IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 23, 2019, accepted November 8, 2019, date of publication November 20, 2019, date of current version December 5, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2954704

Demodulation Band Optimization in Envelope Analysis for Fault Diagnosis of Rolling Element Bearings Using a Real-Coded Genetic Algorithm

VIGNESHWAR KANNAN, HUAIZHONG LI, AND DZUNG VIET DAO
School of Engineering and Built Environment, Griffith University, Gold Coast, QLD 4222, Australia
Corresponding author: Huaizhong Li (h.li@griffith.edu.au)

ABSTRACT Envelope analysis is a commonly used technique in fault diagnosis of rolling element bearings. The selection of a suitable frequency band for demodulation in envelope analysis has traditionally relied on the expertise of diagnosis technicians. The manual selection does not always give the best possible results in revealing the defect frequencies. To overcome this problem, a new demodulation band optimization approach is proposed which is based on a real-coded genetic algorithm with a novel fitness function and crossover selection process. The fitness function uses the ratio between fault frequency peaks and the maximum peak not corresponding to defects in the envelope spectrum. The crossover selection process uses the triangle series method to divide the probability of individuals in the population based on the fitness score obtained. The proposed method is assessed using vibration signals from two different rotor-bearing systems, i.e., a bearing testrig with seeded defects and the Case Western Reserve University bearing dataset. For all the cases, the method can find the optimized demodulation bands successfully for bearing fault detection. The method is further benchmarked with a well-established fast kurtogram approach which proves the effectiveness and superior capability of the developed algorithm, though the computational complexity needs improvement in future work.

INDEX TERMS Condition monitoring, fault diagnosis, genetic algorithms, ball bearings.

I. INTRODUCTION

Rolling element bearings (REB) are widely used in rotating machinery which are essential for the operation of many industries [1]. Bearing defects may occur due to inadequate lubrication, external contaminants, incorrect operating conditions, etc. Damage of bearing components account for about 45% of rotating machinery failures [2]. This damage could lead to the halt of production in an industrial setting causing significant economic losses. The operation of machinery with damaged bearings may also pose a risk to those present in the immediate vicinity. It is vital for the rotor-bearing system to be in good condition to ensure proper functioning of the machine.

There has been extensive research in the development of condition monitoring technologies for the detection of bearing defects. The most widely used condition monitoring technique is based on vibration analysis where accelerometers are used to monitor bearing vibrational behavior. This has been preferred by many industries due to the ability of applying various signal processing techniques to extract fault information and also for its immediate reaction to sudden fluctuations [3]. Of the many signal processing techniques, envelope analysis or High Frequency Resonance Technique (HFRT) has been used prominently for several decades [4]. The periodic contact of the defective component with other surfaces during operation typically generates impulses at a higher frequency compared to other machine vibrations. Envelope analysis allows for the separation of these vibration components from the rest of the signal by demodulating a band or range of frequencies that correspond to the impulse [5]. The spectrum generated through this process will indicate a spike in amplitude in the frequencies corresponding to the periodic impulse.

Selection of the optimal demodulation frequency band is considered as a significant and challenging step in bearing fault diagnosis [6]. Technical expertise is traditionally required for the appropriate selection. This is because the

The associate editor coordinating the review of this manuscript and approving it for publication was Ming Luo.

parameters can greatly influence the envelope spectrum's indication of the fault and can even produce misleading results if wrongly selected [7]. McFadden and Smith [5] presented several common methods that were used in selecting these parameters. One involved the selection of a resonance of interest as the central frequency and ±5% of this was set as the bandwidth. Another method involved adjusting the bandwidth to cover an entire dominant resonance of interest. It is stated however, that despite the widespread use of these methods, there was no indication that the parameter selection was ideal [5].

Many attempts have been reported to find an optimal demodulation band selection technique for envelope analysis. Bechhoefer and Menon [8] studied an incident involving the oil cooler fan bearing of a helicopter and manually tested different envelope windows in the frequency domain. The search for an optimal envelope window was conducted by modifying the lower frequency and the bandwidth, within a specified range and increment. Boškoski and Urevc [9] presented a fault detection methodology where healthy bearing data was not available. Spectral kurtosis was first used to find the likelihood of a fault, then envelope analysis was performed in the bandpass filter maximizing spectral kurtosis value to isolate the fault. Bechhoefer *et al.* [10] used spectral kurtosis for the window parameter selection and an average energy algorithm to measure the performance of the envelope window selected. Nevertheless, these methods were usually time consuming and still required knowledge and experience in condition monitoring and fault diagnosis. The popularity of the use of spectral kurtosis led to the development of the kurtogram which allowed for its visual representation in terms of frequency and frequency resolution [11]. This method, however, was seen as time consuming due to all the possible combinations of frequency resolutions and center frequencies that could be used. This was accounted for through the use of the fast kurtogram technique proposed by Antoni [12] which has since been used as a common benchmarking tool by many researchers in bearing fault detection. The autogram method was also used for further enhancement of the kurtogram for bearing diagnosis [13]. Instead of the filtered time signal kurtosis, the squared envelope of demodulated signal is taken and the unbiased autocorrelation is calculated from which kurtosis is found for a visual representation. Another recent method for the selection of an optimal frequency band is the Distcsgram which is not as influenced by interferences as the fast kurtogram [14]. A downside however is that as defect locations are not known, it could be computationally taxing for a trial and error process determining fault location.

The trend of using some form of metric to measure performance gained popularity and may have been an inspiration for the use of metaheuristic optimization algorithms in fault detection. This includes genetic algorithm (GA) in which the functioning is largely inspired by nature. GAs generate a population of solutions which are assessed based on a criterion set to determine the best in that generation. The algorithm will then continue searching for better solutions in following generations by exchanging parameter information from the most successful solutions to obtain better result while also introducing mutations occasionally to maintain diversity among the population. This process could potentially select better frequency ranges than what is obtained through a brute force method using a multilevel filterbank as a combination of consecutive bands can also be analyzed. Zhang and Randall [15] proposed a method in which fast kurtogram was used to obtain a rough estimate of parameters which were then further optimized by a genetic algorithm for the selection of an ideal bandpass filter. Kang *et al.* [16] also used a genetic algorithm for the selection of an optimal demodulation band, however, real numbers from 0 to 1 were used for the parameters instead of binary values. The defect severity was measured as a ratio of residual-to-defect frequency components and therefore was used as the fitness score in the algorithm [16]. A majority of studies using genetic algorithm for demodulation band selection programmed their genes as binary components as opposed to real numbers. Gaffney et al. [17] compared a binary to real-coded genetic algorithm and stated that real-coded genetic algorithms were generally preferred in applications where the parameter space variables are continuous. With real coding, less storage is needed and a more accurate representation of the optimized solution can be obtained [18]. In order to implement a genetic algorithm which uses real-coding, some adjustments will have to be made to the crossover and mutation process from the more canonical approach.

This paper presents a study to utilize a real-coded genetic algorithm in the fully automated optimization of the demodulation band for fault detection in REB using envelope analysis. A novel fitness function and crossover selection process is introduced. The proposed method is tested on vibration signals from two different rotor-bearing systems under different conditions in order to assess its capability for fault detection. The computational burden of the algorithm is inspected through assessing the time taken to determine the optimal frequency band in each case. The proposed approach is also benchmarked by comparing with the results obtained using the fast kurtogram method.

II. METHODOLOGY

A. BEARING DIAGNOSIS USING ENVELOPE ANALYSIS

Envelope analysis generally allows for periodic impulses in REB to be better visualized [19]. It works by first obtaining the frequency spectrum for the raw signal using a method like FFT. The frequencies of impact of the bearing defect, noise generated, and structural resonance are all shown for the portion of the raw signal analyzed in the frequency domain. From this, a frequency range is chosen for the amplitude demodulation process by using a bandpass filter. Envelope analysis demodulation has been achieved using various techniques with Hilbert transform being the most widely adopted approach. The Hilbert transform of a signal as seen in (1) is the representation of phase shifting Fourier components on its

frequency spectrum by $\pm\pi/2$ [20]. The technique is generally better than other filters such as analogue and real-time digital filters [4].

$$\hat{x}(t) = \frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{x\tau}{t-\tau}d\tau \tag{1}$$

The band selected for demodulation is typically of a higher frequency where the structural resonance amplifies the impulses of the defect present [21]. Once the analytic time signal is obtained, the modulus is calculated. The envelope spectrum can then be obtained for analysis of diagnostic information by performing Fourier transform on the signal once again. HFRT is now considered a common benchmarking method in the fault diagnosis of REB. Depending on the selection of the demodulation band, the envelope spectrum indicates a high amplitude relative to the rest of the spectrum in certain frequencies. The type of defect is usually determined by checking for a correlation of the prominent peaks with theoretical defect frequencies (TDF) and its harmonics in the spectrum. Equations (2)-(5) can be used to calculate the TDF using the geometry of the bearing and its rotational frequency where n is the number of rolling elements, d is the rolling element diameter, D is the pitch diameter of bearing, θ is the contact angle, and f_r is the shaft frequency [4]. For special cases where both inner race and outer race are rotating, f_r is the relative speed difference between the inner and outer race.

Ball pass frequency, outer race (BPFO):

$$BPFO = \frac{nf_r}{2}\left(1 - \frac{d}{D}\cos\theta\right) \tag{2}$$

Ball pass frequency, inner race (BPFI):

$$BPFI = \frac{nf_r}{2}\left(1 + \frac{d}{D}\cos\theta\right) \tag{3}$$

Ball spin frequency (BSF):

$$BSF = \frac{Df_r}{2d}\left(1 - \left(\frac{d}{D}\cos\theta\right)^2\right) \tag{4}$$

Fundamental train frequency (FTF):

$$FTF = \frac{f_r}{2}\left(1 - \frac{d}{D}\cos\theta\right) \tag{5}$$

In comparison to raw vibration signals and the frequency spectrum, the envelope spectrum provides more information regarding bearing defects. Although defect frequencies can still be identified from the frequency domain, the envelope spectrum allows for the detection of defects in conditions where the fault frequencies may be buried in background noise or the shaft speed is too low to detect the frequencies [22]. In order to do this however, an appropriate frequency band needs to be selected for the demodulation process.

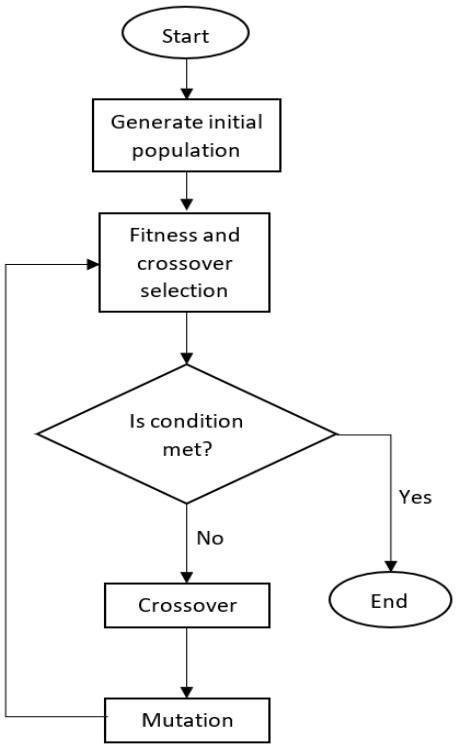


FIGURE 1. Flowchart showing the general operation of a typical genetic algorithm.

B. GENETIC ALGORITHM

The genetic algorithm is a widely adapted technique for optimization of complicated parameter selection tasks. It is an evolutionary algorithm that mimics the process of natural selection in order to find an optimized solution to a problem [23]. A typical genetic algorithm's operation can be seen in Fig. 1 and is further explained as follows.

Firstly, an initial population is randomly generated where each individual represents different combinations of parameters. The number of parameters in each individual is fixed and these combinations are called chromosomes. The parameters for optimization are called genes and are the variables that make each chromosome distinct by producing different solutions. The viability of each chromosome in solving the problem can be evaluated using a fitness function. The fitness function is used to assess each of the chromosomes in the population against a certain criterion and find the most suitable solution to the problem for that generation. The chromosomes with a good fitness score are given a higher probability of selection for crossover compared to those with a bad score. Crossover is the exchange of genetic information from two parent chromosomes selected based on their fitness score to produce offspring. Genetic crossover is typically set to occur between some of the more fit chromosomes in the population to produce offspring that possess a combination of genes that have achieved a good fitness score with other genes. The new offspring produced from the crossover of

parents replaces other chromosomes that have a low fitness score in the population. This allows the algorithm to learn by removing the chromosomes with a low fitness from the population allowing only the fittest to crossover. Examples of traditional crossover selection methods include single-point and two-point crossover as further explained in [24], [25]. After the crossover event occurs between multiple parent chromosomes, the new generation is once again evaluated with the fitness function for crossover. This process repeats either until convergence of the population, a set number of generations or if a condition specified is met. Convergence is when the population largely possesses the same genes meaning that new offspring may not have a significant effect on the fitness score of chromosomes in the following generations. Mutations, or the random alteration of parameters, are occasionally introduced to newly produced offspring as a means of introducing some diversity into the gene pool. This ensures that the population does not converge sooner than it should by reaching a solution that could perhaps be further optimized [26].

C. OPTIMIZATION OF BANDPASS FILTER FOR ENVELOPE ANALYSIS

A demodulation band is selected in the frequency domain for envelope spectrum analysis to better visualize the defect frequencies caused by the impact of the fault during operation of the testrig. The method utilizes genetic algorithms in the selection of this band for the analysis of the envelope spectrum. This is done by first generating an initial population of chromosomes for the genetic algorithm. In this study, the number of chromosomes was set to be 40 as this was a reasonable population size that would not be very time consuming to process and still have diversity among individuals. Each chromosome in the population has two genes corresponding to the upper and lower limit of the passband which are randomly assigned a value between the constraints specified. The constraints are shown in (6) and (7) as follows, where F_{max} is the maximum frequency possible on the spectrum, x_a is the first and x_b is the second gene of a chromosome. In such a way, the value of genes are represented in the form of real numbers instead of binary numbers.

$$0 \leq x_a \leq F_{\max} \tag{6}$$

$$x_a \leq x_b \leq F_{\max} \tag{7}$$

Although it is common to select a central frequency and the passband width as the parameters for optimization, the upper and lower limit of the passband were chosen with the constraints above as it naturally introduces a bias to the population reducing the likelihood of randomly selecting lower frequency bands. The algorithm can benefit from this bias as the peaks in the lower range of the frequency spectrum typically correspond to the vibrations produced at the frequency of the shaft speed where the desired fault information cannot be extracted.

The chromosomes in the population are used to compute the fitness score obtained by their respective genes. The fitness function implemented aims to maximize the amplitude of the fault frequency compared to other frequencies on the envelope spectrum. This is done by generating the envelope spectrum from each chromosome using Hilbert transform. Once the array of data is acquired for each demodulation band in the population, the fitness of that chromosome is calculated. Equation (8) below is used to calculate the fitness score (f_s) where A_{max} is the highest amplitude that corresponds with one of the TDFs achieved in the envelope spectrum. A tolerance is assigned when searching for peaks corresponding to TDFs to account for any slip that may occur causing a slight variation in recorded data. The term s_{max} is the maximum amplitude obtained in the envelope spectrum before the lower tolerance limit of A_{max}.

$$f_s = \frac{A_{\max}}{s_{\max}} \tag{8}$$

When the demodulation band selected is too narrow, s_{max} is typically higher than any defect frequencies present and the envelope spectrum's amplitude distribution appears positively skewed. The ratio of A_{max} and s_{max} is therefore chosen as the fitness function in order to avoid the phenomenon described and also maximize the defect frequency amplitude with respect to the rest of the signal.

The triangle number series principle is used to distribute the probability of selection of a chromosome in such a way that the fittest element has the highest probability for crossover and the probability lowers according to each chromosome's fitness rank. The fittest half of the population is carried into the next generation with a probability of crossover whereas the other half is given zero chance of crossover and is overwritten by offspring. This means 20 new chromosomes will be generated in each iteration. The probability is divided among the fittest half of the population by using the triangle number series as shown in the simple example below with a population size (S_{pop}) of 8. The triangle series number (n) respective to half the population size minus 1 is found using (9) and (10).

$$n = \frac{S_{pop}}{2} - 1 = \frac{8}{2} - 1 = 3 \tag{9}$$

$$T_n = \frac{n(n+1)}{2} \rightarrow T_3 = \frac{3(3+1)}{2} = 6 \tag{10}$$

$$P_x = \frac{\left(\left(\frac{S_{pop}}{2} - x\right) \times \frac{S_{pop}/2}{T_n}\right) + 1}{S_{pop}} \tag{11}$$

Equation (11) was written so that the sum of the numerator for the fittest half of the population would add up to the total population size. Note that x here denotes the rank of an individual's crossover probability. Table 1 shows an example of the crossover probability distribution among the chromosomes for a population size of 8. The fittest chromosome will have the highest probability for crossover, and the probability percentage will decrease the lower the chromosome's fitness compared to the rest of the population.

TABLE 1. Distribution of probability among a population of 8 chromosomes ordered from fittest to least fit.

Population	Crossover probability	
P_1	0.3750	
P_2	0.2917	
P_3	0.2083	
P_4	0.1250	
P_5	0.0417	Set to zero
P_6	−0.0417	
P_7	−0.1250	
P_8	−0.2083	
$\sum_{x=1}^{8} P_x$	1	

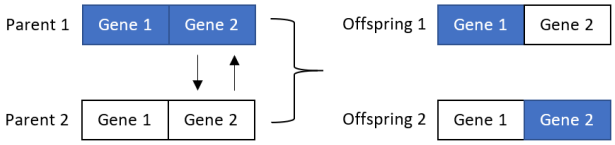


FIGURE 2. Crossover of genes to produce two offspring.

If the two chromosomes selected for crossover happen to be the same, the second parent is reselected with the same probabilities until a different chromosome is chosen. In each crossover that occurs, two offspring will be produced for the next generation replacing the least fit half of the chromosomes of the initial population. As there are only two genes in each chromosome, the algorithm is made so that two offspring are produced in each crossover allowing each gene of the parents an equal chance to carry over in the next generation as shown in Fig. 2.

Occasionally the gene corresponding to the lower limit parameter of the offspring is larger than that of the upper limit parameter for the passband resulting in an error. Although the occurrence of this is rare, a mutation is forced when this occurs replacing the second gene with a randomly assigned value between the first gene and the maximum frequency possible on the spectrum. In order to account for any biases this may create, the second offspring for the same parents has the mutation occur in its first gene to randomly select a value between zero and its second gene. Once all the offspring overwrite half the population with the lowest fitness score, the chromosomes of the population are re-evaluated based on the fitness function and the process is repeated.

A mutation probability is also set to introduce some diversity to the otherwise stagnant gene pool of the population. The mutation is introduced into the population during the crossover process and affects one of each offspring's genes. The first offspring's first gene and second offspring's second gene are mutated by replacing what should exist with a randomly generated value that falls within the constraints for the respective gene in the chromosome. After a certain point, a majority of the population will possess the same genes so the introduction of a mutation according to the mutation rate set will allow for some variance in the population. The ideal mutation rate varies from case to case. It is common to set it to a lower number as this provides a good balance between introducing diversity into the population and fully utilizing the genetic algorithm's ability to carry over useful genes into the next generation. A 10%, or 0.1, mutation rate was chosen for the purpose of the test as there are only two genes in each chromosome. The high mutation rate accounts for the fact that convergence will be reached sooner in a population with only two genes in a chromosome.

After a set number of iterations, the execution of the algorithm will terminate and output the most suitable combination of demodulation upper and lower band limits found. Alternatively, the algorithm can also be set to terminate if convergence is expected prior to running the set number of iterations. This is done by checking the fittest chromosome's two genes for change in magnitude. If no change is found for more than 30% of the set iterations, it can be assumed that the algorithm has reached a state where crossover of the fittest individuals led to the current population largely possessing the same genetic makeup. This would mean that there would be no significant variation in the optimal parameters identified by the algorithm. Mutations may have little to no effect on the population at this point and therefore it would be wise to automatically terminate the program and reduce computation time to arrive at the most optimal result possible.

III. RESULTS AND DISCUSSIONS

The proposed envelope analysis with the GA-based demodulation band optimization was implemented using MATLAB R2018a. The algorithm was assessed by using vibration signals obtained from a bearing testrig with seeded Outer Race Fault (ORF) and Inner Race Fault (IRF) bearings running at various shaft speed as described below.

A. EXPERIMENTAL SETUP

The testrig used for the experiment was a SpectraQuest Machinery Fault Simulator-Lite as shown in Fig. 3. An accelerometer, PCB Piezotronics model 352C65 with a sensitivity of 10.29 mV/m/s^2 was attached on the Y-axis of bearing housing 1 to measure the vibrations produced when the testrig was in operation. 5 second vibration signals were collected by using an ECON MI-7004 dynamic signal analyzer at a sampling frequency of 96 kHz. Envelope analysis with the proposed demodulation band selection approach was programmed using MATLAB and was executed on a DELL laptop with 8 GB RAM and an i5 8350U CPU.

Rexnord ER12KCL ball bearings with a seeded IRF or ORF were used in bearing housing 1 of the testrig for testing. The single localized defect seeded on the IRF bearing was 1.905 mm and the ORF bearing was 2.54 mm in diameter. The geometric parameters of the bearings are given in Table 2.

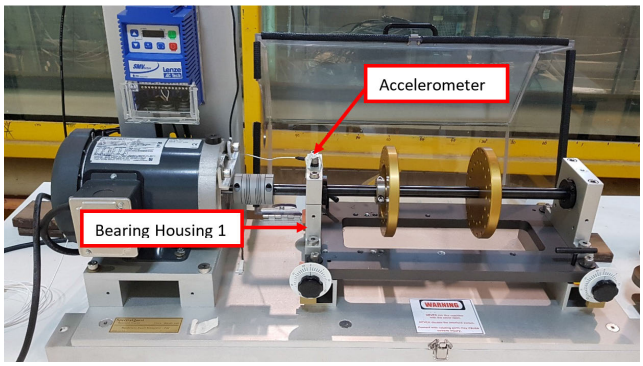


FIGURE 3. Experimental setup with SpectraQuest Machinery Fault Simulator-Lite.

TABLE 2. Geometric parameters of Rexnord ER12KCL ball bearings.

Pitch diameter (**D**)	33.5 mm
Rolling element diameter (**d**)	7.938 mm
Number of rolling elements (**n**)	8
Contact angle ($\boldsymbol{\theta}$)	0°

TABLE 3. Shaft frequencies and the corresponding theoretical defect frequencies (TDF). Unit: Hz.

Set shaft frequency	Actual shaft frequency	BPFO	BPFI	BSF	FTF
17	16.69	50.94	82.58	33.24	6.37
27	26.68	81.43	132.01	53.14	10.18
37	36.67	111.92	181.44	73.03	13.99
47	46.65	142.38	230.82	92.91	17.80
57	56.64	172.88	280.24	112.81	21.61

The testrig was set to run at shaft frequencies of 17 Hz to 57 Hz at intervals of 10 Hz. Note that the actual shaft speed was slightly lower than what was set. Table 3 lists the set shaft frequencies, actual shaft frequencies, and the corresponding TDFs calculated from (2)-(5) using the actual shaft frequencies.

B. DIAGNOSIS OF OUTER RACE FAULT

For the testing case of an ORF bearing at a set shaft frequency of 47 Hz, the envelope spectrum generated using the proposed method is shown in Fig. 4(a), where the optimal demodulation band was selected by the real-coded genetic algorithm. The envelope spectrum clearly shows the ORF as the amplitudes relating to the defect frequency (BPFO = 142.38 Hz) and its harmonics are large enough to be distinguished from other frequencies present. It demonstrated that the fitness function used in the genetic algorithm could successfully find the solution with the highest amplitude in the respective fault frequency. The automatically selected frequency band to generate the envelope spectrum is shown in Fig. 4(b). The selection of the frequency band was based on

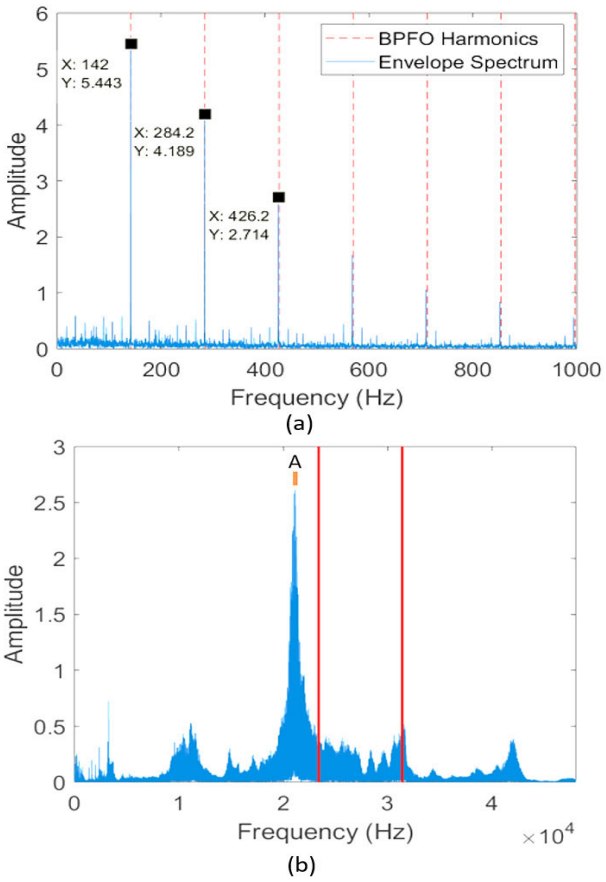


FIGURE 4. (a) Envelope spectrum and (b) frequency spectrum band selection for the ORF bearing at shaft speed of 47 Hz.

the evaluation made through the fitness function used in the real-coded genetic algorithm. The best individual in the final generation was used as the selected solution. It can be seen that the selected band is not coincident with a traditionally assigned band with a center at a structural resonance such as location A.

Fig. 5(a) shows the fitness score of the best individual in each generation. It can be seen that the score does not further increase after convergence at the 25th iteration. The highest fitness score achieved for this case using the proposed algorithm was approximately 9.2. Fig. 5(b) shows the variation of the selected demodulation band in each iteration. The passband selected was 23390.6 to 31384.4 Hz and it did not change after convergence. The fast convergence can be attributed to the design of the genetic algorithm where each chromosome in the population only has two genes. The diversity of the population was maintained by using a high mutation rate. Additionally, real-coded genetic algorithms are generally considered to be fast as they use less storage than a standard binary genetic algorithm.

C. DIAGNOSIS OF INNER RACE FAULT

The method was applied for an IRF bearing with a shaft speed of 47 Hz and the envelope spectrum obtained is shown

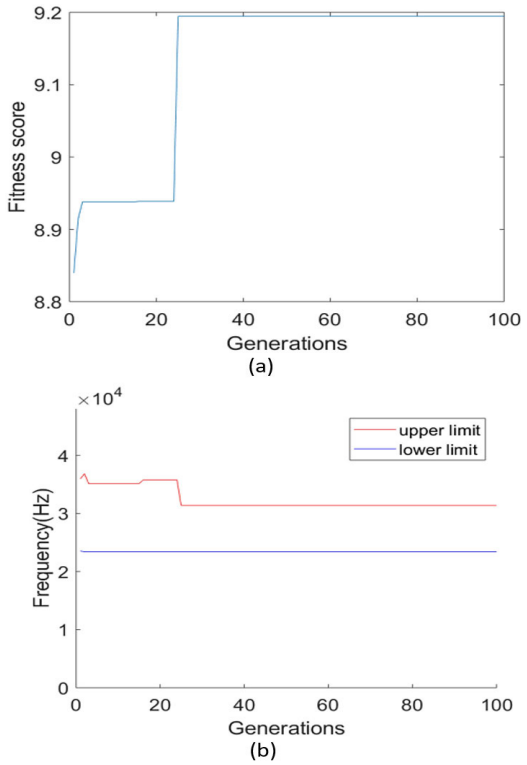


FIGURE 5. (a) Fitness score of the best individual and (b) best passband found within 100 iterations for an ORF bearing at a shaft speed of 47 Hz.

in Fig. 6(a). The amplitudes corresponding to the IRF frequency are generally higher than the rest of the envelope spectrum peaks. As an IRF bearing is used, prominent sidebands can also be noticed on either side of the BPFI harmonics spaced at the magnitude of the shaft frequency harmonics. This lowers the fitness score of the solution, however, the BPFI amplitudes are still clearly distinguishable. Fig. 6(b) shows the demodulation band selected by the algorithm for the IRF bearing at 47 Hz. The frequency band can be seen focusing around a dominant resonance, location A, similar to the band selection method that was mentioned in [5].

Fig. 7(a) shows the fitness score of the best individual in each generation and it can be observed that it converges after about 27 iterations. The highest fitness score found through the algorithm was approximately 1.67. The variation in optimal passband selection for each iteration can be seen in Fig. 7(b). The optimal band selected was 17840.4 to 23297.1 Hz and was found within the first 27 iterations. The IRF bearing case also converges in a small number of iterations similar to the ORF bearing for the same reasons.

D. COMPARISON OF DEMODULATION BANDS AND PERFORMANCE

Fig. 8(a) and 8(b) graphically represent the variation in demodulation bands obtained for ORF and IRF bearings respectively for the five sets of different shaft frequencies. The highlighted sections show the maximum and minimum points the limits have been computed to for each bearing.

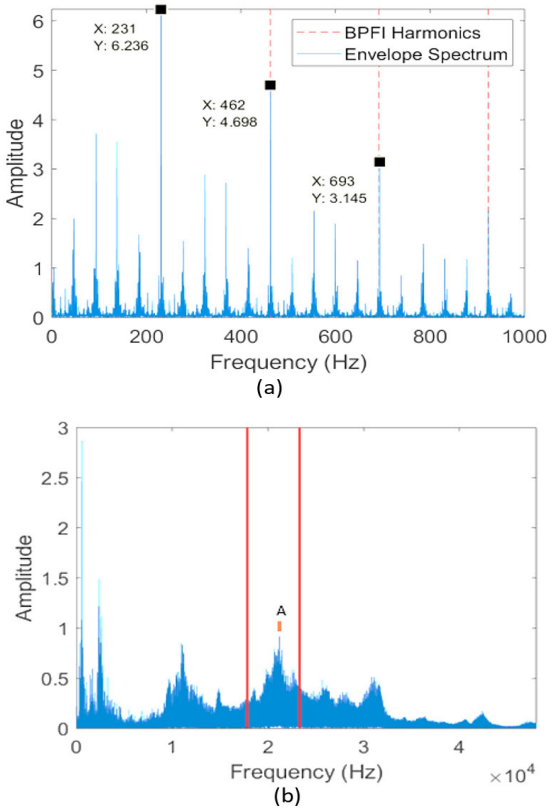


FIGURE 6. (a) Envelope spectrum and (b) frequency spectrum band selection for the IRF bearing at shaft speed of 47 Hz.

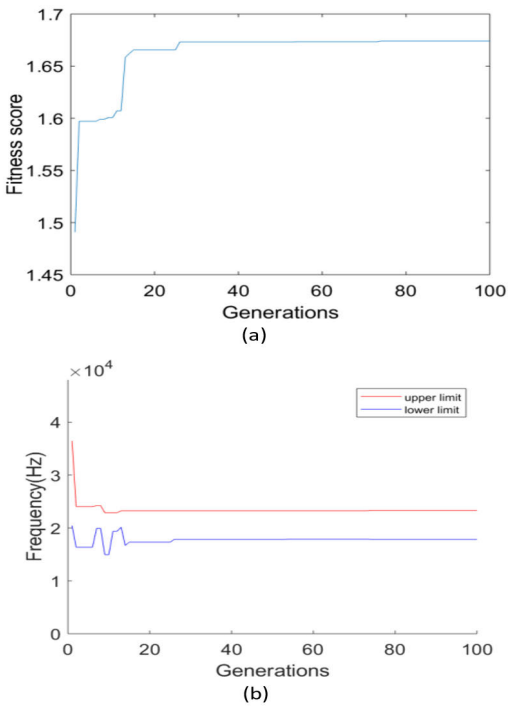


FIGURE 7. (a) Fitness score of the best individual, and (b) best passband found within 100 iterations for an IRF bearing at a shaft speed of 47 Hz.

The ORF bearing's narrowest bandwidth was observed for the case where the shaft frequency was 17 Hz. The widest bandwidth computed for the same bearing was in a shaft

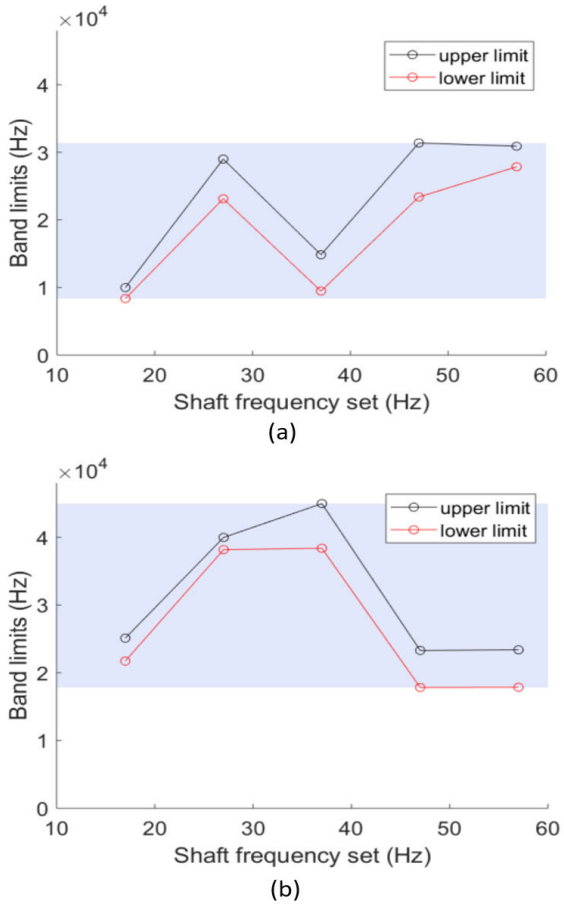


FIGURE 8. Comparison of optimal band limits computed for (a) an ORF bearing, and (b) IRF bearing at five different shaft frequencies.

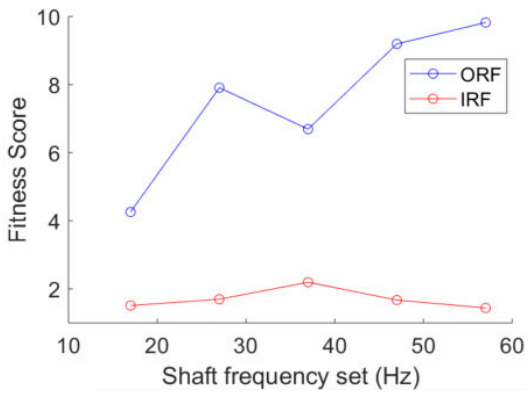


FIGURE 9. Fitness scores obtained for each optimal demodulation band selected.

TABLE 4. Fitness scores and time taken for 100 iterations using the proposed method for ORF and IRF bearings.

	Set Shaft Frequency (Hz)	17	27	37	47	57
ORF	Fitness score	4.26	7.91	6.69	9.20	9.82
	Time (s)	342.04	344.19	343.15	336.02	333.43
IRF	Fitness Score	1.51	1.70	2.20	1.67	1.44
	Time (s)	338.64	350.83	360.61	339.49	358.02

frequency of 47 Hz. Unlike the ORF bearing, the IRF bearing's narrowest and widest bandwidths computed were from the shaft frequencies 27 and 37 Hz respectively. No direct correlation could be made regarding selection of optimal bands with the operating frequency of the shaft. An important point worth noting however is that the highlighted sections in Fig. 8(a) and 8(b) are at a high frequency with the IRF bearing's optimal range reaching much higher than that of the ORF bearing.

The fitness scores obtained for the two bearings at the shaft frequencies tested are shown in Fig. 9. The fitness score of the IRF bearing is much lower than that of the ORF bearing for all shaft speeds due to the sidebands present on either side of the BPFI harmonics spaced by the magnitude of the shaft frequency harmonics. It was also noticed that the general shape of the fitness score variation relates to the band selection for both bearings in Fig. 8. The increase and decrease of the optimal band's position, disregarding magnitudes, closely resembled the variation observed in the respective fitness scores. The higher bandpass filters selected through the real-coded genetic algorithm gave a higher fitness score. However, this does not necessarily mean that the manual selection of a higher filter will always produce a better result as it is also greatly dependent on the bandwidth and the frequencies present within.

As shown in Table 4, all cases tested using the proposed method achieved reasonable fitness scores. This meant that the algorithm and fitness function were able to accurately select optimal passbands based on the evaluation made with the fitness function. ORF bearing tests obtained a very high fitness score but the IRF bearings cases were not quite as high due to the sidebands of the fault frequency present in the envelope spectrum. It is also important to note that in most cases tested, convergence is either achieved within 30 iterations or the fitness score increases minutely not making a significant difference. This means that for most cases, a reasonable fitness score can be achieved in 30 iterations. On average, the time taken to run a hundred iterations of the algorithm is 345 seconds on a DELL laptop with 8 GB RAM and an i5 8350U CPU. If convergence, or an acceptable solution, can be reached in about 30 iterations then the approximate time taken for this will be about 103 seconds.

It is demonstrated that the proposed approach can automatically find the optimized demodulation bands corresponding to the acquired vibration signals under different shaft frequencies. Although the algorithm achieved a fitness score large enough to distinguish the presence of faults, the performance could be further improved if the fitness score limitation was due to convergence at a local optimum rather than a global optimum. Methods such as vibrational genetic algorithm [27]

TABLE 5. Fitness scores for data analyzed from testrig using fast kurtogram method for ORF and IRF bearings.

	Set Shaft Frequency (Hz)	17	27	37	47	57
ORF	Fitness score	2.5	2.19	2.53	1.03	1.21
	Time (s)	2.67	2.10	2.42	2.06	2.08
IRF	Fitness Score	1.05	1.08	1.48	1.22	1.08
	Time (s)	2.26	4.37	2.50	2.77	0.49

TABLE 6. CWRU dataset envelope analysis comparison using proposed method and fast kurtogram.

	Defect size (mm)		0.1778	0.5334
ORF	Proposed algorithm	Fitness score	15.61	1.54
		Time (s)	35.38	117.52
	Fast kurtogram	Fitness score	3.54	0.98
		Time (s)	1.17	0.83
IRF	Proposed algorithm	Fitness score	4.23	2.34
		Time (s)	112.30	128.64
	Fast kurtogram	Fitness score	0.08	1.44
		Time (s)	0.83	1.06

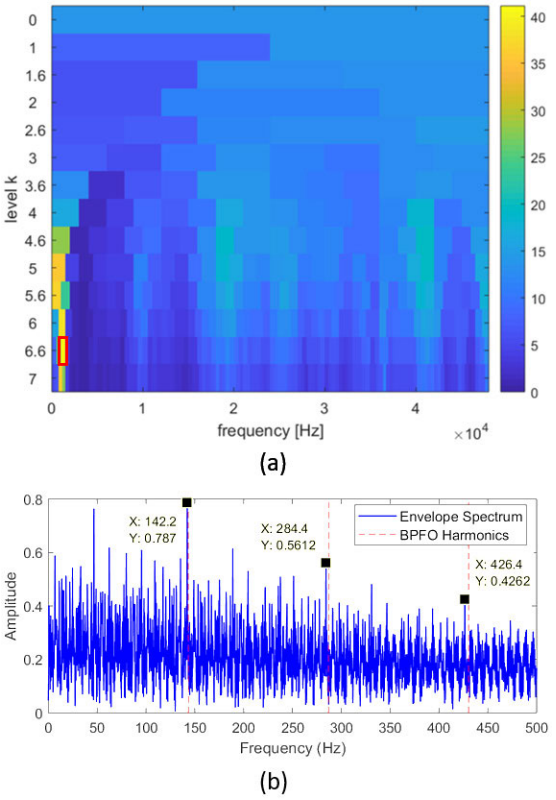


FIGURE 10. (a) Fast kurtogram where the red box indicates the selected combination and (b) envelope spectrum of ORF bearing at shaft frequency of 47 Hz.

or adaptive genetic algorithm [28] can ensure that the population diversity is maintained. The downside to using these methods however would be a significant increase in computation time due to the additional processes introduced to the real-coded genetic algorithm.

E. BENCHMARKING AND ADDITIONAL TESTING

The proposed algorithm was benchmarked with an established spectral analysis tool. The fast kurtogram as in [12] was used to analyse the same set of testing data as listed in Table 3. The resultant fitness function values as introduced in Eq. (8) are listed in Table 5. Comparing with the results obtained using the proposed method as listed in Table 4, it can be seen that the proposed method outperforms the fast kurtogram method for both ORF and IRF bearings over all shaft frequencies. Fig 10 shows the kurtogram and envelope spectrum from the fast kurtogram method for a case with an ORF bearing at a shaft frequency of 47 Hz. It can be seen that in the envelope spectrum, although the defect related frequencies can be observed, other frequencies are still prominent, which makes it difficult to identify the defect frequencies. As a comparison, the envelope spectrum by using the proposed approach, as shown in Fig 4, shows clearly the bearing defect frequencies associated with the ORF. It demonstrated the effectiveness of the new method.

The CWRU bearing dataset [29] has been used by many researchers in the field of bearing fault diagnosis and is therefore considered a standard set of reference data for the testing of new algorithms. The proposed method and the fast kurtogram [12] were once again used to obtain an optimal envelope spectrum from selected cases in CWRU's dataset as listed in Table 6. The dataset used was for drive end bearing faults at a sampling frequency of 48 kHz, and the cases selected for analysis were the largest and smallest defect size available for both bearings with an IRF and ORF. Note that the ORF data used had the defect in the 6 o'clock position. The largest and smallest defect sizes were used as these cases give a good indication of the performance of the algorithms.

The comparison of the fitness scores obtained showed that the performance of the proposed algorithm is better than the benchmark in all cases tested in terms of the detection accuracy. Thei algorithm is able to achieve a much higher fitness score specifically for cases with the smaller defect for both types of bearings. As for the cases with the larger bearing defects, the proposed algorithm also resulted in a greater fitness score, although the gap is not so significant. The superior performance in comparison to the benchmark for both the dataset collected and the one from CWRU's testrig indicated the adaptability of the algorithm. It was found that the proposed algorithm is capable of converging or completing its maximum iterations much sooner while using

CWRU's bearing dataset. This is due to the smaller sampling frequency in this dataset, which means that the algorithm will generally compute faster. A very high sampling rate is not required for the accurate analysis of accelerometer signals and therefore a good envelope spectrum result can still be achieved in a reasonable time. It should be noted that the benchmarking technique was able to generate results much faster than the proposed algorithm. Nevertheless, the prominence of defect frequencies in the proposed algorithm was significantly greater for all cases. It is expected the proposed method will be further improved in future study to reduce the computational cost.

IV. CONCLUSION

This paper proposed a real-coded genetic algorithm with a novel fitness function and crossover selection method to automate the optimal selection of bandpass filter parameters for envelope analysis to diagnose rolling element bearing defects. The ratio between fault frequency peaks and the maximum peak not corresponding to defects in the envelope spectrum is used as the fitness function to evaluate the solutions obtained by the algorithm. The triangle series method was used to divide the probability of each individual in the population being selected for crossover based on the fitness scores obtained. A generally fast convergence was achieved due to the design of the GA. The algorithm allowed for the distinction of defect related frequencies in a fully automated way. Comparison to a benchmarking test using fast kurtogram proved the effectiveness and superior capability of the developed algorithm to better accentuate defect related frequencies. Additional testing on another rotor-bearing system dataset was able to demonstrate the adaptability of the proposed algorithm to measurement data from different platforms.

REFERENCES

[1] F. Cong, J. Chen, G. Dong, and M. Pecht, "Vibration model of rolling element bearings in a rotor-bearing system for fault diagnosis," *J. Sound Vib.*, vol. 332, no. 8, pp. 2081–2097, Apr. 2013.

[2] J. B. Ali, N. Fnaiech, L. Saidi, B. Chebel-Morello, and F. Fnaiech, "Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals," *Appl. Acoust.*, vol. 89, no. 3, pp. 16–27 Mar. 2015.

[3] R. B. Randall, *Vibration-Based Condition Monitoring: Industrial, Automotive and Aerospace*. Hoboken, NJ, USA: Wiley, 2011.

[4] R. B. Randall and J. Antoni, "Rolling element bearing diagnostics—A tutorial," *Mech. Syst. Signal Process.*, vol. 25, no. 2, pp. 485–520, Feb. 2011.

[5] P. D. McFadden and J. D. Smith, "Vibration monitoring of rolling element bearings by the high-frequency resonance technique—A review," *Tribol. Int.*, vol. 17, no. 1, pp. 3–10, Feb. 1984.

[6] W. A. Smith, Z. Fan, Z. Peng, H. Li, and R. B. Randall, "Optimised spectral kurtosis for bearing diagnostics under electromagnetic interference," *Mech. Syst. Signal Process.*, vol. 75, pp. 371–394, Jun. 2016.

[7] S. Tyagi and S. K. Panigrahi, "An improved envelope detection method using particle swarm optimisation for rolling element bearing fault diagnosis," *J. Comput. Des. Eng.*, vol. 4, no. 4, pp. 305–317, Oct. 2017.

[8] E. Bechhoefer and P. Menon, "Bearing envelope analysis window selection," in *Proc. Annu. Conf. Prognostics Health Manage. Soc.*, 2009, pp. 1–7.

[9] P. Boškoski and A. Urevc, "Bearing fault detection with application to PHM Data Challenge," *Int. J. Progn. Heal. Manag.*, vol. 2, no. 1, p. 32, 2011.

[10] E. Bechhoefer, M. Kingsley, and P. Menon, "Bearing envelope analysis window selection using spectral kurtosis techniques," in *Proc. IEEE Conf. Prognostics Health Manage. (PHM)*, Jun. 2011, pp. 1–6.

[11] J. Antoni and R. Randall, "The spectral kurtosis: Application to the vibratory surveillance and diagnostics of rotating machines," *Mech. Syst. Signal Process.*, vol. 20, no. 2, pp. 308–331, 2006.

[12] J. Antoni, "Fast computation of the kurtogram for the detection of transient faults," *Mech. Syst. Signal Process.*, vol. 21, no. 1, pp. 108–124, Jan. 2007.

[13] A. Moshrefzadeh and A. Fasana, "The Autogram: An effective approach for selecting the optimal demodulation band in rolling element bearings diagnosis," *Mech. Syst. Signal Process.*, vol. 105, pp. 294–318, May 2018.

[14] Q. Ni, K. Wang, and J. Zheng, "Rolling element bearings fault diagnosis based on a novel optimal frequency band selection scheme," *IEEE Access*, vol. 7, pp. 80748–80766, 2019.

[15] Y. Zhang and R. B. Randall, "Rolling element bearing fault diagnosis based on the combination of genetic algorithms and fast Kurtogram," *Mech. Syst. Signal Process.*, vol. 23, no. 5, pp. 1509–1517, Jul. 2009.

[16] M. Kang, J. Kim, B.-K. Choi, and J.-M. Kim, "Envelope analysis with a genetic algorithm-based adaptive filter bank for bearing fault detection," *J. Acoust. Soc. Amer.*, vol. 138, no. 1, pp. EL65–EL70, Jul. 2015.

[17] J. Gaffney, D. A. Green, and C. E. M. Pearce, "Binary versus real coding for genetic algorithms: A false dichotomy?" *Anziam J.*, vol. 51, pp. 347–359, Jun. 2010.

[18] R. L. Haupt and S. E. Haupt, *Practical Genetic Algorithms*, 2nd ed. Hoboken, NJ, USA: Wiley, 2004.

[19] W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study," *Mech. Syst. Signal Process.*, vols. 64–65, pp. 100–131, Dec. 2015.

[20] F. R. Kschischang, *The Hilbert Transform*. Toronto, ON, Canada: Univ. of Toronto 2006.

[21] Z. Fan and H. Li, "A hybrid approach for fault diagnosis of planetary bearings using an internal vibration sensor," *Meas. J. Int. Meas. Confed.*, vol. 64, pp. 71–80, Mar. 2015.

[22] A. Fernandez, "Demodulation or envelope analysis," Power-MI, 2017.

[23] W. Liu, W.-C. Gao, Y. Sun, and M.-J. Xu, "Optimal sensor placement for spatial lattice structure based on genetic algorithms," *J. Sound Vib.*, vol. 317, nos. 1–2, pp. 175–189, Oct. 2008.

[24] J. Magalhães-Mendes, "A comparative study of crossover operators for genetic algorithms to solve the job shop scheduling problem," *WSEAS Trans. Comput.*, vol. 12, no. 4, pp. 164–173, 2013.

[25] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Reading, MA, USA: Addison-Wesley, 1989.

[26] J. Carr, *An Introduction to Genetic Algorithms*. Walla, WA, USA: Whitman College, 2014.

[27] A. Hacioğlu and I. Özkol, "Vibrational genetic algorithm as a new concept in airfoil design," *Aircr. Eng. Aerosp. Technol.*, vol. 74, no. 3, pp. 228–236, Jun. 2002.

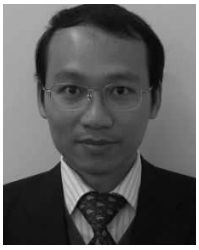
[28] M. J. Mahmoodabadi and A. R. Nemati, "A novel adaptive genetic algorithm for global optimization of mathematical test functions and real-world problems," *Eng. Sci. Technol. Int. J.*, vol. 19, no. 4, pp. 2002–2021, 2016.

[29] Case Western Reserve University. *Bearing Data Center*. Accessed: Aug. 2, 2019. [Online]. Available: https://csegroups.case.edu/bearingdatacenter/home

VIGNESHWAR KANNAN was born in Pondicherry, India, in 1997. He received the Dip.Eng. degree from the Queensland Institute of Business and Technology, Australia, in 2015, and the B.Eng. degree (Hons.) in mechanical engineering from Griffith University, Australia, in 2018, where he is currently pursuing the Ph.D. degree. In 2017, he worked as an Undergraduate Engineering Project Trainee with Ashok Leyland, India, where he participated in projects involving analysis and mitigation of cabin defects. His current research is on the intelligent condition monitoring and fault diagnosis of rolling element bearings. His research interests include condition monitoring and fault diagnosis, rotor-bearing systems, and autonomous control.

HUAIZHONG LI received the B.E. degree from Tsinghua University, Beijing, China, in 1988, the M.E. degree from Xi'an Jiaotong University, Xi'an, China, in 1991, and the Ph.D. degree from the National University of Singapore, Singapore, in 2002.

From 2001 to 2008, he was a Senior Research Engineer with the Singapore Institute of Manufacturing Technology, Singapore, and an Associate Principal Engineer with Vestas Technology Research and Development Singapore, Singapore, from 2008 to 2010. From 2011 to 2014, he was a Lecturer with the School of Mechanical and Manufacturing Engineering, UNSW, Australia. He is currently a Senior Lecturer with the School of Engineering and Built Environment, Griffith University, Australia. He has authored or coauthored more than 100 publications, including refereed technical articles in international journals, conference proceedings, and scholarly book chapters. His current research interests include advanced manufacturing technologies, machine tool dynamics, vibration monitoring, analysis and control, and mechatronics. He was a recipient of the 2015 Thatcher Bros Prize by the Institution of Mechanical Engineers (IMechE), U.K.



DZUNG VIET DAO received the Ph.D. degree in micro electro mechanical systems (MEMS) from Ritsumeikan University, Japan, in 2003.

He has served as a Postdoctoral Research Fellow (2003–2006), a Lecturer (2006–2007), and a Chair Professor (2007–2011) with Ritsumeikan University. In 2011, he joined Griffith University-School of Engineering and Built Environment, Australia, where he is currently an Associate Professor and also the Head of Mechanical and Mechatronic Engineering Discipline. His research interests include advanced manufacturing, MEMS sensors and actuators, as well as mechatronics. He has published 175 journal articles, over 175 conference papers, five book/book chapters, and 17 patents. He has secured over AU$5.5 M research funding from governments, industries, and universities.