A novel technique for selecting mother wavelet function using an intelligent fault diagnosis system

J. Rafiee, P.W. Tse, A. Harifi, M.H. Sadeghi

1. Introduction

The application of a condition-based maintenance strategy used for detection and diagnosis of incipient faults can help us to minimize avoidable costs and impediments caused by the need to perform ad hoc repairs. Gearboxes consisting of three major components (gears, bearings, and shafts) are considered to be one of the essential systems on account of its influential applications in a wide range of industries (e.g., machine tools and vehicles). As more than half of the failures in gearboxes are due to gear defects (Yesilyurt, 2004), the gear faults have been considered as a case study for the proposed technique.

In general, acoustic emission (Wu, Chiang, Chang, & Shiao, 2008) and vibration analysis (Tse, Gontarz, & Wang, 2007) are two main centers of fault detection and diagnosis systems in machine condition monitoring. In this paper, vibration signals established as the way of recording required dataset because of the ease of measurement and the rich contents of information. To analyze raw vibration signals, a simultaneous time–frequency analysis method which maps one-dimensional signals into a two-dimensional space of time and frequency was applied because of non-stationarity of vibration signals. Wavelet analysis, the most popular one for analyzing the non-stationary signals (Peng et al., 2005), overcomes the drawbacks of other techniques by means of analytical functions that are local in both time and frequency (Tse, Yang, & Tam, 2004). Therefore, in pre-processing and feature extraction phase of the research, wavelet transform (WT) (Tse, Peng, & Yam, 2001) has been used to take out proper features. To bolster this recommendation, in early 1990, Leducq applied wavelet transform to analyze the hydraulic noise of the centrifugal pump which is probably the first paper in application of wavelet transform in machine diagnostics (Peng & Chu, 2004); Momoh and Dias (Momoh & Dias, 1996), moreover, used both fast Fourier transform (FFT) and wavelet transform (WT) to obtain processed vibration signals to identify and pinpoint the location of the faults in power machinery.
distribution systems. The prior research demonstrates the wavelet transform as the most reliable technique for gear condition monitoring and a capable technique to recognize the incipient failures and to identify different types of faults simultaneously.

The most indispensable challenge in wavelet analysis is the selection of the mother wavelet function as well as the decomposition level of signal. As a result of the use of dyadic discrete wavelet transform (DWT), orthogonal wavelets have been applied in this research; among them Daubechies (DB) wavelets have been widely implemented as it matches the transient components in vibration signals. Another main issue in wavelet analysis is the order of the mother wavelet function, which was previously determined by trial-and-error methods based on intrinsic characteristics of the data in several papers (Wang & McFadden, 1995; Samanta & Al-Balushi, 2003; Tse et al., 2004; Liu, 2005; Kar & Mohanty, 2006; Saravanan et al., 2007; Rafiee, Arvani, Harifi, & Sadeghi, 2007). To more explanations, the range of DB2 and Db20 has been extensively used in machine condition monitoring. Wang and McFadden (1995) used DB4 orthogonal wavelets to disclose abnormal transients generated by early gear damage from gearbox vibration signal. Sung et al. and also Gaborson used DB20 and DB4, as the mother wavelet, in their works, respectively (Kar & Mohanty, 2006). Samanta and Al-Balushi (2003) and Rafiee et al. (2007) processed vibration signals through discrete wavelet transform (DWT).
Fig. 3. Raw vibration signals of four gearbox conditions recorded during one revolution of input shaft.

Fig. 4. Synchronized vibration signals of four gearbox conditions recorded during one revolution of input shaft.
using DB4 to obtain the wavelet coefficients. The most important issue as it has been shown in Fig. 1, is that it can not be functional if the adjacent wavelet functions (e.g. DB6 and DB7) are compared as they do not differ from each other tremendously.

In general, the current fault detection and diagnosis techniques are predominantly based on intelligent systems, modeling using classical techniques in time or frequency domain, and statistical analysis. Fuzzy-based (Wang & Hu, 2006), ANN-based (Wuxing, Tse, Guicai, & Tielin, 2004), GA-based (He, Guo, & Chu, 2001) methods as well as intelligent hybrid systems such as fuzzy-GA-based (Lei et al., MSSP 2007), (Lo, Chan, Wong, Rad, & Cheung, 2007), Nero-fuzzy-based (Lei et al., ESWA 2007), Nero-GA-based (Samanta, 2004) systems can be categorized as the intelligent fault diagnosis systems. In this category, ANN-based systems consisting of signal preprocessing, feature extraction and recognition process have been introduced as one of the most noteworthy methods in several papers. To name a few, Li, Chow, Tipsuwan, and Hung (2000) presented an approach for motor rolling bearing fault diagnosis using both simulation and experiment to acquire vibration signals which was used to obtain a feature vector to adjust ANN weights. Samanta and Al-Balushi (2003) introduced a new technique for fault detection of rolling bearing elements through ANNs which was obtained from statistical properties of time-domain vibration signals of the rotating machinery with normal and defective bearings. In 2003, Samanta, Al-Balushi, and Al-Araimi (2003) also enhanced the proposed procedure using GA to meticulously optimize the mentioned feature vector for gear fault detection using the experimental vibration data of a gearbox system. Inspired by Jack and Nandi (2002) innovations, the paper presents an optimized ANN-based system for identifying the gear faults of a gearbox system. It also presents a novel solution to find the best mother wavelet function for fault classification purpose as well as the best level of decomposing the vibration signals by wavelet analysis in machine condition monitoring and related areas. In addition, the small structure optimized network has improved the stability and reliability of the system in practical implementations.

2. Summary of the developed procedure

Since gears are the most challenging components in condition monitoring, a complex gearbox system has been considered as the case study. Of paramount importance is to synchronize the raw vibration signals which usually don’t have the same length in a complete revolution of the shaft. To overtake the impediment, piecewise cubic hermite interpolation (Rafiee et al., 2007) was suggested to synchronize the signals. Afterwards, wavelet packet analysis (Liu, 2005) has been considered to preprocess the synchronized signals for presenting a distinguished feature vector. Then, selecting of the mother wavelet function and decomposition level has been reformed by means of genetic algorithm as an obstacle in this regard. To present a more accurate fault classification system, number of neurons in hidden layer has been also chosen to be optimized by GA.

2.1. Experimental set-up

A proper data acquisition process can extensively improve the sensibility for the detection of failures. Therefore, selecting an appropriate method is important for acquiring vibration signal data. The experimental set-up (Rafiee et al., 2007) to collect dataset in this research consists of a four-speed motorcycle gearbox containing the oil during acquiring dataset, an electrical motor with a constant nominal rotation speed of 1420 RPM, a load mechanism, multi-channel pulse analyzer system, a triaxial accelerometer, tachometer and four shock absorbers under the bases of test-bed. The vibration signals were recorded by mounting the accelerometer on the outer surface of the gearbox’s case near input shaft of the gearbox. Ten different configurations were synthesized and tested. For each configuration, three different fault conditions were tested that were slight-worn, medium-worn, broken-tooth of gear. For evaluating the meticulousness of the technique, two very similar models of worn gear have been taken into account with partial difference. The rotational speed of the system was measured by tachometer used as a measure to compensate the fluctuations owing to uncertainties of the load mechanism which was a constant static load. Moreover, the signals were also sampled at 16384 Hz lasting eight seconds. Fig. 2 clearly shows all parts of fault simulator and related components in detail.

2.2. Preprocessing of raw vibration signals

It is important to pre-process any raw data before use since it contains redundant information. The pre-processing of vibration signals was involved in synchronization of the signals and computation of standard deviation of wavelet packet coefficients per-
formed in the following two steps to extract the feature vector. Accurate pre-processing of the data to feed neural network can make ANN training more efficient because of a considerable reduction of the dimensionality of the input data and therefore improving the network performance.

2.2.1. Synchronization using piecewise cubic hermite interpolation

In the recorded signals, as a consequence of non-synchronous attribute of vibration signals, the number of data-points per each revolution may not be exactly equal from one revolution to another one, due to shaft speed fluctuations which can influence the time-domain average adversely. To overcome the drawback, piecewise cubic hermite interpolation (P.C.H.I.) (Fritsch & Carlson, 1980) was exploited to resample the obtained data in each revolution as it outperforms the other methods (Rafiee et al., 2007). A menace of interpolation is that the fitting function may tend to demonstrate unsolicited fluctuations, essentially following random errors in the data and oscillations of the data points. This problem is particularly persisting if the fitting function is a polynomial defined over the entire region of interest, and becomes more critical as the number of data points increases. Consequently, P.C.H.I. was implemented as the interpolating scheme. In this method, the same sample spans in each revolution could be guaranteed. In this paper, the average length of the 150 segmented sample signals was applied to synchronize the whole signal set. Given that $h_k$ is the length of the $k$th subinterval,

$$h_k = x_{k+1} - x_k$$  \hspace{1cm} (1)

Then the first difference, $d_k$ is calculated by the following equation,

$$d_k = (y_{k+1} - y_k)/h_k$$  \hspace{1cm} (2)

Assume that $d_k$ designate the slope of the interpolant at $x_k$. 

![Wavelet packet coefficients of a sample synchronized broken-tooth signal in four levels.](image)

Fig. 6a. Wavelet packet coefficients of a sample synchronized broken-tooth signal in four levels.
for the piecewise linear interpolant, \( d_k = \delta_{k-1} \) or \( \delta_k \).

Suppose the following function on the interval \( x_k \leq x \leq x_{k+1} \), expressed in term of local variables \( s = x - x_k \) and \( h = h_k \),
\[
P(x) = \frac{3hs^2 - 2s^3}{h^3} y_{k+1} + \frac{h^2 - 3hs^2 + 2s^3}{h^3} y_k + \frac{s^2(s - h)}{h^2} d_{k+1} + \frac{s(s - h)^2}{h^2} d_k
\]  
(4)

which is a cubic polynomial in \( s \), and, hence in \( x \), that fulfilled four interpolation conditions, two of those on function values and the rest two on the possibly unknown derivative values. The raw and synchronized signals have been depicted in Figs. 3 and 4, respectively. As shown, the signal lengths which are not equal in raw ones have been precisely synchronized after implementing of P.C.H.I.

2.2.2. Wavelet packet analysis

Feature extraction, aimed to take out certain characteristics from original measured signals, plays the vital role in neural networks. In simple words, badly implemented feature extraction or improper features will be led to poor classification results even using the potential classifiers. Accordingly, special attention has been paid to the selection of the feature vector or input layer in ANNs.

Wavelet analysis, which is mainly categorized into continuous and discrete types, includes transforming a signal from the time
domain into time–frequency domain. Wavelet transform (WT) offers a satisfactory time and poor frequency resolution at high frequencies and a satisfactory frequency and poor time resolution at low frequencies (Peng et al., 2005). It is used to detect transient components because it is able to simultaneously impart time and frequency structures.

Continuous wavelet transform (CWT), a function of two parameters entitled “translation and scale”, bears a remarkable data redundancy for the signal or function to be analyzed. By skipping several steps rather than continuously adjusting two parameters, in other words, by reducing the volume of calculations, discrete wavelet transform (DWT) has been defined to analyze the signals with a smaller set of scales and specific number of translations at each scale (Soman & Ramachandran, 2004), as efficient as CWT with identical accuracy. In DWT, the signals are decomposed into a hierarchical structure of detail and approximations at limited levels as follows:

\[ f(t) = \sum_{i=1}^{N} D_i(t) + A_j(t) \]  

where \( D_i(t) \) denotes the wavelet detail and \( A_j(t) \) stands for the wavelet approximation at the \( j \)th level.

Wavelet packet analysis (Avci, Turkoglu, & Poyraz, 2005), a generalization of wavelet decomposition offering a richer range of possibilities for signal analysis, is based upon DWT. Every wavelet detail component is further decomposed to get their own

<table>
<thead>
<tr>
<th>Table 1</th>
<th>GA parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of generation</td>
<td>200</td>
</tr>
<tr>
<td>Population</td>
<td>30</td>
</tr>
<tr>
<td>Chromosome length</td>
<td>12</td>
</tr>
<tr>
<td>Selection operator</td>
<td>Roulette</td>
</tr>
<tr>
<td>Fitness normalization</td>
<td>Rank</td>
</tr>
<tr>
<td>Elitism</td>
<td>1</td>
</tr>
<tr>
<td>Crossover</td>
<td>Pc = 0.8, two-point, uniform</td>
</tr>
<tr>
<td>Mutation</td>
<td>Pm = 0.1, Gaussian, mean = 0.0, std = 1.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2</th>
<th>GA variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable name</td>
<td>Range</td>
</tr>
<tr>
<td>Daubechies order</td>
<td>DB2–DB20</td>
</tr>
<tr>
<td>Decomposition level</td>
<td>3–6</td>
</tr>
<tr>
<td>Number of the hidden-layer neurons</td>
<td>10–35</td>
</tr>
</tbody>
</table>
approximation and detail components as shown in Fig. 5. Wavelet packet includes a set of linearly combined usual wavelet functions. Therefore, wavelet packet inherits the attributes of their corresponding wavelet functions such as orthonormality and time–frequency localization. In this research, after synchronizing the vibration signals, wavelet packet has been used to pre-process the signals. A wavelet packet is a function with three indices of integers \(i, j, k\) which are the modulation, scale and translation parameters, respectively,

\[ w_{ij}^k(t) = 2^{j/2} \sqrt{\sum_{k} h(k) \psi(2^j t - k)} \quad i = 1, 2, 3, \ldots \]  \(6\)

The wavelet functions \(\psi^j\) are determined from the following recursive equations:

\[ \psi^{2j}(t) = \sqrt{2} \sum_{k} h(k) \psi(2^j t - k) \]  \(7\)

\[ \psi^{2j+1}(t) = \sqrt{2} \sum_{k} g(k) \psi(2^j t - k) \]  \(8\)

The original signal \(f(t)\) after \(j\) level of decomposition are defined as follows:

\[ f(t) = \sum_{i=1}^{2j} f_i^j(t) \]  \(9\)

while the wavelet packet component signal \(f_i^j(t)\) are stated by a linear combination of wavelet packet functions \(\psi_{ij}^k(t)\) as follows:

\[ f_i^j(t) = \sum_{k} c_{ij}^k(t) \psi_{ij}^k(t) \]  \(10\)

where the wavelet packet coefficients \(c_{ij}^k(t)\) are calculated by:

\[ c_{ij}^k(t) = \int f(t) \psi_{ij}^k(t) dt \]  \(11\)

providing that the wavelet packet functions satisfy the orthogonality:

\[ \psi_{ij}^m(t) \psi_{ij}^n(t) = 0 \quad \text{if} \quad m \neq n \]  \(12\)

In this paper, standard deviation of wavelet packet coefficients (Rafiee et al., 2007) has been proposed to identify the faults as the feature vector for ANN training.

The wavelet-based techniques are absolutely dependent upon the mother wavelet function. Ingrid Daubechies invented what is called ‘compactly supported orthonormal wavelets’– thus making discrete wavelet analysis practical. In machine condition monitoring, Daubechies family functions are often selected for signal analysis and synthesis arbitrarily by trial and error (Wang and McFadden, 1995; Samanta and Al-Balushi, 2003; Tse et al., 2004; Liu, 2005; Peng et al., 2005; Kar and Mohanty, 2006; Saravanah et al., 2007; Rafiee et al., 2007. In other words, there is no computational logic behind the selection of Daubechies order which is therefore considered to be optimized for fault classification in this research. It is of paramount significance to note that calculating of
the wavelet coefficients is directly dependent on the shape of the mother wavelet such that the correlation between mother wavelet and signal is calculated as wavelet coefficients. Consequently, the standard deviation of wavelet coefficients (Rafiee et al., 2007) is among ideally suited features for fault detection and classification.

As demonstrated in Fig. 6, the time and frequency contents of the signals have not been lost with down-sampling of the lengthy signals into 16 sub-signals.

In Daubechies family functions, the difference between two wavelet functions with adjacent order (e.g. DB5 and DB6) is slight and they are similar to each other as depicted in Fig. 1. Therefore, the research has been focused more on a novel technique not only to eliminate the mentioned shortcoming but also to acquire the more efficient ANN-based system by proposing an accurate feature vector.

2.3. Artificial neural networks

Artificial neural networks (Haykin, 1998) or parallel-distributed-processing systems are made of simple processing units, called neurons which are a rough imitation of architecture of biological nervous system. ANNs have been shown to be suited to model highly complex and nonlinear phenomena. ANNs are also useful in machine condition monitoring since they can learn the system’s normal operating conditions and determine if incoming signals are significantly different as it is handy when there is lots of training data available and can manage the imprecision of condition monitoring process thanks to adaptability dynamism. Unlike the classic digital-processing techniques, ANNs possess the ability to perform parallel processing of data, cope with noisy data and adapt to different circumstances.

The most popular neural network is the multi-layer perceptron, which is a feed-forward network and frequently exploited in fault detection and diagnosis systems, which has found an immense popularity in machine condition monitoring applications as it constitutes more than 90% (Bartelmus, Zimroz, & Batra, 2003) of the current ANNs in the field, and therefore, in this study, has been used as the base of failure classifier system. The feed-forward network first learns through already existing data by iteratively changing its weights using the ‘back propagation’ of error algorithm which is performed by iteratively seeking to minimize the root-mean-square-error (RMSE) using the gradient search technique (Kwak & Ha, 2004).

An important network design issue is the selection and implementation of the network configuration. In the proposed system, the number of nodes in hidden layer, usually determined by trial-and-error methods in several papers, has been optimized using GA to provide the maximum throughput. Hyperbolic tangent sigmoid transfer function was applied to all net layers with Resilient Backpropagation training algorithm (Riedmiller & Braun, 1993). A matrix of a data set with 800 processed elements, comprising $5 \times 4 \times 40$ sampled data for every four gear conditions.

![Wavelet coefficients in one sample broken-tooth signal in level 4](image_url)

Fig. 12. Wavelet packet coefficients in one sample broken-tooth signal in level 4 (decomposition tree and standard deviation of the coefficients in title).
and five gearbox configurations, was applied to the network as the feature vector (Rafiee et al., 2007). The initial weights were, in addition, obtained randomly in a range of \((-10, 10)\). Error function was chosen to be least mean square. In many applications, training the network to perform fault detection system is performed off-line (Chow & Sharpe, 1993). The obtained signals for 10 configurations were divided into two groups; each used for training and testing of the proposed ANN. Fig. 7 depicts schematically the utilized ANN structure.

2.4. Genetic algorithm

Genetic algorithms have become popular following Holland’s work (Holland, 1975). GAs consisting of continuous and binary forms are designed to efficiently search huge, non-linear, discrete and poorly understood search spaces, where expert knowledge is scarce or difficult to model and traditional optimization techniques has failed (He et al., 2001). Typically, a simple GA consists of three operations: (1) parent selection, (2) crossover, and (3) mutation.

In this research, Roulette wheel selection scheme has been applied among the selection operators (Wong & Nandi, 2004). Two-point crossover was used for each chromosome of the chromosome-pair having a 50% chance of selection, the two parents selected for crossover exchange information lying between two randomly generated points within the binary string.

In addition that most of the successful applications of GAs (Tang, Man, Kwong, & He, 1996) is greatly dependent on finding a suitable method for encoding the chromosome, the creation of a fitness function to rank the performance of a specific chromosome is also of paramount importance for the success of the training process. The genetic algorithm rates its own performance around that of the fitness function; consequently, if the fitness function does not adequately take account of the desired performance features, the genetic algorithm is unable to meet the deserved requirements of the user.

The proposed chromosome includes five genes for Daubechies mother wavelet function, two genes for decomposition level of signals, and five genes pertain to the number of neurons of hidden

Fig. 13. Standard deviation of wavelet packet coefficients for 90 segmented sample signals in each condition (decomposition level: 4, gearbox conditions: normal gearbox, slight-worn, medium-worn, broken-tooth).
layer which is depicted in Fig. 8. Furthermore, the following linear equation was used as the fitness function $F$.

$$F = \alpha \cdot p + \beta \cdot t$$

(13)

where $\alpha$ and $\beta$ are the weights, $p$ and $t$ denote network performance and training time, respectively. In general, the network performance which is directly related to the accuracy of the network outweighs the training time constraint due to the significance of the system accuracy and off-line attribute of the network in the fault detection and diagnosis systems. Therefore, to optimize the system, network performance has more impact on fitness function than training time; the weights were applied where $\alpha$ was assumed to be 0.7 vs. $\beta$ which was assigned to 0.3. As the more training time has the more negative impact on the whole fault identification system negative sign has been considered for $\beta$ in fitness function. In simple words, the network performance and training time have inverse proportion on the fitness function. Depending on different case studies, these values for constant coefficients $\alpha$ and $\beta$ may deliberately be selected. For example, perhaps the training time may not be very important for a fault identification system having offline training process. Accordingly, the weight of training time can verge to be zero in that circumstance.

To sum up, GA was used to optimally search Daubechies order, decomposition level of the signals, and the number of neurons in hidden layer. Fig. 9 depicts the whole scheme of the fault identification system. Tables 1 and 2 explore, respectively, the parameters and variables of the GA exploited in the research.

3. Results and discussion

In this paper, by running GA, DB11, level 4 and 14 neurons have been selected as the best values for Daubechies order, decomposition level, and the number of nodes in hidden layer, respectively. Decomposition tree of wavelet packet has been depicted in Fig. 10 and related wavelet coefficients of two sample signals for normal gearbox and broken-tooth gear have been illustrated in Figs. 11 and 12. With close scrutiny to the amplitude of the wavelet packet coefficients, it has been shown that there is a considerable correlation between DB11 and the failure condition (broken-tooth gear). The values of standard deviation of wavelet packet coefficients in 16 sub-signals are mostly higher than those of the normal one. On the other hand, as shown in Fig. 13, the fluctuations of the standard deviations in different sample segmented signals are considerable and this is the most appropriate factor for training the neural networks. Stability and convergence of the network are directly dependent on the distinguished feature vector with enough samples in ample ranges. As a matter of fact, the range of the samples in feature vector is more important than the number of samples in training that...
To validate the estimated values, the performance of the neural network with three above-mentioned parameters has been also calculated manually with several trials and errors. Related results which are matched with the estimation of GA have been shown in Table 3. Although, the network structure is compact in decomposition level of 3 (due to 8-element network input), obviously no satisfactory results were obtained. In spite of the higher performance and better results, the decomposition level of 5 leads to a larger number of neurons of the hidden layer (32-element network input) which leads to the complexity of the system. Therefore, fourth decomposition level was determined to be the most appropriate level for this case study because of the better outcomes than those of the others and the introduction of a well-formed network structure. The data in Table 3 has been calculated for each case without genetic algorithm. In addition, the optimal number of the neurons of the hidden layer was determined using several trials and errors runs for each case in Table 3.

Hence, the end results obtained using genetic algorithm not only does the high capabilities of this algorithm prove in order to find the optimal solution but also verifies the soundness of the parameters for the adjustment of the algorithm presented in Table 1.

Table 4 puts forward the results of a diagnostics system without using GA and manually with 14 neurons in hidden layer with different function orders and decomposition levels to better clarify the aptness of the results of the previous case (DB11 with a decomposition level of 4).

---

**Table 3**

Various combinations of the three optimized parameters and related results

<table>
<thead>
<tr>
<th>Wavelet function order</th>
<th>Decomposition level</th>
<th>No. of hidden-layer neurons</th>
<th>Samples for testing</th>
<th>Slight-worn</th>
<th>Medium-worn</th>
<th>Broken-tooth</th>
<th>Faultless</th>
<th>Best net structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>DB 8</td>
<td>3</td>
<td>13</td>
<td>5 x 4 x 75</td>
<td>94</td>
<td>90</td>
<td>98.06</td>
<td>96.66</td>
<td>8:10:4</td>
</tr>
<tr>
<td>DB 9</td>
<td>3</td>
<td>15</td>
<td>5 x 4 x 75</td>
<td>93.33</td>
<td>92.66</td>
<td>97.13</td>
<td>92.33</td>
<td>8:12:4</td>
</tr>
<tr>
<td>DB 10</td>
<td>3</td>
<td>13</td>
<td>5 x 4 x 75</td>
<td>95.66</td>
<td>93.33</td>
<td>97.33</td>
<td>95.66</td>
<td>8:10:4</td>
</tr>
<tr>
<td>DB 11</td>
<td>3</td>
<td>13</td>
<td>5 x 4 x 75</td>
<td>98.53</td>
<td>97.66</td>
<td>99.33</td>
<td>97.66</td>
<td>8:10:4</td>
</tr>
<tr>
<td>DB 12</td>
<td>3</td>
<td>15</td>
<td>5 x 4 x 75</td>
<td>97.33</td>
<td>96.66</td>
<td>98.2</td>
<td>96.73</td>
<td>8:12:4</td>
</tr>
<tr>
<td>DB 13</td>
<td>3</td>
<td>15</td>
<td>5 x 4 x 75</td>
<td>93.33</td>
<td>95.33</td>
<td>98.06</td>
<td>99</td>
<td>8:12:4</td>
</tr>
<tr>
<td>DB 8</td>
<td>4</td>
<td>17</td>
<td>5 x 4 x 75</td>
<td>94.33</td>
<td>100</td>
<td>99.66</td>
<td>96</td>
<td>16:14:4</td>
</tr>
<tr>
<td>DB 9</td>
<td>4</td>
<td>19</td>
<td>5 x 4 x 75</td>
<td>92</td>
<td>98.46</td>
<td>99.13</td>
<td>94.66</td>
<td>16:14:4</td>
</tr>
<tr>
<td>DB 10</td>
<td>4</td>
<td>17</td>
<td>5 x 4 x 75</td>
<td>100</td>
<td>98.66</td>
<td>100</td>
<td>93.33</td>
<td>16:14:4</td>
</tr>
<tr>
<td>DB 11</td>
<td>4</td>
<td>14</td>
<td>5 x 4 x 75</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>16:13:4</td>
</tr>
<tr>
<td>DB 12</td>
<td>4</td>
<td>15</td>
<td>5 x 4 x 75</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>16:13:4</td>
</tr>
<tr>
<td>DB 14</td>
<td>4</td>
<td>14</td>
<td>5 x 4 x 75</td>
<td>97.86</td>
<td>100</td>
<td>99</td>
<td>100</td>
<td>16:14:4</td>
</tr>
<tr>
<td>DB 8</td>
<td>5</td>
<td>24</td>
<td>5 x 4 x 75</td>
<td>97.33</td>
<td>96.86</td>
<td>100</td>
<td>98.66</td>
<td>32:21:4</td>
</tr>
<tr>
<td>DB 9</td>
<td>5</td>
<td>25</td>
<td>5 x 4 x 75</td>
<td>90.66</td>
<td>92</td>
<td>99.4</td>
<td>97.53</td>
<td>32:22:4</td>
</tr>
<tr>
<td>DB 10</td>
<td>5</td>
<td>25</td>
<td>5 x 4 x 75</td>
<td>96</td>
<td>100</td>
<td>100</td>
<td>98.66</td>
<td>32:22:4</td>
</tr>
<tr>
<td>DB 11</td>
<td>5</td>
<td>23</td>
<td>5 x 4 x 75</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>32:21:4</td>
</tr>
<tr>
<td>DB 12</td>
<td>5</td>
<td>23</td>
<td>5 x 4 x 75</td>
<td>100</td>
<td>96</td>
<td>100</td>
<td>100</td>
<td>32:23:4</td>
</tr>
<tr>
<td>DB 13</td>
<td>5</td>
<td>23</td>
<td>5 x 4 x 75</td>
<td>90.66</td>
<td>100</td>
<td>94</td>
<td>96.66</td>
<td>32:23:4</td>
</tr>
</tbody>
</table>

---

**Table 4**

Various arrangements of the three optimized parameters with a constant number of neurons and the related results

<table>
<thead>
<tr>
<th>Wavelet function order</th>
<th>Decomposition level</th>
<th>No. of hidden-layer neurons</th>
<th>Samples for testing</th>
<th>Slight-worn</th>
<th>Medium-worn</th>
<th>Broken-tooth</th>
<th>Faultless</th>
<th>Net structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>DB 8</td>
<td>4</td>
<td>14</td>
<td>5 x 4 x 75</td>
<td>97</td>
<td>93.33</td>
<td>100</td>
<td>96.66</td>
<td>16:11:4</td>
</tr>
<tr>
<td>DB 9</td>
<td>4</td>
<td>14</td>
<td>5 x 4 x 75</td>
<td>96.33</td>
<td>97.33</td>
<td>100</td>
<td>100</td>
<td>16:11:4</td>
</tr>
<tr>
<td>DB 10</td>
<td>4</td>
<td>14</td>
<td>5 x 4 x 75</td>
<td>98.66</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>16:11:4</td>
</tr>
<tr>
<td>DB 11</td>
<td>4</td>
<td>14</td>
<td>5 x 4 x 75</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>16:11:4</td>
</tr>
<tr>
<td>DB 12</td>
<td>4</td>
<td>14</td>
<td>5 x 4 x 75</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>16:11:4</td>
</tr>
<tr>
<td>DB 13</td>
<td>4</td>
<td>14</td>
<td>5 x 4 x 75</td>
<td>96</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>16:11:4</td>
</tr>
<tr>
<td>DB 8</td>
<td>5</td>
<td>14</td>
<td>5 x 4 x 75</td>
<td>89.66</td>
<td>91</td>
<td>96.66</td>
<td>95</td>
<td>32:11:4</td>
</tr>
<tr>
<td>DB 9</td>
<td>5</td>
<td>14</td>
<td>5 x 4 x 75</td>
<td>94.33</td>
<td>89.66</td>
<td>95.33</td>
<td>96.66</td>
<td>32:12:4</td>
</tr>
<tr>
<td>DB 10</td>
<td>5</td>
<td>14</td>
<td>5 x 4 x 75</td>
<td>95.06</td>
<td>97</td>
<td>98.66</td>
<td>100</td>
<td>32:11:4</td>
</tr>
<tr>
<td>DB 11</td>
<td>5</td>
<td>14</td>
<td>5 x 4 x 75</td>
<td>95</td>
<td>98.66</td>
<td>96.66</td>
<td>98</td>
<td>32:11:4</td>
</tr>
<tr>
<td>DB 12</td>
<td>5</td>
<td>14</td>
<td>5 x 4 x 75</td>
<td>95.66</td>
<td>94</td>
<td>96.33</td>
<td>96.66</td>
<td>32:11:4</td>
</tr>
<tr>
<td>DB 13</td>
<td>5</td>
<td>14</td>
<td>5 x 4 x 75</td>
<td>93.66</td>
<td>91.66</td>
<td>96.66</td>
<td>94</td>
<td>32:11:4</td>
</tr>
</tbody>
</table>
4. Conclusion

An intelligent gear fault identification system was developed and implemented to pinpoint the gearbox faults; Three parameters were recognized to be of paramount significance to be optimized using GA i.e. Daubechies wavelet function order, decomposition level, and number of neurons in hidden layer of ANN, which play a key role in size and performance of the network and the soundness of the feature vector. The Feature vector was also obtained from optimized standard deviation of wavelet packet coefficients acquired from DB11 at fourth level of decomposition following a piecewise cubic hermite interpolation for synchronizing. Eventually, a MLP network with well-formed and optimized structure (16:14:4) and remarkable accuracy was presented providing the capability to identify slight-worn, medium-worn and broken-tooth of gears faults perfectly in 100% of the five tested configurations of gearbox. By the proposed system which has been elaborately depicted in Fig. 15, the drawback of implementing mother wavelet function through trial-and-error based methods has been also improved for fault classification purposes.

Acknowledgements

This work described in this paper was partially supported by Vibration and Modal Analysis Lab at University of Tabriz and partially supported by a grant from the Research Grants Council of Hong Kong Special Administrative Region, China (Project no. CityU 120506). The authors would like to write in memoriam of two dedicated mentors, Professor Vahhab Pirouzpanah and Professor Samad Nami Notash who devoted more than four decades to enlighten a myriad of students. They would also like to thank the anonymous reviewers for spending their valuable time to review the current research.

References

Liu, B. (2005). Selection of wavelet packet basis for rotating machinery fault

algorithm for automatic fault detection in HVAC systems. *Applied Soft

NASA conference Publication 10189.

condition monitoring and fault diagnostics. *Mechanical Systems and Signal
Processing, 18*, 199–221.

its application in vibration signal analysis. *Journal of Sound and Vibration, 286*,
187–205.

Hilbert–Huang transform and wavelet transform: Application to fault
974–988.

monitoring of a gearbox using artificial neural network. *Mechanical Systems and

backpropagation learning: The RPROP algorithm. In *Proceedings of the IEEE
international conference on neural networks*, San Francisco.

Samanta, B. (2004). Gear fault detection using artificial neural networks and
support vector machines with genetic algorithms. *Mechanical Systems and Signal
Processing, 18*, 625–644.

diagnostics of rolling element bearings using time-domain features. *Mechanical

and support vector machines with genetic algorithm for bearing fault detection.
*Engineering Applications of Artificial Intelligence, 16*, 657–665.

Saravanan, N. et al. (2007). A comparative study on classification of features by SVM
and PSVM extracted using Morlet wavelet for fault diagnosis of spur bevel gear

Prentice-Hall of India.


recovering multiple sources of vibration signals in machine fault diagnosis.


using artificial neural network and genetic algorithm. *Signal Processing, 84*,
351–365.

Wu, J. D., Chang, P. H., Chang, Y. W., & Shiao, Y. J. (2008). An expert system for fault
diagnosis in internal combustion engines using probability neural network.

Xie, L. (2004). The application of the conditional moments analysis to gearbox
fault detection-a comparative study using the spectrogram and scalogram.