be seen as a classification problem. Typically, a fault pattern is described using a set of features extracted from fault signals, some of which might be irrelevant or redundant. The feature set usually is high-dimensional. Dimension reduction, a technique commonly used in classification, is expected to be helpful in machine condition monitoring because of the following reasons. First, there is an associated cost with feature measurements. From the expense-saving point of view, it is desirable to reduce the number of features. Second, it is often observed that the accuracy of classifiers can be significantly improved using fewer features [**9**]. Third, it is easier for humans to interpret the induced concept when the number of features is smaller, since concepts that make use of many features are usually hard to understand.

Dimension reduction generally falls into two paradigms: semantics-destroying and semantics-preserving. The first one refers to constructing new features with a linear or non-linear transformation from the original feature set. Methods such as the principal component analysis, independent component analysis, and linear discriminant analysis belong to this paradigm. The main drawback of these methods is that the constructed features do not have true meanings. The second paradigm, also called feature selection, attempts to retain the meaning of the original feature set. According to the evaluation procedure, feature selection can be classified into two categories: the filter method and the wrapper method. If a classifier is used as the evaluation function, then it is called the wrapper method. However, the wrapper method usually has a high time-complexity, and may not be feasible in real-world applications. The filter method (e.g. relief-F [**10**], focus [**11**], and soap [**12**]) is independent of classifiers, and tends to be more applicable.

Rough set, launched by Pawlak [**13**], has been shown to be an effective filter-based tool for feature selection. Its main advantage is that it requires no additional parameters to operate other than the supplied data. However, the classical rough set model is best suited to nominal features that are discrete values. In condition monitoring, it is often the case that real values (i.e. numerical features) must be dealt with. For this reason, some researchers [**14**, **15**] introduced discretization algorithms to partition real-valued features into several intervals, and then treated them as nominal features. But this may cause information loss because the membership degrees of numerical values to discretized values are not considered. To overcome this problem, Jensen and Shen [**16**] presented a fuzzy-rough feature selection algorithm which generalized the dependency function defined in the classical rough set model into the fuzzy case. Generating a fuzzy-equivalence relation from numerical features is time consuming, and, however, is itself an open problem [**17**]. Another way of solving the problem of information loss in the discretization is by introducing the neighbourhood concept [**18**, **19**]. Hu *et al.* [**20**] proposed a neighbourhood rough set model for the heterogeneous feature (i.e. the nominal feature and the numerical feature coexist) selection. A feature evaluation function, called neighbourhood dependency, was used to evaluate the significance of a subset of features.

One problem with applying the neighbourhood rough set model in fault diagnosis is determining the neighbourhood size. In reference [**20**], the numerical features were treated as a vector, with which a distance was calculated and compared to the given neighbourhood size to determine the neighbourhood of a sample. From the measurement point of view, a neighbourhood size reflects a feature's tolerance level to noise, and thus can be regarded as a measurement of noise level [**21**]. However, noise levels are not the same for features measured by different accelerometers. Usually it is difficult to find a common neighbourhood value for all features.

In this article, the authors demonstrate that using one neighbourhood size for all features may overestimate or underestimate a feature's dependency degree. The neighbourhood rough set model is then modified by setting different neighbourhood sizes for different features. The modified model is applied to feature selection in the fault diagnosis of slurry pump impellers. Comparisons are conducted with the raw feature sets and feature subsets selected by the original neighbourhood rough set model. Furthermore, the feature subsets generated by the modified neighbourhood rough set model are explained by the corresponding flow patterns. The effects of accelerometer locations are also discussed.

The rest of the article is organized as follows. Section 2 introduces the original neighbourhood rough set model, discusses its disadvantage based on which modification is conducted. Section 3 describes the experimental set-up and the feature extraction from the vibration data in a slurry pump. Experimental results are shown and discussed in section 4. The conclusion is presented in section 5.

2 NEIGHBOURHOOD ROUGH SET MODELS

2.1 A neighbourhood rough set model with a common neighbourhood size

Let IS $= \langle U, A \rangle$ be an information system, where U is a non-empty and finite set of samples $\{x_1, x_2, \ldots, x_n\}$ called a universe, and A is a set of features $\{a_1, a_2, \ldots, a_f\}$, also called attributes, which characterize the samples. $\langle U, A \rangle$ is also called a decision table if $A = C \cup D$ where C is the set of condition features which describe the samples' characteristics, and D is the decision features which classify the samples' labels. Given $x_i \in U$, $B \subseteq C$, the neighbourhood of x_i in feature space B is defined as

$$\delta_B(x_i) = \{x_j | x_j \in U, \quad \Delta_B(x_i, x_j) \leqslant \delta\} \tag{1}$$

where δ is the neighbourhood size and Δ is a distance function. If B is a nominal feature set, then $\delta = 0$. The general distance function, called the Minkowsky distance, is defined as

$$\Delta_B(x_i, x_j) = \left[\sum_{k=1}^{n} (x_{ik} - x_{jk})^{\mathrm{P}} \right]^{1/p} \tag{2}$$

In the above equation, x_{ik} and x_{jk} are the values of feature k for samples x_i and x_j, respectively, and p is a real number that is not <1. (a) If $p = 1$, it is called the Manhattan distance, (b) if $p = 2$, it is called the Euclidean distance, and (c) if $p = \infty$, it is called the Chebychev distance.

If a feature set contains both numerical features B_1 and nominal features B_2, the neighbourhood of sample x induced by B_1 and B_2 is defined as follows [**20**]

$$\delta_{B_1 \cup B_2}(x_i) = \{x_j | x_j \in U, \Delta_{B_1}(x_i, x_j) \leqslant \delta \cap \Delta_{B_2}(x_i, x_j) = 0\} \tag{3}$$

A neighbourhood relation N on the universe can be written as a relation matrix $(r_{ij})_{n \times n}$ where

$$r_{ij} = \begin{cases} 1, & \Delta(x_i, x_j) \leqslant \delta \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

An information system is called a neighbourhood information system, denoted by NIS $= \langle U, A, N \rangle$, if there is a feature in the system generating a neighbourhood relation on the universe. More specifically, if $A = C \cup D$, where C is the set of condition features, and D is the decision feature represented by integers, $1, 2, \ldots, m$, a neighbourhood information system is also called a neighbourhood decision system denoted by NIS $= \langle U, C \cup D, N \rangle$. Suppose $X_1, X_2, \ldots, X_m$ are the object subsets with labels $1, 2, \ldots, m$, respectively. $\delta_B(x_i)$ is the neighbourhood information of sample x_i generated with a feature subset $B \subseteq C$. The lower and upper approximations of decision D with respect to B are defined as

$$\begin{aligned} \underline{N_B} D &= \cup_{i=1}^{m} \underline{N_B} X_i \\ \overline{N_B} D &= \cup_{i=1}^{m} \overline{N_B} X_i \end{aligned} \tag{5}$$

where

$$\begin{aligned} \underline{N_B} X_i &= \{x_i | \delta_B(x_i) \subseteq X_i, x_i \in U\} \\ \overline{N_B} X_i &= \{x_i | \delta_B(x_i) \cap X_i \neq \phi, x_i \in U\} \end{aligned}$$

The lower approximation of the decision is also called the positive region of the decision, denoted by $\mathrm{POS_B}(D)$. It is the subset of objects whose neighbourhood consistently belongs to one of the decision labels. The dependency degree of D to B is defined as

$$\gamma_B(D) = \frac{|\mathrm{POS}_B(D)|}{|U|} \tag{6}$$

where $|\bullet|$ is the cardinality of a set. $\gamma_B(D)$ reflects the ability of B to approximate D (i.e. it reflects the significance of B to D). Let $a \in C - B$; the contribution of a single feature, a, to the approximation of D can be expressed as

$$\mathrm{sig}_B(a, B, D) = \gamma_{B \cup a}(D) - \gamma_B(D) \tag{7}$$

A detailed survey on the neighbourhood rough set model can be seen in [**20**].

2.2 A neighbourhood rough set model with multiple neighbourhood sizes

From equations (3) to (6), it can be seen that the neighbourhood size δ largely influences the neighbourhood of a sample and correspondingly $\mathrm{POS}_B(D)$ and $\gamma_B(D)$. Thus, it is a key parameter which influences the feature evaluation. Hu [**21**] states that the neighbourhood size can be interpreted as the tolerance level for noise which is inevitably contained in the signals collected. Moreover, the noise contained in the signals are not the same considering that signals are usually collected by different accelerometers in a condition monitoring system. Thus, the features obtained from those signals might be subjected to different noise levels and thus should have different neighbourhood sizes. If the same neighbourhood size is used for two features with different noise levels separately, the significance of each feature may be wrongly evaluated. This can be shown as follows.

Given a neighbourhood decision system, NIS $= \langle U, C \cup D, N \rangle$, features $B_1, B_2 \in C$ at the noise levels of δ_1 and δ_2, respectively. If B_2 is noisier than B_1, then

$\delta_1 < \delta_2$. The neighbourhoods of x_i in feature B_1 and feature B_2 are defined as

$$\delta_{B_1}(x_i) = \{x_j | x_j \in U, \Delta_{B_1}(x_i, x_j) \leqslant \delta_1\}$$
$$\delta_{B_2}(x_i) = \{x_j | x_j \in U, \Delta_{B_2}(x_i, x_j) \leqslant \delta_2\} \tag{8}$$

If the neighbourhood size, δ, is used to compare the significances of B_1 and B_2, when a smaller value is chosen, say $\delta = \delta_1$, then $\delta_2 > \delta$; thus

$$\begin{aligned}\delta_{B_1}(x_i)|_\delta &= \delta_{B_1}(x_i)|_{\delta_1} \\ \delta_{B_2}(x_i)|_\delta &= \{x_j | x_j \in U, \Delta_{B2}(x_i, x_j) \leqslant \delta\} \\ &\subseteq \{x_j | x_j \in U, \Delta_{B2}(x_i, x_j) \leqslant \delta_2\} = \delta_{B2}(x_i)|_{\delta_2}\end{aligned} \tag{9}$$

Therefore

$$\underline{N_{B_1}} X_i|_\delta = \underline{N_{B_1}} X_i|_{\delta_1}, \quad \underline{N_{B_2}} X_i|_\delta \supseteq \underline{N_{B_2}} X_i|_{\delta_2} \tag{10}$$

So

$$\gamma_{B_1}|_\delta = \gamma_{B_1}|_{\delta_1}, \quad \gamma_{B_2}|_\delta > \gamma_{B_2}|_{\delta_2} \tag{11}$$

This means that the significance of B_2 is overestimated. Similarly, if a larger value of δ is chosen, say $\delta = \delta_2$, then $\gamma_{B_1}|_\delta < \gamma_{B_1}|_{\delta_1}$ (i.e. the significance of B_1 is underestimated). Feature B_i can be properly estimated only if $\delta = \delta_i$. Therefore, it is necessary to apply different neighbourhood sizes for different features based on estimations of their noise levels. Thus, the modified neighbourhood of sample x_i is given below.

Given a neighbourhood decision system, $\text{NIS} = \langle U, C \cup D, N \rangle$, $B_1, B_2 \in C$. The neighbourhood of x_i in features B_1 and B_2 is defined as

$$\begin{aligned}\delta_{B_1 \cup B_2}(x_i) = \{x_j | x_j \in U, \Delta_{B_1}(x_i, x_j) \\ \leqslant \delta_1 \cap \Delta_{B_2}(x_i, x_j) \leqslant \delta_2\}\end{aligned} \tag{12}$$

where δ_1 and δ_2 are the proper neighbourhood sizes for B_1 and B_2, respectively. If B_i is a nominal feature, then $\delta_i = 0$. The difference between the present work and [**20**] is the calculation of neighbourhood for numerical features. In reference [**20**], equations (1) and (2) are used for neighbourhood calculation, whereas in this article, equation (12) is used. With the neighbourhood defined as equation (12), it can be easily concluded that

$$\delta_{B_1 \cup B_2}(x_i) = \delta_{B_1}(x_i) \cap \delta_{B_2}(x_i) \tag{13}$$

Equation (13) enables us to calculate the neighbourhood of each feature independently and the intersection will be the neighbourhood of the feature combination.

2.3 A feature selection algorithm

The objective of feature selection is to find an optimum feature subset such that the classification problem has the highest classification accuracy and does not have any irrelevant or redundant features. There are generally two kinds of search approaches for feature selection: hill-climbing (or greedy) methods and stochastic methods. The hill-climbing methods usually employ feature significance as heuristic knowledge. They start off with an empty set and then adopt forward feature selection or start with all features and then conduct backward elimination. Forward feature selection is employed here since it involves a smaller amount of computational time. It adds in turn, one at a time, the most significant feature from the candidate set to the selected feature subset, until the significance of the selected feature subset does not grow any more.

Feature evaluation is another key part in feature selection algorithm. As discussed in section 2.1, the dependency degree (equation (6)) reflects the approximation ability of a feature set, whose calculation needs the value of cardinality of the positive region. In practice, an object necessarily belongs to the positive region with respect to a feature set M, if it belongs to the positive region with respect to its subset $B \subseteq M$. Thus for a neighbourhood decision system, $\text{NIS} = \langle U, C \cup D, N \rangle$, when a new feature, a, is added, the objects in $\text{POS}_B(D)$ do not need to be computed when $\text{POS}_{B \cup a}(D)$ is being computed. Let $S = U - \text{POS}_B(D)$, which is the set of samples in U out of the positive region; by constructing a new decision table $\text{DT} = \langle S, B \cup a, D \rangle$, the significance of the feature a is reflected by the cardinality of the positive region with respect to $B \cup a$ in sample space S

$$S_a = |\text{POS}_{B \cup a}(D)|_U - |\text{POS}_a(D)|_U = |\text{POS}_{B \cup a}(D)|_S \tag{14}$$

With equation (14), feature significance can be calculated faster. A forward feature selection algorithm is formulated as below.

A forward feature selection algorithm

Algorithm 1

Input: U– the set of samples;
C– the set of conditional features;
D– the set of decision features.
Output: *red*– the set of selected features

1. $\{\} \rightarrow red, U \rightarrow S$//Initialization. S is the set of samples out of the positive region
2. for each $a_i \in C - red$, $\{\} \rightarrow pos_{a_i}$

   for each $O_j \in S$, compute the neighbourhood of O_j, $\delta(O_j)$, using equation (12)

   if $\exists X_k \in U/D$ such that $\delta(O_j) \subseteq X_k$,
   then $pos_{a_i} \cup O_j \rightarrow pos_{a_i}$

3. select a_k that $|pos_{a_k}| = \max_i |pos_{a_i}|$
   if $pos_{a_k} \neq \phi$,
   then $red \cup a_k \rightarrow red$, $S - pos_{a_k} \rightarrow S$.
   if $S \neq \phi$, then go to step 2.
   else go to step 4.
   else go to step 4.
4. **return** *red*

3 VIBRATION DATA COLLECTION AND FEATURE EXTRACTION FOR A SLURRY PUMP

3.1 Experimental set-up [8]

The schematic diagram of the experimental test rig is shown in Fig. 1. This rig consists of the following key components.

1. Slurry pump: Weir/Warman 3/2 CAH slurry pump with impeller C2147 (8.4″).
2. Motor: 40 HP drive motor complete with a variable frequency drive.
3. Data acquisition system: a 12-channel National Instruments SCXI system.
4. PLC control panel: designed to control and monitor the operation of the system.
5. Sensors: two thermocouples, one microphone, three tri-axial accelerometers, two pressure sensors for the inlet and the outlet, and a differential pressure sensor for flowrate measurement. In this article, the focus is on vibration signals only.
6. Computer: a Dell Inspiron 9200 laptop computer for data collection via Labview.
7. Others: seal water pump, inlet pressure control tank, sand addition tank, safety rupture disk, various valves, pipes, and the glycol cooling system.

Three tri-axial vibration accelerometers were mounted on the system, one on the side of the outlet (accelerometer 1), one on the top of the impeller casing (accelerometer 2), and one on the top of the bearing casing (accelerometer 3), as shown in Fig. 2.

Two of the common damage modes in a slurry pump are the leading edge damage and the trailing edge damage [**2**, **8**], as illustrated in Fig. 3. Experiments

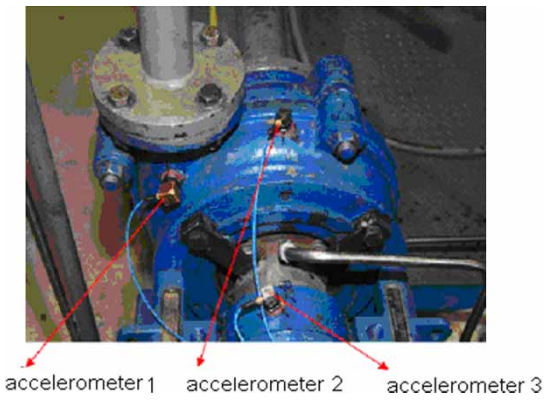

Fig. 2 Locations of the three accelerometers

have been conducted with one undamaged impeller and two damaged impellers. Three damage levels – slight, intermediate, and server – were fabricated, one for each damage mode; however, only slight damage is discussed in this article because we want to focus on the initial damage detection. Vibration data were collected at a pump speed of 2600 r/min with both water and slurry (specific gravity 1.17) as the medium. The sampling rate was 9 kHz, and the duration of each sampling period was 5 min.

3.2 Feature extraction from vibration data

According to the technical bulletin of Warman [**22**], the likely vibration frequencies in a slurry pump include the pump frequency, its second, third, and fourth harmonics, and the vane passing frequency. Also it is shown in reference [**23**] that the jet-wake effect which may occur at the impeller outlet will result in the second harmonic of the vane passing frequency. The impeller in this experiment has five vanes; therefore the pump frequency, its second, third, fourth, fifth, and tenth harmonics are chosen as the valuable frequencies.

Figures 4(a) and (b) show three vibration signals corresponding to no damage, the trailing edge damage and the leading edge damage in time and frequency domains, respectively. These three signals were all measured by accelerometer 2 in the y-direction. It can

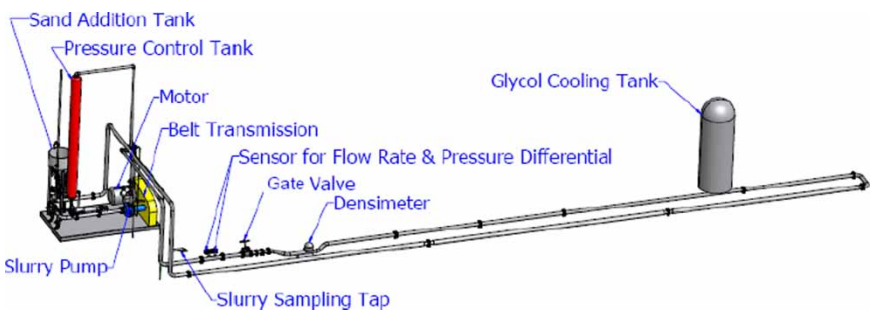

Fig. 1 Schematic of the laboratory test-rig pump loop

Fig. 3 Impellers used in the experiment

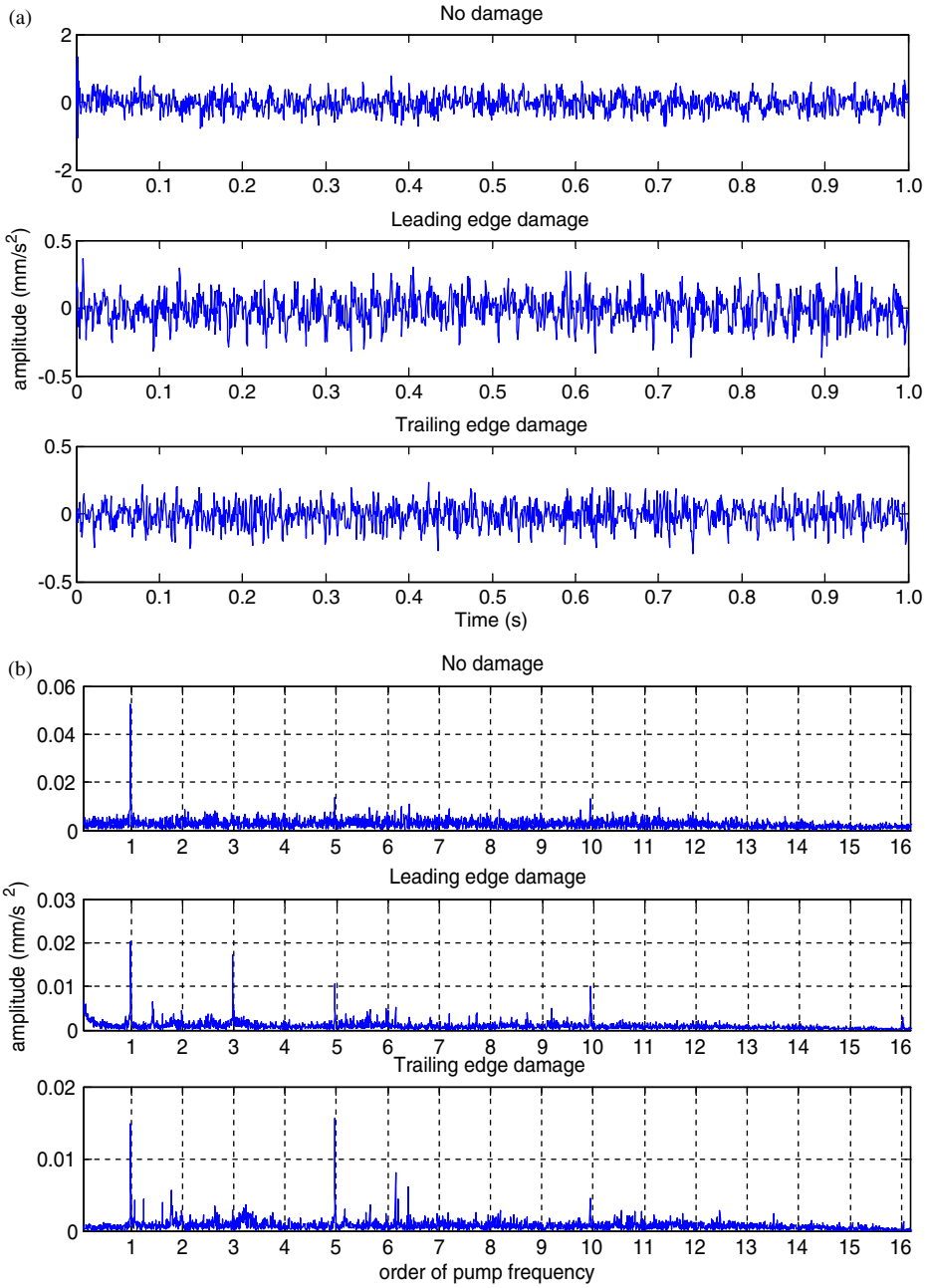

Fig. 4 (a) Three vibration signals in the time domain and (b) the frequency spectrum of the three vibration signals

be seen that most peaks are located at the chosen valuable frequencies, while some peaks show up at fractional frequencies, e.g. 6.2 times the pump frequency. It is not certain as to whether the fractional frequencies are related to impeller damage, since in some samples they also appear in the frequency spectra under undamaged conditions. Because of the uncertainty of its source, here it was not taken as a valuable frequency.

From the frequency spectra in Fig. 4(b), it can be seen that the amplitudes of undamaged impellers at different frequencies are different from those of damage impellers. Considering the fact that the amplitudes may be influenced by the amount of damage, and since the distinction of damage modes is being considered, it is not appropriate to use the value of amplitude directly as features. Thus the amplitude ratios are chosen as features.

As mentioned before, the pump frequency, its second, third, fourth, fifth, and tenth harmonics are chosen as valuable information, whereas all the other frequencies are treated here as noise. The noise amplitude is evaluated with equation (15), where j denotes all the frequencies that are not treated as valuable frequency, and A_j stands for its corresponding amplitude

$$\text{Noise} = \sqrt{\frac{1}{N}\sum_{j=1}^{N} A_j^2} \tag{15}$$

For each set of vibration data, seven features shown below have been extracted. Since there are three tri-axial accelerometers in this experiment, the total number of features will be $7 \times 3 \times 3 = 63$

$$A_{\text{total}} = A_{1X} + A_{2X} + A_{3X} + A_{4X} + A_{5X} + A_{10X} + \text{Noise} \tag{16}$$

$$\text{feature } i = \frac{A_{iX}}{A_{\text{total}}}, \quad \text{where } i = 1, 2, 3, 4 \tag{17}$$

$$\text{feature } 5 = \frac{A_{5X}}{A_{\text{total}}} \tag{18}$$

$$\text{feature } 6 = \frac{A_{10X}}{A_{\text{total}}} \tag{19}$$

$$\text{feature } 7 = \frac{\text{Noise}}{A_{\text{total}}} \tag{20}$$

4 EXPERIMENTAL RESULTS AND DISCUSSION

Experiments were conducted with water and slurry used separately as the medium. The whole dataset consists of 190 samples. A detailed description is given in Table 1.

As discussed in section 3.2, 63 features were obtained from each vibration signal. Considering that in real applications, the density of the slurry may be set as different values, it is valuable for us to analyse the influence of density on the results of fault diagnosis. Thus a nominal feature is added, i.e. 0 (water), 1 (slurry with specific gravity 1.17), which makes the total number of features to be 64. The sources of the features are listed in Table 2. For each source, the features are generated by and in order of equations (17) to (20).

Table 1 Data description

	Medium – water	Medium – slurry	Total
No damage	31	32	63
Leading edge damage	31	31	62
Trailing edge damage	32	33	65
Total	94	96	190

To evaluate the performance of feature selection, a probabilistic neural network (PNN), a K-nearest neighbour (KNN), and a support vector machine (SVM) were used as the learning algorithms. The key parameters for these classifiers were set as follows. For the PNN, the spread of radial basis functions varied from 0.01 to 1 with a step size of 0.01. For the KNN, the number of neighbourhood points varied from 1 to 5 with a step size of 1. For the SVM, the Gaussian radial basis function was used as the kernel function, and the cost of the constraint violation C, varied from 0 to 100 with a step size of 1. The raw 190 samples were partitioned into the training, validation, and test sets as shown in Table 3. In the case of an SVM, for example, the training set was used to train the classifier for each selected C value to find the optimal weights of the classifier. The validation set was used to measure the performance of the classifier with the specified C value and the optimized weight values. The classifier with the C value and the corresponding optimized weights that had the best classification performance on the validation data set was selected to be the final classifier. The test set was then used to assess the classification performance of this fully trained classifier, and its classification performance was then compared

Table 2 Sources of features

Feature label	Sources	Feature label	Sources
1–7	Accelerometer 1, x axis	36–42	Accelerometer 2, z axis
8–14	Accelerometer 1, y axis	43–49	Accelerometer 3, x axis
15–21	Accelerometer 1, z axis	50–56	Accelerometer 3, y axis
22–28	Accelerometer 2, x axis	57–63	Accelerometer 3, z axis
29–35	Accelerometer 2, y axis	64	0-water; 1-slurry

Table 3 Data division to training, validation, and test sets

	Training set	Validation set	Test set
Total number of samples	64	63	63

with the performance of this classifier when different feature sets were used.

4.1 Performance of the neighbourhood rough set model with multiple neighbourhood sizes

As discussed in section 2.2, a proper neighbourhood size should be used for each feature. From the measurement point of view, the neighbourhood size reflects a feature's tolerance level to noise, and thus can be regarded as a measurement of noise level [**21**]. The noise mentioned here is generated in the process of measurement and quantization. In this article, it is assumed that the vibration data measured by the same accelerometer in the same direction (for convenience, the term 'channel' will be used hereafter) are subjected to the same noise level. Furthermore, the noise is assumed to be white noise; thus the noise level, and therefore the neighbourhood size, for features generated by each channel is the same. In the following paragraph, how to calculate the neighbourhood sizes for the vibration signals will be discussed.

To calculate the neighbourhood size, first the root mean square of the noise amplitude is calculated. In equation (15), the definition of noise is used, which was used in the signal processing field, that is, noise can be considered as any data that are not generated by the sources of interest, and j is a frequency generated by such an irrelevant source. It is possible that equation (15) contains a frequency, say k, which does not contribute to noise in the general definition used in the previous paragraph. But considering the limited number of k and the large value of N, equation (15) can be regarded as a reasonable approximation of the RMS noise amplitude. Furthermore, taking the normalization into consideration, equation (20) can be used as an estimation of the neighbourhood size.

Using equation (20), a value for the neighbourhood size was calculated for each sample of data. The average value for the 190 samples was used as the neighbourhood size for features generated by signals collected from a specific channel. The results are shown in Table 4. It can be seen that features from different channels have different neighbourhood sizes.

With the neighbourhood size determined, the forward selection algorithm in section 2.3 was then applied. In the first step, feature No. 1 was selected because it had the highest dependency degree (as shown in Fig. 5). Then in the second step, five other features (No. 12, No. 27, No. 30, No. 43, and No. 51) were selected, each of which generated a dependency

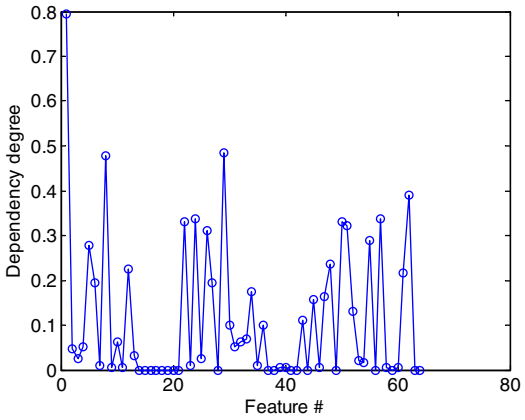

Fig. 5 The dependency degree for each feature

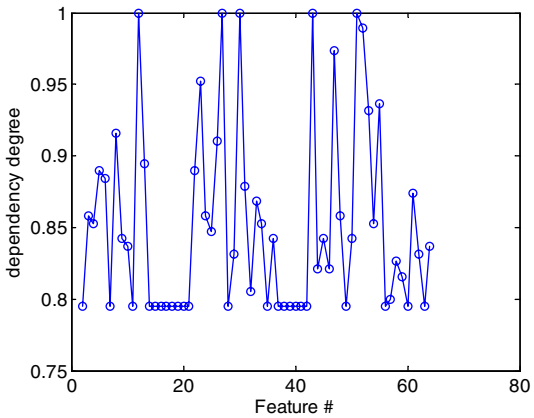

Fig. 6 The dependency degrees for feature subsets consisting of feature No. 1 and one of the remaining features

degree equal to one (shown in Fig. 6) together with feature No. 1. Thus the search stopped at the second step and five feature subsets were generated. Table 5 shows the performance of the five feature subsets. Three classifiers (SVM, KNN, and PNN) are used separately to test the performance of each feature subset. The corresponding average classification accuracies are shown in the last row. The first two feature subsets comprised of features No. 1 and No. 12 (denoted as [No. 1, No. 12]), and features No.1 and No. 27 (i.e. [No. 1, No. 27]) are comparable, and give the two highest average classification accuracies 99.56 per cent and 99.33 per cent, respectively.

To illustrate the effect of the feature selection by the modified neighbourhood rough set model, it is compared with the case which uses the raw features and

Table 4 Calculated neighbourhood sizes for features

Feature label	1–7	8–14	15–21	22–28	29–35	34–41	43–49	50–56	57–63	64
Noise level	0.023	0.027	0.064	0.019	0.021	0.050	0.026	0.020	0.045	0

Table 5 Classification accuracy (%) of feature subsets selected by the neighbourhood rough set model with multiple neighbourhood sizes

	[No. 1, No. 12]	[No. 1, No. 27]	[No. 1, No. 30]	[No. 1, No. 43]	[No. 1, No. 51]
SVM	99.73	100	98.66	97.04	99.73
KNN	99.73	100	95.16	91.94	98.66
PNN	99.21	98.00	96.01	91.79	98.81
Average	99.56	99.33	96.61	93.59	99.07

Table 6 Classification accuracy (%) of the raw feature subsets (without feature selection)

	[No. 1–No. 64] (all three sensors)	[No. 1–No. 21] (accelerometer 1)	[No. 22–No. 42] (accelerometer 2)	[No. 43–No. 63] (accelerometer 3)
PNN	86.06	89.38	88.88	76.17
KNN	83.06	89.25	90.32	79.84
SVM	83.06	91.13	91.40	79.57
Average	84.06	89.96	90.20	78.53

the case in which feature selection is conducted using the original neighbourhood rough set model.

4.1.1 Comparison with the raw features

Without feature selection, feature subsets that consist of all features obtained from different accelerometers are tested directly by the classifiers; the corresponding classification accuracies are given in Table 6. Comparing Tables 6 and 5, it is obvious that feature selection increases the classification accuracy from around 84.06 per cent to 99.56 per cent.

4.1.2 Comparison with the neighbourhood rough set model with a common neighbourhood size

For the original neighbourhood rough set, a common neighbourhood size should be chosen for all the numerical features. Table 4 shows that the noise levels ranged from 0.019 to 0.064 in this experiment. Since the proper value of the neighbourhood size for all numerical features is not known, different values of the neighbourhood size, from 0.01 to 0.07 with a step size of 0.02, were tested. The results are shown in Table 7. It can be seen that different neighbourhood sizes generate different feature reduction subsets, and correspondingly different classification results. Overall, the classification accuracies are smaller here than those in Table 5. The optimum feature subset is [No. 1, No. 27], which gives the highest accuracy, 99.33 per cent, in Table 7. This feature subset is one of the two optimum feature subsets (i.e. [No. 1, No. 12] and [No. 1, No. 27]) generated in Table 5. This supports the idea that if a neighbourhood size is properly selected, the original neighbourhood rough set model might also produce acceptable results. The problem is that it is not easy to determine the proper neighbourhood size by enumeration especially for complex systems. Furthermore, it is noticeable to say that the average neighbourhood size in Table 4 is calculated to be 0.033, which is close to the common neighbourhood size, 0.03, that results in the feature subset [No. 1, No. 27] in Table 7. This supports the idea that the closer the neighbourhood size is to the actual noise level, the higher the classification accuracy. This, on the other hand, proves the feasibility of equation (20) as a neighbourhood size approximation calculation formula.

As discussed in section 2.2, when the neighbourhood size used is smaller than its real value, the feature's significance may be overestimated. Thus in the feature selection process, some bad features may be included. It can seen in Table 7 that for $\delta = 0.01$, a total of nine feature subsets are generated, some of which perform badly with even a classification accuracy of only 63.94 per cent. On the other hand, if the neighbourhood size used is larger than the value it should be, the feature's significance may be underestimated. Thus in the feature selection process, some good features may be missed. For example, in the case

Table 7 Classification accuracy (%) of feature subsets selected by the neighbourhood rough set model with a common neighbourhood size

δ	Feature subset	SVM	KNN	PNN	Average	
0.01	[No. 1, No. 12]	99.73	99.73	99.21	99.56	90.79
	[No. 1, No. 27]	100	100	98.00	99.33	
	[No. 1, No. 30]	98.66	95.16	96.01	96.61	
	[No. 1, No. 36]	87.10	87.63	88.85	87.86	
	[No. 1, No. 43]	97.04	91.94	91.79	93.59	
	[No. 1, No. 47]	67.20	63.44	61.18	63.94	
	[No. 1, No. 51]	99.73	98.66	98.81	99.07	
	[No. 1, No. 52]	80.91	80.11	76.86	79.29	
	[No. 1, No. 61]	97.58	97.31	98.63	97.84	
0.03	[No. 1, No. 27]	100	100	98.00	99.33	
0.05	[No. 1, No. 12, No. 8]	98.12	100	99.39	99.17	
0.07	[No. 1, No. 12, No. 8, No. 3]	89.25	88.71	94.69	90.88	

of $\delta = 0.03$, the optimum feature subset [No. 1, No. 12] is missed; as stated in Table 4, the noise level for No. 12 should be 0.027.

4.2 Discussion of classification results

4.2.1 *Physical explanation of the optimal feature subsets*

Table 8 shows the physical meaning of each feature selected by the modified neighbourhood rough set model. The general causes and likely frequencies of the vibration are stated in Table 9.

Of the four related frequencies, $1X$ (the pump frequency) and $2X$ (the second harmonic of the pump frequency) are related to mechanical reasons, while $5X$ (the fifth harmonic of the pump frequency) and $10X$ (the tenth harmonic of the pump frequency) are related to flow patterns inside the pump. For $2X$, we are not sure how to relate it to the impeller damage, even though the physical explanations for the two optimal feature subsets ([No. 1, No. 12], [No. 1, No. 27]), which result in the two highest classification accuracies (shown in Table 5), can be discussed as follows.

The pump frequency may be caused by impeller imbalance [**22**, **24**]. Although in this experiment, damage exists on all vanes, the balance condition of the damaged impeller is not guaranteed to be the same as that of the undamaged impeller. Therefore, pump frequency may contain damage information.

When the trailing edge damage or the leading edge damage occurs, some complex flow patterns such as flow recirculation and jet-wake may show up. [**25**] states that flow recirculation at the impeller inlet and outlet are likely to influence the vane passing frequency ($5X$). Also, according to Hodkiewicz and Norton [**27**], the pressure pulsations are affected by the gap between the impeller and the volute casing, which leads us to conclude that when the trailing edge damage occurs, the gaps change and, as a result, vibration at $5X$ might be influenced. According to reference [**22**], the impeller is deigned to have the scallops at the trailing edges to ensure an even discharge across the width of the impeller. Obviously, when trailing edge happens, the even discharge cannot be maintained. It is possible that a jet-wake flow pattern appears, which contributes to $10X$.

Table 8 Physical meaning and location of each selected feature

Feature label	Physical meaning – location
1	Amplitude ratio at $1X$ – accelerometer 1, x-axis
12	Amplitude ratio at $5X$ – accelerometer 1, y-axis
27	Amplitude ratio at $10X$ – accelerometer 2, x-axis
30	Amplitude ratio at $2X$ – accelerometer 2, x-axis
43	Amplitude ratio at $1X$ – accelerometer 3, x-axis
51	Amplitude ratio at $2X$ – accelerometer 3, x-axis

Table 9 Frequencies and likely causes

Frequency	Causes
Pump frequency	Impeller imbalance [**22**, **24**]
$2X$	Mechanical looseness, vee belt drive, misaligned coupling, and bent pump shaft [**22**]
$5X$	Pressure pulsations [**26**], flow recirculation at the impeller inlet or at the impeller outlet [**24**, **26**]
$10X$	Jet-wake effect [**23**, **25**]

4.2.2 *Comparison of accelerometer locations*

From Table 6, it can be seen that accelerometer 3 gives the lowest classification accuracy. The vibration signals from accelerometer 3 contain less information about the impeller condition, because this accelerometer is mounted on the bearing casing, which is relatively far from the impeller; it is therefore quite likely subjected to the complicated high frequency vibration interference from the rolling element bearings. Accelerometer 2 is located on the top of the impeller casing, which is near the impeller outlet where the jet-wake effect is strong [**23**]. This might be the reason as to why feature No. 27 ($10X$) was chosen. Accelerometer 1 is located on the side of the outlet (i.e. near the tongue of the volute liner) where pressure pulsation is strong [**26**]. This explains why feature No. 12 ($5X$) was chosen.

4.2.3 *Influence of medium: water and slurry*

The medium as a nominal feature has the constant neighbourhood size, 0, in both the modified and the original neighbourhood rough set models, but it is not included in the selected feature subsets. Figure 7 shows the scatter plot for feature subsets [No. 1, No. 64], from which it can be seen that when the medium is slurry (i.e. feature No. 64 is 1), the leading edge damage and the trailing edge damage can be clearly

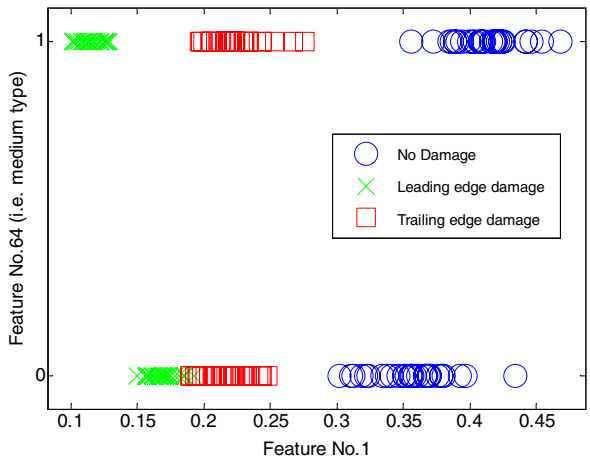

Fig. 7 Scatter plot for feature subsets (No. 1 and No. 64)

classified from the no damage case by feature No. 1. However, for water (i.e. feature No. 64 is 0), the cross symbols of the leading edge damage case are partly mixed up with those of the no damage case. From Fig. 6, it can be seen that the medium type provides some information; however, the information provided is not adequate to generate a dependency degree that is high enough. That is why No. 64 is not chosen.

Though No. 64 was not selected based on the specific data set analysed, the analysis of the medium as a nominal feature still provides us with useful information on the influence of liquid density on fault diagnosis. First, from Fig. 7, it can be seen that feature No. 1 distinguishes three classes for the slurry case well, although it does not perform so well for the water case. However, as shown in section 4.1, some feature subsets (e.g. [No. 1, No. 12]) can classify three classes well if water and slurry cases are considered together. This, on the other hand, indicates that when the pumped liquid density varies, more features may be needed to classify slight damages on the impeller leading edge and trailing edge. Second, this analysis helps us to understand the applicable density range of the classification model. If the medium type is included in the final feature subset, it means that the medium will play a significant role in the classification model. The model generated by data of the water case may not work for the slurry case. If, however, the medium type is not chosen, this means that the classification model may work well for both the water case and the slurry case. For example, based on the data analysed in this study, we may say that the trained classification model is suitable for specific density ranges below 1.17.

5 CONCLUSIONS

This article analyses the problem of applying a neighbourhood rough set model to condition monitoring systems. To overcome the problem of improper estimation of dependency degree, the neighbourhood rough set model is modified with multiple neighbourhood sizes. The modified model is applied to the feature selection in the process of fault diagnosis of slurry pump impellers. Three impeller conditions (no damage, leading edge damage, and trailing edge damage) – are classified. Results show that feature subsets generated by the modified neighbourhood rough set model can achieve higher classification accuracy than do the raw features and features selected by the original neighbourhood rough set model. Based on the experimental results, the following conclusions can be drawn.

1. Multiple neighbourhood sizes should be considered in the neighbourhood rough set model. If a common neighbourhood size is used for all features, the dependency degree tends to be overestimated or underestimated.
2. Feature selection based on the modified neighbourhood rough set model is useful in condition monitoring of a slurry pump impeller. After feature selection, the classification accuracy for the undamaged impeller, the slight trailing edge damage impeller, and the slight leading edge damage impeller increases up to 99.56 per cent.
3. Accelerometers located at the top of the impeller casing and at the side of the outlet contain more information on impeller condition, while the accelerometer on the bearing casing contains less information.
4. The pump frequency, the vane passing frequency, and the second harmonic of the vane passing frequency provide important information on the detection of initial damage on the impeller leading edge and trailing edge, and therefore should be paid attention when monitoring the impeller condition.

ACKNOWLEDGEMENT

This research was supported by Syncrude Canada Ltd and the Natural Sciences and Engineering Research Council of Canada. Qinghua Hu is supported by the National Natural Science Foundation of China under grant 60703013. The authors would like to thank Mr Dan Wolfe for his help in the experiment. The comments and suggestions from the reviewers and the Editor are very much appreciated.

REFERENCES

1 **Liu, J., Xu, H., Qi, L.,** and **Li, H.** Study on erosive wear and novel wear-resistant materials for centrifugal slurry pumps. In Proceedings of the ASME Heat Transfer/Fluids Engineering Summer Conference, Charlotte, NC, USA, 2004.

2 **Khalid, Y. A.** and **Sapuan, S. M.** Wear analysis of centrifugal slurry pump impellers. *Ind. Lubr. Tribol.*, 2007, **59**(1), 18–28.

3 **Cornelius, S.** Pump condition monitoring through vibration analysis. In Proceedgins of the Pumps: Maintenance, Design and Reliability Conference 2008 – IDC Technology, Johannesburg, South Africa, 2008.

4 **Jiang, W., Spurgeon, S. K., Twiddle, J. A., Schlindwein, F. S., Feng, Y.,** and **Thanagasundram, S.** A wavelet cluster-based band-pass filtering and envelope demodulation approach with application to fault diagnosis in a dry vacuum pump. *Proc. IMechE, Part C: J. Mechanical Engineering Science*, 2007, **221**(C11), 1279–1286. DOI: 10.1243/09544062JMES544.

5 **Wang, L.** and **Hope, A. D.** Bearing fault diagnosis using multi-layer neural networks. *Insight: Non-Destr. Test. Cond. Monit.*, 2004, **46**(8), 451–455.

6 **Wang, J.** and **Hu, H.** Vibration-based fault diagnosis of pump using fuzzy technique. *Measurement*, 2006, **39**(2), 176–185.

7 **Chudina, M.** Noise as an indicator of cavitation in a centrifugal pump. *Acoust. Phys.*, 2003, **49**(4), 463–474.

8 **Zuo, M. J., Wu, S. Y., Feng, Z. G., Aulakh, A.,** and **Miao, C. X.** Condition monitoring of slurry pumps for wear assessment. Final report, technical report, University of Alberta, 2007.

9 **Samanta, B., Al Balushi, K. R.,** and **A1Araimi, S. A.** Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection. *Eng. Appl. Artif. Intell.*, 2003, **16**(7–8), 657–665.

10 **Kira, K.** and **Rendell, L. A.** The feature selection problem: traditional methods and a new algorithm. In Proceedings of the Tenth National Conference on *Artificial intelligence*, 1992, pp. 129–134 (MIT Press, Cambridge, MA, USA).

11 **Almuallim, H.** and **Dietterich, T. G.** Learning with many irrelevant features. In Proceedings of the Ninth National Conference on *Artificial intelligence*, 1991, pp. 547–552 (MIT Press, Cambridge, MA, USA).

12 **Ruiz, R., Aguilar, J.,** and **Riquelme, J.** SOAP: efficient feature selection of numeric attributes. In Proceedings of the Eighth Ibero-American Conference on *AI: Advances in artificial intelligence*, Spain, 2002, pp. 233–242.

13 **Pawlak, Z.** *Rough sets: theoretical aspects of reasoning about data*, 1991 (Kluwer academic publisher, Dordrecht).

14 **Ching, J. Y., Wong, A. K. C.,** and **Chan, K. C. C.** Class-dependent discretization for inductive learning from continuous and mixed-mode data. *IEEE Trans. Pattern Anal. Mach. Intell.*, 1995, **17**(7), 641–651.

15 **Yu, D. R., Hu, Q. H.,** and **Bao, W.** Combining rough set methodology and fuzzy clustering for knowledge discovery from quantitative data. *Proc. Chin. Soc. Electr. Eng.*, 2004, **24**(6), 205–210.

16 **Jensen, R.** and **Shen, Q.** Fuzzy-rough set assisted attribute selection. *IEEE Trans. Fuzzy Syst.*, 2007, **15**(1), 73–89.

17 **Yeung, D. S., Chen, D., Tsang, E. C. C., Lee, J. W. T.,** and **Zhao, W. X.** On the generalization of fuzzy rough sets. *IEEE Trans. Fuzzy Syst.*, 2005, **13**(3), 343–361.

18 **Jin, W., Tung, A. K. H., Han, J.,** and **Wang, W.** Ranking outliers using symmetric neighborhood relationship. In Proceedings of the 2006 Pacific-Asia Conference on *Knowledge discovery and data mining* (PAKDD'06), Singapore, April 2006, pp. 577–593.

19 **Shin, H.** and **Cho, S.** Invariance of neighborhood relation under input space to feature space mapping. *Pattern Recognit. Lett.*, 2005, **26**(6), 707–718.

20 **Hu, Q. H., Yu, D. R., Liu, J.,** and **Wu, C.** Neighborhood rough set based heterogeneous feature subset selection. *Inf. Sci.*, 2008, **178**(18), 3577–3594.

21 **Hu, Q. H.** *Rough computation models and algorithms for knowledge discovery from heterogenous data*. PhD Thesis, School of Energy, Harbin Institute of Technology, Harbin, China, 2008.

22 Warman Group Development. Impeller balancing and pump vibration. *Tech. Bull.*, 1991, **1**, 3.

23 **Zheng, Q.** and **Liu, S.** Numerical investigation on the formation of jet-wake flow pattern in centrifugal compressor impeller. In Proceedings of the ASME, International Gas Turbine Institute Turbo Expo (Publication) IGTI, 6 B, 2003 pp. 755–764.

24 **Dyson, G.** Impeller rerate to reduce hydraulically generated vibration. In Proceedings of the 22nd International Pump Users Symposium, 2005, pp. 10–15.

25 **Schiavello, B.** and **Cicatelli, G.** Vibration field problem resolved with analytical diagnostics approach and innovative impeller design. In Proceedings of the 24th International Pump Users Symposium, 2007, pp. 39–49.

26 **Majidi, K.** Numerical study of unsteady flow in a centrifugal pump. *J. Turbomach.*, 2005, **127**(2), 363–371.

27 **Hodkiewicz, M. R.** and **Norton, M. P.** The effect of change in flow rate on the vibration of double-suction centrifugal pumps. *Proc. IMechE, Part E: J. Process Mechanical Engineering*, 2002, **216**(E1), 47–58. DOI: 10.1243/095440802760075030.