Smart Structures and Systems, *Vol. 27, No. 4 (2021) 623-640*
DOI: https://doi.org/10.12989/sss.2021.27.4.623

Optimum feature selection for SHM of benchmark structures using efficient AI mechanism

Ramin Ghiasi [1a], Mohammad Reza Ghasemi*[1] and Tommy H.T. Chan [2b]

[1] *Department of Civil Engineering, Faculty of Engineering, University of Sistan and Baluchestan, Zahedan, Iran*
[2] *Civil Engineering and Built Environment School, Queensland University of Technology (QUT), Brisbane, QLD, Australia*

(*Received March 9, 2020, Revised November 8, 2020, Accepted December 8, 2020*)

Abstract. Structural Health Monitoring (SHM) is rapidly developing as a multi-disciplinary technology solution for condition assessment and performance evaluation of civil infrastructures. It consists of three parts: data collection, data processing (feature extraction/selection), and decision-making (feature classification). In this research, for effectively reducing a dimension of SHM data, various methods are proposed such as advanced feature extraction, feature subset selection using optimization algorithm, and effective surrogate model based on artificial intelligence methods. These frameworks enhance the capability of the SHM process to tackle with uncertainties and big data problem. To reach such goals, a framework based on three main blocks are proposed here: feature extraction block using wavelet pocket relative energy (WPRE), feature selection block using improved version of binary harmony search algorithm and finally feature classification block using wavelet weighted least square support vector machine (WWLS-SVM). The capability of the proposed framework is compared with various well known methods for each block. Results will be presented using metrics of precision, recall, accuracy and feature-reduction. Furthermore, to show the robustness of the proposed methods, six well-known benchmark datasets of SHM domain are studied. The results validate the suitability of the proposed methods in providing data reduction and accelerating damage detection process.

Keywords: structural health monitoring; feature extraction; feature selection; surrogate model; SHM benchmarks; data reduction

1. Introduction

Structural health monitoring (SHM) can be defined as the use of on-structure, non-destructive sensing systems to collect structural responses in order to monitor the performance and evaluate the safety of a structure for decision making and advancing the current practice in structural design, maintenance and rehabilitation (Chan *et al*. 2011). The goal of SHM is to improve the safety and reliability of aerospace, civil, and mechanical infrastructure by detecting damage before it reaches a critical state. To achieve this goal, technology is being developed to replace qualitative visual inspection and time-based maintenance procedures with more quantifiable and automated damage assessment processes. A more detailed general discussion of SHM can be found in Boller *et al*. (2009), Chan *et al*. (2011).

In vibration-based SHM, damage identification is performed from vibration signals measured simultaneously at different locations of the structure (Gomes *et al*. 2018). Damage detection can be performed in the time domain from the raw sensor data or in the feature domain, in which damage-sensitive features are first extracted from the time series, this process is referred to as feature extraction (Zhong *et al*. 2006).

Fundamentally, the feature-extraction process is based on fitting some model, either physics based or data-based, to the measured system response data. The parameters of these models or the predictive errors associated with such models then become the damage-sensitive features (An *et al*. 2015, Monavari *et al*. 2018, Wu and Jahanshahi 2018).

Another importing step in extracting the useful information and signal processing is Feature selection (FS) (Kashef *et al*. 2018, Kashef and Nezamabadi-pour 2015). FS is generally used in machine learning, especially when the learning task involves high-dimensional datasets. The primary purpose of feature selection is to choose a subset of available features, by eliminating features with little or no predictive information and also redundant features that are strongly correlated (Kashef *et al*. 2018, Kashef and Nezamabadi-pour 2015, Liu and Yu 2005). The availability of large amounts of data represents a challenge to classification analysis. For example, the use of many features may require the estimation of a considerable number of parameters during the classification process. Ideally, each feature used in the classification process should add an independent set of information. Often, however, features are highly correlated, and this can suggest a degree of redundancy in the available information which may have a negative impact on classification accuracy (Kashef and Nezamabadi-pour 2015). Thus, the FS approaches is needed to tackle these problems.

Methods for FS are generally divided into three categories: the filter-based, the wrapper-based and the

*Corresponding author, Professor,
E-mail: mrghasemi@eng.usb.ac.ir

ISSN: 1738-1584 (Print), 1738-1991 (Online)

hybrid methods (Kashef and Nezamabadi-pour 2015). The filter based methods use the statistical information of data to select features before the actual learning algorithm. A search strategy is applied to a given dataset to generate a subset of features. The generated subset is then evaluated by some measure, independent of the performance of the learning algorithm. The searching process is performed for finding a subset with relatively the best evaluation measure. The wrapper-based methods use the classification accuracy as a fitness function for subset evaluation. The hybrid methods use the independent measure to decide the best subsets for a given cardinality and use the mining algorithm to select the final best subset between the best subsets (Cai *et al*. 2018). FS algorithms have been reviewed in Guyon and Elisseeff (2003) and Liu and Yu (2005). For a large number of features, evaluating all states is computationally non-feasible and therefore metaheuristic search methods are required. Furthermore, from the optimization point of view, feature selection is a difficult combinatorial optimization problem (Gu *et al*. 2018). Firstly, since the size of the feature subset is not known a priori, the dimensionality of the decision space is non-reducible. Secondly, since the features may have complementary or contradictory interactions with each other, the decision space is non-separable. Due to the inefficiency of traditional search approaches in solving complex combinatorial optimization problems various metaheuristics have been proposed, such as Particle Swarm Optimization (PSO) (Chuang *et al*. 2011), Genetic Algorithm (GA)-based attribute reduction (Oh *et al*. 2004), Gravitational Search Algorithm (GSA) (Rashedi and Nezamabadi-pour 2014). These methods attempt to achieve better solutions by applying knowledge from previous iterations. In this paper we used binary version of Harmony Search (BHS) algorithm to find a subset of the most important features for accomplishing a particular machine learning task. However, the traditional BHS does not perform well for large-scale optimization problems, which degrades the effectiveness of BHS for feature selection when the number of features dramatically increases. Hence, Improved Binary Ant System Harmony Search (IBASHS) is proposed, for solving high-dimensional feature selection problems that can be considered as a combinatorial optimization problem.

In our previous work (Ghiasi *et al*. 2019, 2018, Ghiasi and Ghasemi 2018a, b), we have proposed a new framework for processing of measured structural health monitoring data and damage detection of large scale structures. In this paper, we intend to extend this framework by adding desirable features selection scheme which help to selecting appropriate subset of damage features. Furthermore, the efficacy of various feature extraction method will be examined. To reach to this goal, IBASHS is proposed as metaheuristic algorithms for feature subset selection.

The standard Harmony Search algorithm (HS) mimics music improvisation process to solve optimization problems (Geem *et al*. 2001). However, it is not suitable for binary representations. This is due to the pitch adjusting operator not being able to perform the local search in the binary space. To extend HS to solve the binary-coded problems more effectively and efficiently, the original BASHS algorithm which is proposed in Wang *et al*. (2011) will be improved in this paper.

The main focus of this research is facilitating the processing of big data in SHM (Cremona and Santos 2018). Accordingly, the integrated system consists of three block will be proposed in this paper. Firstly, wavelet packet decomposition (WPD) (Ghiasi *et al*. 2016, Han *et al*. 2005) is applied to the structural response signals under ambient vibration, and feature vectors are obtained via a feature extraction based on wavelet energy spectrum. Subsequently, the best feature subset is selected by the IBASHS algorithm based on a four desirability index: Pearson product-moment correlation coefficient (Onwuegbuzie *et al*. 2007), max-relevance and min-redundancy (Peng *et al*. 2005), distance evaluation technique (Nguyen *et al*. 2008) and F-score (Huang 2009). In the final step, selected feature is employed for training the Wavelet Weighted Least Squares Support Vector Machine (WWLS-SVM) algorithm.

To assess the efficiency of Wavelet Packet Transform (WPT) as a feature extraction method, it is compared with statistical characteristics and autoregressive (AR) parameters of vibration signals (Wang and Ong 2009). Furthermore, for measuring the accuracy of proposed surrogate model based on WWLS-SVM, similar model is constructed based on Extreme Learning Machine (ELM) (Huang *et al*. 2006) and Group Method of Data Handling (GMDH) neural network (Beheshti *et al*. 2017) in order to inspect pros and cons of each algorithm.

Various levels of structural damage detection including the occurrence, location and severity of the damages are studied through a numerical analysis based on the six benchmark dataset as: Four-story structure of IASC-ASCE SHM group (Das and Saha 2018, Johnson *et al*. 2003), three-story frame aluminum structure of the Los Alamos National Laboratory (Figueiredo *et al*. 2009), wooden bridge model (Kullaa 2011) and three different model of beam structure (Kullaa 2014, 2009). Finally, the efficacy of using Binary Genetic Algorithm (BGA), Binary Particle Swarm Optimization (BPSO), Advanced Binary ACO (ABACO) (Kashef and Nezamabadi-pour 2015), and Improved Binary Gravitational Search Algorithm (IBGSA) compared to IBASHS as main algorithm for feature selection, will also be investigated.

The paper is organized as follows: In Section 2, three main blocks of proposed method will be defined. Feature extraction method utilizing WPT, statistical characteristics and AR parameters of dynamic signals are described in section 3. Optimization algorithm including HS and the procedure for constructing the IBASHS algorithm are presented in section 4. In Section 5, a description of the intelligent classification algorithm, utilizing WWLS-SVM, ELM and GMDH will be given. SHM benchmark datasets are described in section 6. Evaluation of proposed method on benchmark data sets studied in Section 7 and Conclusions are presented in Section 8.

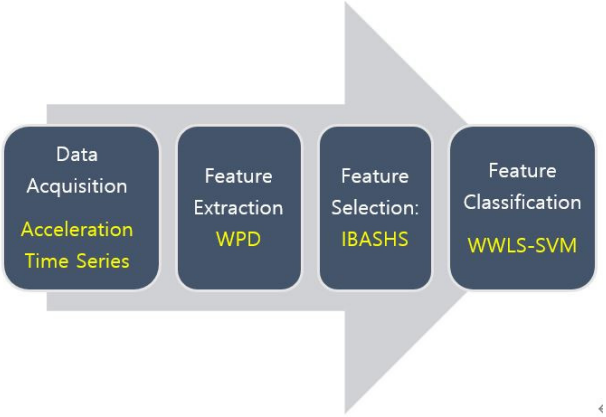


Fig. 1 Summary of damage detection approach

2. Damage detection procedure based on the proposed algorithm

Fig. 1 presents a summary of proposed novel method based on WPD, IBASHS algorithm and WWLS-SVM that is used to look for an optimal feature set. The proposed method consists of three main blocks:

(A) Feature Extraction Block
Feature extraction block is implements using methods that will describe in section 3
(B) Feature Selection Block
The feature selection procedure will be described in section 4.
(C) Feature Classification Block

In the final block, well trained surrogate model based on Artificial Intelligence (AI) methods of section 5 is applied to classify and identify the structural condition of samples based on the given principles

3. Feature extraction blok

This section gives a brief overview of three damage-sensitive features extraction techniques based on time-series analysis. The techniques will then be applied to measured acceleration data and comparison result will be present in section 7.1.

3.1 Wavelet Packet Transform (WPT)

Wavelet packet transform (WPT) is a technique to decompose a signal repeatedly into successive low and high frequency components (Mallat 1989). In WPT both approximations and details are decomposed further, which results flexible and wide base for the analysis of signals. The wavelet packets are alternative basis functions formed by linear combinations of the usual wavelet functions. As a result, each of mother wavelet functions can be used in WPT. In this study, one of these mother wavelets, Battle-Lemarie, is adopted (Daubechies 1992).

The WPT of a time domain signal $f(t)$ can be calculated using a recursive filter-decimation operation (Coifman and Wickerhauser 1992). After j-levels of decomposition, the original signal $f(t)$ can be expressed as

$$f(t) = \sum_{i=1}^{2^j} f_j^i(t) \tag{1}$$

$$f_j^i(t) = \sum_{i=1}^{2^j} C_{j,k}^i(t)\psi_{j,k}^i(t) \tag{2}$$

Herein, a linear combination of wavelet functions $\psi_{j,k}^i(t)$ can express the component signal $f_j^i(t)$. Integers $i.j$ and k are the parameters of the modulation, scale and translation, respectively; $C_{j,k}^i(t)$ and $\psi_{j,k}^i(t)$ are defined as the wavelet packet coefficient and the wavelet packet function. The wavelet packet coefficients can be obtained from

$$C_{j,k}^i = \int_{-\infty}^{\infty} f(t)\psi_{j,k}^i(t)dt \tag{3}$$

For structural damage detection purposes, frequency domain information tends to be more important, thus often requiring a high level of WPT to detect the sudden changes in the signals. Once the WPT is carried out, the energies of these decomposed component signals can be utilized for structural condition assessment. These component energies are defined as

$$E_j^i = \int_{-\infty}^{\infty} f_j^i(t)^2 dt \tag{4}$$

It can be shown that, when the mother wavelet is semi-orthogonal or orthogonal (Han *et al.* 2005), the signal energy E_f is the summation of the j-level component energies as follows

$$E_f = \int_{-\infty}^{\infty} f^2(t)dt = \sum_{i=1}^{2^j} E_j^i \tag{5}$$

Generally, we use relative energy to indicate the damage feature, thus, the relative energy E_i in i-frequency band can be expressed as

$$E_i = \frac{E_j^i}{E_f} \tag{6}$$

Battle-Lemarie is adopted as the basis wavelet package function in this paper in order to decompose the signals to be analyzed into different frequency bands and to make each frequency band energy independent and irredundant (Daubechies 1992). Several optional measuring nodes are selected, and vibration signals from these nodes are analyzed by using the WPT.

Apart from the reliance on the mother wavelet function, the wavelet-based methods link up with the decomposition level at which the wavelet analysis must be carried out. Specification of an appropriate level is not known in advance and depends on a wide range of parameters

including the characteristics of the structure, the nature of the signal and the type, location and severity of the damage, etc. Several researchers have suggested trying different decomposition levels (Ravanfar 2017).

In this paper based on our previous work (Ghiasi *et al*. 2016) The level of decomposition of the wavelet packet is determined using both healthy and damaged structural models through a trial and error sensitivity analysis and set on 7 levels. The frequency band energy is then calculated and normalized. The wavelet package relative energy of the signals from sensor s is

$$E_p^s = \{E_m, m = 1, \dots, M\} \tag{7}$$

Where, $s = 1,2, \dots . S$, p is the acquiring number, $p = 1,2,3, \dots, P$.

The wavelet package relative energy (WPRE) E_p^s of the signals from sensor s is combined to obtain the fused feature vector (Coifman and Wickerhauser 1992)

$$E_p = \{E_p^1, E_p^2, \dots, E_p^s\} \tag{8}$$

3.2 Statistical Features (SF)

Time-domain vibrational signals from sensors can be pre-processed to form features vector by the function shown in Table 1. The features of each sensor are: root mean square, variance, skewness, kurtosis, crest factor and the maximum frequency signal (Widodo *et al*. 2007). These features represent the energy, the vibration amplitude and the time series distribution of the signal in time-domain (Widodo *et al*. 2007).

3.3 Autoregressive model

In statistics and signal processing, an autoregressive (AR) model is a representation of a type of random process that is used to describe/model and predict certain time-varying processes in nature, economics, etc. (Figueiredo *et al*. 2009). The AR model with p autoregressive parameters, AR(p), can be written as (Wang and Ong 2009)

Table 1 Time-domain features

Feature	Function
Root mean square	$rms = \sqrt{\frac{\sum_{n=1}^{N}(x(n))^2}{N}}$
Variance	$var = \sigma^2 = \frac{\sum_{n=1}^{N}(x(n) - mean(x))^2}{(N-1)}$
Skewness	$skewness = \frac{\sum_{n=1}^{N}(x(n) - mean(x))^3}{(N-1)\sigma^3}$
Kurtosis	$kurtosis = \frac{\sum_{n=1}^{N}(x(n) - mean(x))^4}{(N-1)\sigma^4}$
Crest factor	$crest = \frac{\max \lvert x(n) \rvert}{rms}$
Maximum value	$max = \max \lvert x(n) \rvert$

$$x_i = \sum_{j=1}^{p} \varphi_j x(i-j) + e_i \tag{9}$$

where x_i is the measured signal at discrete time index i, and e_i is an unobservable random error (or residual error) at the i^{th} signal value. The unknown AR parameters, φ_j, can be estimated by using either least squares or the Yule-Walker equations (Moyo and Brownjohn 2002). Only the former is used in this paper.

The AR model can be used in SHM as a damage-sensitive extractor based on two approaches: (1) using the AR parameters φ_j; and (2) using the residual errors. The first approach consists of fitting an AR model to signals from undamaged and damaged structure. The AR parameters φ_j are then used as damage-sensitive features. The second approach consists of using the AR model, with parameters estimated from the baseline condition, to predict the response of data obtained from a potentially damaged structure. The residual error, which is the difference between the measured and predicted signal, is calculated at time i as follows

$$e_i = x_i - \hat{x}_i \tag{10}$$

where $\hat{x}_i$ is the predicted i^{th} signal value. This approach is based on the assumption that damage will introduce either linear deviation from the baseline condition or nonlinear effects in the signal, so the linear model developed with the baseline data will no longer predict the damaged system's response accurately. As a result, the residual errors associated with the damaged system will increase (Figueiredo *et al*. 2009). Note that for a fitted AR (p) model, the residual errors can only be computed for $i > p$ time points. The order of the AR model is an unknown value. A high-order model may perfectly match the data, but will not generalize to other data sets. On the other hand, the underlying physical system response is not necessarily captured by a low-order model. Several techniques are used in this study to determine the optimum model order, such as Akaike's information criterion (AIC) (Bishop 1995) and root mean squared error (RMSE).

4. Feature selection block

In second block, best subset of extracted features will be selected using IBASHS, based on different indices to measure desirability of features. This procedure will be described in following sub-sections.

4.1 Harmony Search (HS) Algorithm

Harmony Search (HS) algorithm is an meta-heuristic algorithm that Geem *et al*. first proposed in2001 (Geem *et al*. 2001), which is an inspired by the process of music players searching for a perfect state of harmony. The harmony in music is analogous to the solution vector of the optimization problem, and the musician improvisations depending on the aesthetic standard are analogous to the

local and global search mechanisms of HS. By imitating the improvisation process, HS searches for the global optimal solution based on harmony memory considering and the pitch adjusting rule. HS has attracted more and more attention due to its simplicity, flexibility and efficiency and has been successfully applied to various fields of science and engineering (Manjarres *et al*. 2013).

In order to meet the requirements of various problems, researchers improved HS and proposed the enhanced variants primarily from two aspects: (1) tuning the parameter setting (Mahdavi *et al*. 2007); (2) hybridizing HS with other optimization algorithms (Wang *et al*. 2011). Alia and Mandava surveyed the improved works on HS, details of which can be found in (Moh'd Alia and Mandaa, 2011).

The procedures of harmony search can be briefly described as follows (Geem *et al*. 2001). Firstly, the control parameters such as harmony memory size (HMS), harmony memory considering rate (HMCR) and the pitch adjusting rate (PAR), are set. The harmony memory (HM) which consists of HMS harmony vectors is then initialized. A new harmony vector $x' = [x'_1, x'_2, \dots, x'_N]$ is then improvised by memory consideration rule, pitch adjustment rule and a random selection as Fig. 2.

In Fig. 2, x'_i represent the i-th component of x', x_i^j is the i-th component of the j-th harmony vector in HM. UB_i and LB_i are the lower and upper bounds for the decision variable x'_i. The new candidate harmony is compared to the worst harmony vector of HM and stored in the HM if recorded better. This process is repeated until the termination criteria are satisfied.

For solving binary-coded problems, binary format of HS algorithm, firstly proposed by Geem (2005), is used to solve the water pump switching problem. In this algorithm, the candidate value for each decision variable is "0" or "1". Although researchers have attempted to solve the binary-coded problems with HS, the pitch adjustment operator's dysfunctions spoil HS performance in binary space and it is therefore essential to further study the binary-coded HS to improve its global search capability (Afkhami *et al*. 2013). In traditional HS, the pitch adjustment operator (PAO) chooses an adjacent value from the HM to perform a local search. However, for binary optimization problems, the solution only has two values, i.e., "0" and "1" which means that the PAR is degraded to a mutation operator. So, the performances of basic binary HS with or without PAO (Geem 2005) are spoiled due to the lack of a powerful local search operator.

```
for each i ∈ [1.N] do
  if rand_1() ≤ HMCR then /*memory consideration*/
     begin
   x_i = x_i^j,      where j ∈ (1, ..., HMS)
  if rand_2() ≤ PAR then /*pith adjustment*/
     begin
     x'_i = x'_i ± r × bw. where r ∈ (0.1) and bw is an
     arbitrary distance bandwidth
     endif
  else /*random selection*/
     x'_i = LB_i + r × (UB_i − LB_i)
  endif
done
```

Fig. 2 Improvising a new harmony

Inspired by the Ant system mechanism, Wang *et al*. (2011) proposed the BASHS with the new Harmony Memory Consideration Operator (HMCO) and PAO was developed. In BASHS, the global best harmony vector H^{gb} and the current iteration best harmony vector H^{ib} are employed to cooperate with HMCO and PAO to perform the global and the local search, respectively. For more details on process of BASHS algorithm, readers are addressed to original source (Wang *et al*. 2011). In this paper, the original version of BASHS will be improved.

The original BASHS algorithm uses fixed values for PAR and HMCR which are adjusted in initialization step and cannot be changed during new generations. Its main drawback lies on a high number of iterations needed to find the global optimum (Mahdavi *et al*. 2007, Wang *et al*. 2011).

This paper develops an Improved Binary Ant System Harmony Search (IBASHS) algorithm for optimization problems. The key difference between IBASHS and original BASHS methods is in the way of adjusting PAR and HMCR. The performance of BASHS algorithm was modified by eliminating its nature of embedded fixed value of PAR and HMCR, and allow for a dynamic variation of them in improvisation step, leading to an improved BASHS algorithm. Its dynamic alteration is in conjunction with generation number as shown in Fig. 3, expressed with the following relation (Mahdavi *et al*. 2007)

$$PAR(gn) = PAR_{min} + \frac{(PAR_{max} - PAR_{min})}{NI} \times gn \quad (11)$$

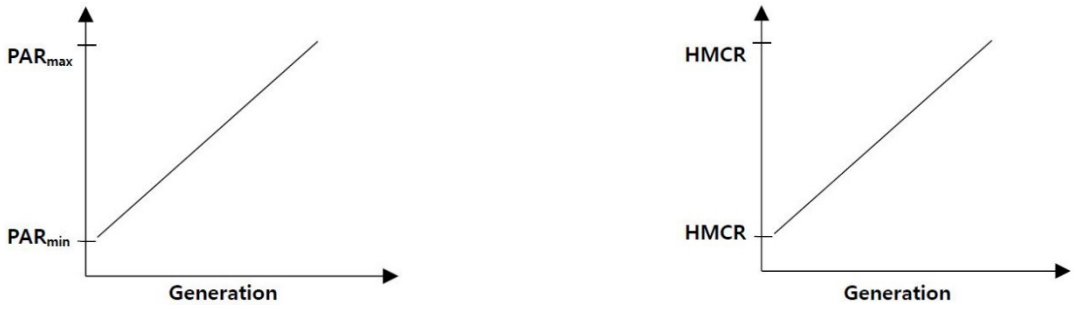
$$HMCR(gn) = HMCR_{min} + \frac{(HMCR_{max} - HMCR_{min})}{NI} \times gn \quad (12)$$

Where

$PAR(gn)$: pitch adjusting rate for each generation
PAR_{min}: minimum pitch adjusting rate
PAR_{max}: maximum pitch adjusting rate
NI: number of solution vector generations
gn: generation number
$HMCR(gn)$: harmony memory considering rate for each generation
$HMCR_{min}$: minimum harmony memory considering rate
$HMCR_{max}$: maximum harmony memory considering rate

4.2 Creating harmony vector

As mentioned earlier, given the original set of size n, feature selection problem is to find a minimal subset of salient features of size $p\,(p < n)$, such that the classification accuracy is maximized. In this study, IBASHS algorithm is designed for the feature selection. For this purpose, Harmony vector has been formatted in a binary form that each data in this vector represent features of vibrational signal of the sensors. By this procedure, the value of '1' means the feature is selected and the value of



(a) Variation of PAR (b) Variation of HMCR

Fig. 3 Variation of PAR and HMCR against generation number

'0' indicates that the feature is not selected.

At the end of each iteration, each harmony vector has a solution in the form of a binary vector with the same length as the number of the features, where 1 means selecting and 0 means deselecting the corresponding feature. This process continues for all iterations and at last, the best feature subset with the least classification error of the classifier is suggested as the best result.

4.3 Fitness function

Fitness function is an important factor for the speed and the efficiency of IBASHS algorithm. In this study, the fitness function of IBASHS is developed based on surrogate model accuracy and the efficiency of selected subset of features. The surrogate model (WWLS-SVM) accuracy is obtained by the evaluation of the test data classification using the trained model. By using this fitness function, efficiency of selected subset of features are evaluated using various indices to measure desirability of features. These indices will be explained in section 3.4. IBASHS selects the vector with the smallest fitness value after the completion conditions satisfied. The fitness function of IBASHS is formed as follows

$$F = W \times (Classification\ Accuracy)^{-1} + (1 - W) \times Features\ Desirability \quad (13)$$

where W is weighting factor between 0 to 1.

4.4 Measure the desirability of features

A desirability value, for each feature generally represents the attractiveness of the features, and can be any subset evaluation function like an entropy-based measure or rough set dependency measure (Jensen 2005). Four different methods were attempted in the proposed algorithm to determine the desirability value, including Pearson product-moment correlation coefficient, max-relevance and min-redundancy, Distance evaluation technique and F-score that will be described in following sub-section.

4.4.1 Method 1: Pearson product-moment correlation coefficient

The first method is based on correlation. Correlation is one of the most common and useful statistics that describes the degree of relationship between two variables. Statistics have proposed a number of criteria for estimating correlation. In this work, IBASHS uses the best known Pearson product-moment correlation coefficient (Onwuegbuzie *et al*. 2007) to measure correlation between different features of a given training set. The correlation coefficient r_{ij} between the two features i and j is

$$r_{ij} = \frac{\sum_h (x_i - \bar{x}_i)(x_j - \bar{x}_j)}{\sqrt{\sum_h (x_i - \bar{x}_i)^2} \sqrt{\sum_h (x_j - \bar{x}_j)^2}} \quad (14)$$

where x_i and x_j are the value of features i and j, respectively. The variables x_i and x_j represent the mean values of x_i and x_j, averaged over h samples. If the features i and j are completely correlated, i.e., exact linear dependency exist, then r_{ij} would be 1 or -1. If i and j are completely uncorrelated then r_{ij} would be 0 (Kabir *et al*. 2011).

Method 1 uses the idea of min redundancy, which attempts to select the most distinct features. For convenience, the correlation between features i and j is assumed to be high. In this case, the two features are highly similar. Therefore, one of these features is sufficient to describe the whole set. Hence, if one of them is selected, the probability to select/deselect the other feature can be described as $1 - |r_{ij}|$, that has a low/high value. Again, if the first feature is not selected, presence of the other feature is not necessary too, because they are similar. So, the probability to select/deselect the other feature can be described as $1 - |r_{ij}|$, that has a low/high value.

4.4.2 Method 2: Max-relevance and min-redundancy

Information theory-based feature selection methods select feature subsets that maximize information regarding the class label (Cai *et al*. 2018). The approximate expression and incremental heuristic search approaches are usually employed by such methods because directly calculating mutual information between the feature subset and the class label is difficult. Hence, the idea of max-relevance and min-redundancy as mutual information criteria is used in the second method (Peng *et al*. 2005). Max-relevance is one of the most popular approaches to realize max-dependency in feature selection, i.e., selecting

the features with the highest relevance to the target class $\boldsymbol{c}$. Relevance is usually characterized in terms of correlation or mutual information (Peng *et al*. 2005). As Eq. (15) shows, geometric mean of the two criteria (max-relevance and min-redundancy) is considered to calculate the desirability information of features. Here, $\boldsymbol{cls - cor_j}$, is the correlation between feature $\boldsymbol{j}$ and class labels over all samples. The more this value is close to 1 for a feature, the higher this feature is correlated to the class labels and thus the importance of the feature is more sound (Peng *et al*. 2005).

$$Desirability\ Index = \sqrt{(1 - |r_{ij}|)(|cls - cor_j|)} \quad (15)$$

4.4.3 Method 3: Distance evaluation technique

The within-class distance is given by Nguyen *et al*. (2008)

$$J_c = \sum_{i=1}^{c} p_i J_i \quad (16)$$

$$J_i = (\frac{1}{n_i}) \sum_{k=1}^{n} (x_k^i - m_i)^T (x_k^i - m_i) \quad (17)$$

where class i = 1, …, c; m_i is the mean vector of class i; n_i is the number of samples in class i and p_i is the ratio factor between the number of samples in class i and the total samples. The between-class distance is

$$J_b = \sum_{k=1}^{c} p_i (m_i - m)^T (m_i - m) \quad (18)$$

where m is the mean vector of all classes. In this study, class means the damage pattern**.** The J_c/J_b ratio is used to obtain the optimal features based on the criterion that chooses the smaller within-class distance J_c and the larger between-class distance J_b. The number of features is selected based on minimizing J_c/J_b ratio.

4.4.4 Method 4: Pearson product-moment correlation coefficient

This method is based on F-score which is a measurement to evaluate the discrimination ability of feature $\boldsymbol{i}$. Eq. (19) defines the F-score of the $\boldsymbol{i^{th}}$ feature. The numerator specifies the discrimination among the categories of the target variable, and the denominator indicates the discrimination within each category. A larger F-score implies to a greater likelihood that this feature is discriminative (Huang 2009).

$$F_{score_i} = \frac{\sum_{k=1}^{c} (\bar{x}_i^k - \bar{x}_i)^2}{\sum_{k=1}^{c} \left[\frac{1}{N_i^k - 1} \sum_{j=1}^{N_i^k} (x_{ij}^k - \bar{x}_i^k)^2 \right]} \quad (19)$$

where c is the number of classes and **n** is the number of features; N_i^k is the number of samples of the feature i in class k, ($\boldsymbol{k = 1, 2, \ldots, c}$; $\boldsymbol{i = 1, 2, \ldots, n}$), x_{ij}^k is the $\boldsymbol{j}$-th training sample for the feature $\boldsymbol{i}$ in class k, ($j = 1,2, \ldots, \boldsymbol{N_i^k}$), $\boldsymbol{x_i}$ is the mean value of feature $\boldsymbol{i}$ of all classes and x_{ik} is the mean value of feature $\boldsymbol{i}$ of the samples in class $\boldsymbol{k}$ (Huang, 2009).

5. Feature classification block

In final block of proposed scheme, well trained surrogate model is applied to classify various condition of structure. In these models the input matrix will be selected features and outputs are the corresponding damage condition. The damage assessment results are therefore, will be obtained.

This section gives a brief overview of three well-known AI classification methods that utilized in final block and comparison result of them will be present in section 7.1.

5.1 Wavelet Weighted Least Square Support Vector Machine (WWLS-SVM)

Least square support vector machines (LS-SVM) are a class of kernel-based learning methods and are the least square versions of Support Vector Machines (SVM) (Suykens and Vandewalle 1999). They are a set of related supervised learning methods that analyze data and recognize patterns, and which are used for classification and regression analysis. They were proposed by Suykens and Vandewalle (1999). In this paper, Wavelet Weighted version of LS-SVM (Khatibinia *et al*. 2013) will be used a main surrogate model. In fact, the kernel function of Weighted version of LS-SVM (WLS-SVM) is substituted with a specific kind of wavelet function that proposed by Ghiasi *et al*. (2016). The thin plate spline Littlewood–Paley wavelet is used as the kernel function of WLS-SVM and called WWLS-SVM. For more detailed information on mathematical basis of WWLS-SVM algorithm, readers are referred to original paper (Ghiasi *et al*. 2016, Khatibinia *et al*. 2013).

5.2 Extreme Learning Machine (ELM)

An ELM is a novel and fast learning method based on the structure of multi-layer perceptrons (Huang *et al*. 2006), recently proposed and applied to a large number of classification and regression problems. The ELM approach is a novel way of training feedforward neural networks, with a perceptron structure. The most significant characteristic of the ELM training is that it is carried out just by randomly setting the network weights, and then obtaining the inverse of the hidden-layer output matrix. The advantages of this technique are its simplicity, which makes the training algorithm fast, as well as its outstanding performance compared to other methods of learning, such as classical multilayer perceptrons or basic support vector machines. Moreover, the universal approximation capability of the ELM network, as well as its classification capability, have already been certified (Huang *et al*. 2006).

5.3 Group Method of Data Handling (GMDH) neural network

In recent years, many neural network models have been proposed or employed for various components of structural health monitoring in order to perform pattern classification, function approximation, and regression (Ghiasi *et al*. 2018, Salehi and Burgueno 2018). Among them, GMDH is a neural network that automatically selects its optimum architecture. This feature leads to finding the most accurate or unbiased model. In GMDH, all possible combinations of a pair (x_i, x_j) of input variables (all possible neurons) are considered. Then polynomial coefficients are determined using one of the available minimizing methods such as the principle of least-squares error and singular value decomposition (with training data). Then, neurons that have the least error (for testing data) are kept active, and others are removed. Subsequently, the next layer will be generated, and this process continues up to the last layer (Ivakhnenko and Ivakhnenko 1995).

5.4 Evaluation functions

The features selected by the proposed algorithms are evaluated with the well-known metrics precision, recall, accuracy and feature-reduction. Precision is defined as the ratio of correctly assigned category C samples to the total number of samples classified as category C as in Eq. (20). Recall is the ratio of correctly assigned category C samples to the total number of samples actually in category C as in Eq. (21) (Powers 2011). Let TP_i, FP_i, TN_i, and FN_i indicate a number of samples as follows:

TP_i = the number of test samples correctly classified under i^{th} category (C_i)
FP_i = the number of test samples incorrectly classified under C_i
TN_i = the number of test samples correctly classified under other categories
FN_i = the number of test samples incorrectly classified under other categories

$$precision_i = \frac{TP_i}{TP_i + FP_i} \tag{20}$$

$$recall_i = \frac{TP_i}{TP_i + FN_i} \tag{21}$$

In this paper, classification accuracy (CA) is used to define the quality function of a solution, which is the percentage of samples correctly classified and evaluated as Eq. (22)

$$Accuracy = \frac{Number\ of\ samples\ correctly\ classified}{Total\ number\ of\ samples\ taken\ for\ exprimenta} \tag{22}$$

Another parameter which is used for comparison is the average feature reduction F_r, to investigate the rate of feature reduction

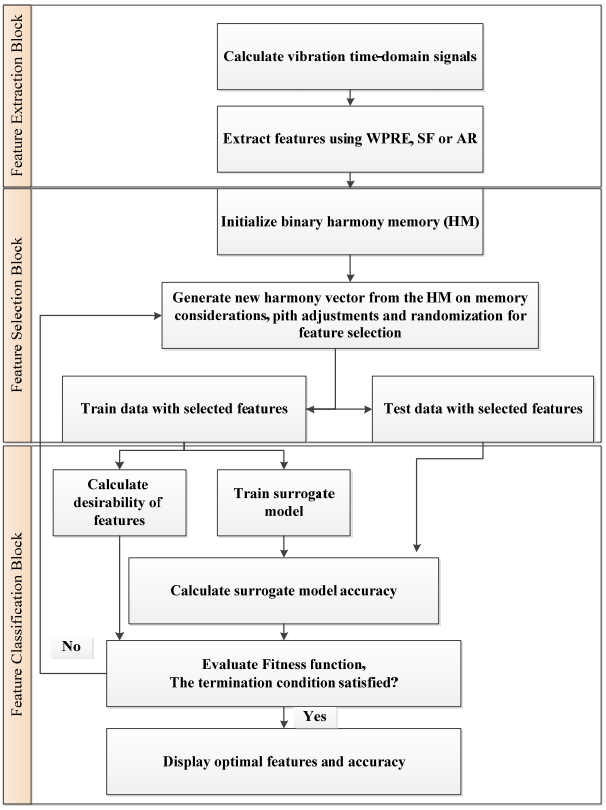


Fig. 4 The flow chart of the proposed system

$$F_r = \frac{n - p}{n} \tag{23}$$

where n is the total number of features and p is the number of selected features by the FS algorithm. F_r is the average feature reduction. The more it is close to 1, the more features are reduced, and the classifier complexity is less.

Fig. 4 presents the overall framework of proposed method based on three defined block.

6. Description of SHM benchmark datasets

A series of experiments are conducted to show the effectiveness of the proposed feature selection algorithm. All experiments were performed on a laptop with 2.40 GHz CPU and 4 Gb of RAM using MATLAB. For experimental studies, we have considered six datasets from benchmark studies in SHM domain were considered, including Four-story structure of ASCE health monitoring benchmark (Das and Saha 2018, Johnson *et al*. 2003), three-story frame aluminum structure of Los Alamos National Laboratory (Figueiredo *et al*. 2009), wooden bridge model (Kullaa 2011, 2009), a beam structure with environmental and operational influences, a beam with a nonlinear breathing crack, and beam with moving loads that presented in 7th European Workshop on Structural Health Monitoring (Kullaa 2014). These datasets have been the subject of many studies in SHM, covering examples of small, medium and high-dimensional datasets (Kullaa 2009, Liu *et al*.

Table 2 The characteristics of SHM benchmark datasets

Name of Data set	Number of classes	Number of experiments
Four-story structure	3	183
Three-story frame structure	6	1700
Wooden bridge	6	210
Beam with spring	6	100
Beam with a nonlinear breathing crack	7	16
Beam with moving load	4	26

2011, Santos *et al*. 2015). The characteristics of these datasets, summarized in Table 2, show a considerable diversity in the number of features, classes, and samples. Brief description of each will be presented in subsection 6.1 to 6.6. Furthermore, number of training and testing samples for each experimentation will be specified in each subsection.

6.1 ASCE health monitoring benchmark

The four-story steel structure shown in Fig. 5 has 12 degrees of freedom (DOF). This structure has a base plan of 2.5 m × 2.5 m and a height of 3.6 m. The quarter-scale symmetrical model of the structure was developed and studied in the Earthquake Engineering Research Laboratory at the University of British Columbia (UBC) (Johnson *et al*. 2003).

The members are hot rolled grade 300 W steel with a nominal yield stress of 300 MPa (42.6 ksi). The excitation is a low-level ambient wind loading at each floor in y-direction. To consider the environmental load uncertainties, the wind loading is modeled as a filtered Gaussian white noise process passed through a sixth order low-pass Butterworth filter with a 100 Hz cutoff. Sensors are installed in each floor on the side end middle column; there are in total 16 sensors. Signals to be analyzed are the acceleration response signals gathered from each floor sensors. The sampling frequency is 100 Hz; the length of the data is 40000.

To perform damage detection using the proposed algorithm, some damages were incorporated into the structure by removing the braces in y-direction. Relative calculation indicates that the damage of beams and the braces in x-direction has low effects on the vibrational responses (Liu *et al*. 2011). Therefore, only the damage of braces in y-direction is investigated in this study. Damage severity is described each time by removing 4 braces, 3 braces, 2 braces and 1 brace, respectively. Furthermore, it is assumed that damages occur in one, two, three and all four floors of the structure, each case of which are analyzed distinctly. For example, when damage is restricted to one floor, there are exerted 4×4 damage scenarios. Therefore all damage scenarios include (Ghiasi *et al*. 2016, Liu *et al*. 2011):

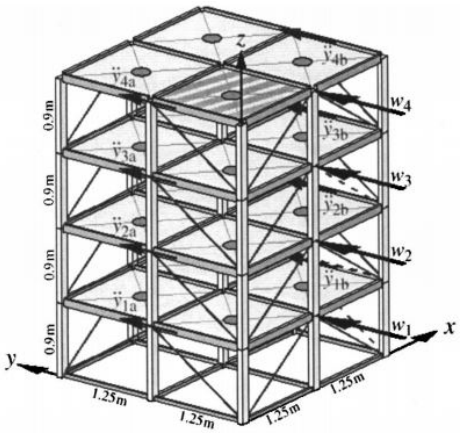
(1) Damages in One floor: 4 ×4 =16
(2) Damages in two floors: 4 ×6 =24
(3) Damages in three floors: 4 ×4 =16
(4) Damages in four floors: 4 ×1 =4

With the above damages, in addition to the case of no damage, there are in total 61 damaged cases. Considering the effect of the environmental noise, random Gaussian white noise with different severity is added to the acceleration responses of the above 61 damage cases. The ratios of the maximum root mean square (RMS) values between the noise and the signal for the 61 cases are 10%, 20% and 30%, respectively. These are named as the samples I to III. Samples I and II are used as training samples and sample III is employed for testing. Therefore, 183 samples are used in simulation of damage identification, including 61 testing samples and 122 training samples.

6.2 Three-story frame aluminum structure

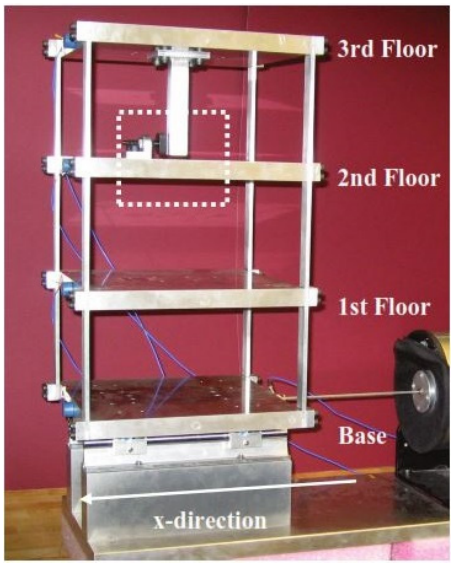
Another standard data sets that used in this study are from a three-story frame aluminum structure reported in Figueiredo *et al*. (2009) and has been intensively used for SHM validation in recent unsupervised damage detection approaches (Gui *et al*. 2017, Monavari *et al*. 2018, Santos *et al*. 2015). Test bed building model is four-degree-of-freedom system with varied practical conditions, including

(a) Real structure

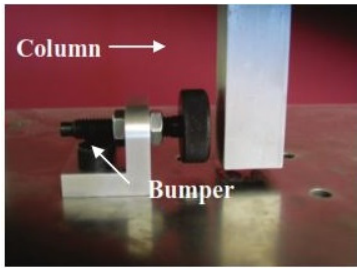


(b) Schematic drawing (Johnson *et al*. 2003)

Fig. 5 Four-story structure of ASCE health monitoring benchmark studies



(a) Three-story Building Structure and Shaker



(b) Adjustable Bumper and Column

Fig. 6 Three-story Frame Structure (Figueiredo *et al*. 2009)

variations in stiffness and mass loading. These variations simulate temperature changes and traffic, respectively. These changes were designed to introduce variability in the fundamental natural frequency up to about 7 percent from the baseline condition, which is within the range normally observed in real-world structures (Figueiredo *et al*. 2009). The experimental natural frequencies and damping ratios of all state conditions as well as the numerical natural frequencies can be found in Figueiredo *et al*. (2009). Raw data were collected by four accelerometers mounted on the structure, as shown in Fig. 6.

A nonlinear damage scenario was introduced by contacting a suspended column with a bumper mounted on the floor below to simulate fatigue crack that can open and close under loading conditions or loose structural connections. The smaller gap between the column and the bumper will result in the higher level of damage. Thus, by adjusting the gap, different levels of damage were created. More details about the test structure can be found in Figueiredo *et al*. (2009).

Acceleration time-series from 17 different structural state conditions were collected, as described in Table 3, where the first 9 state conditions introduce the undamaged and the rest are damaged states. Time-series discretized into 4096 data points sampled at 3.125 ms intervals corresponding to a sampling frequency of 320 Hz.

Data were acquired from 100 separate tests for each structural condition. Based on the test description (Figueiredo *et al*. 2009), state1 is the structure's basic condition (reference state) and states 2-9 include those states with simulated operational and environmental variability. State14 is considered as the most severely damaged one as it corresponds to the smallest gap case, which induces the highest number of impacts. State10 is the least severe damaged scenario and states11–13 represent mid-level damage scenarios. States15–17 are the variant states of either state10 or state13 with mass added effect in order to create more realistic conditions.

Table 3 Data labels of the structural state conditions

Label	Description
State 1	Baseline condition
State 2	Added mass (1.2 kg) at the base
State 3	Added mass (1.2 kg) on the 1st floor
State 4	States 4-9: 87.5% stiffness reduction at various positions to simulate temperature impact (more detail in Figueiredo *et al*. (2009))
State 5	
State 6	
State 7	
State 8	
State 9	
State 10	Gap (0.20 mm)
State 11	Gap (0.15 mm)
State 12	Gap (0.13 mm)
State 13	Gap (0.10 mm)
State 14	Gap (0.05 mm)
State 15	Gap (0.20 mm) and mass (1.2 kg) at the base
State 16	Gap (0.20 mm) and mass (1.2 kg) on the 1st floor
State 17	Gap (0.10 mm) and mass (1.2 kg) on the 1st floor

Training data set is consists of an extracted feature from 10 out of 100 tests of each undamaged state (states1–9) and testing data set is consists of feature from 10 out of 100 tests of each undamaged and damaged state (states1–17).

6.3 A beam structure with environmental and operational changes

The structure is a simple steel beam (Fig. 7) with a length of 1.4 m and a uniform cross-section of 50 mm × 5 mm (Kullaa 2011). The beam is also supported with a spring 612.5 mm far from the support, with the spring

constant k related non-linearly on temperature

$$k = k_0 + aT^3 \tag{24}$$

where k_0 = 100 kN/m, a = –0.8 (with consistent units), and T is temperature with a uniform random distribution between –20 to +40°C. Notice that seasonal variation of temperature would have been probably more realistic. The beam is divided into three sections of equal length. The Young's modulus E_i in the i^{th} section has linear relationship with a corresponding independent and dimensionless environmental variable z_i

$$E_i = E_0 + \sigma_i z_i, \quad i = 1,2,3 \tag{25}$$

where E_0 = 207 GPa, z_i are standardized Gaussian variables: $z_i \sim N(0,1)$, and the standard deviations σ_i of different sections are: $\sigma_1 = 5$ GPa, $\sigma_2 = 3$ GPa, and $\sigma_3 = 7$ GPa (Kullaa 2011). Based on Kullaa (2011) the structure is modelled using 144 simple beam elements and a single spring element.

The structure is excited at three points by independent random excitations with different amplitudes in each measurement. The response is measured at 47 DOF shown with numbered arrow (Fig. 7). In the response analysis, modal superposition with a static correction procedure (Clough and Penzien 1993) is used. At 47 equidistant points along the beam, transverse acceleration is measured. Gaussian noise with standard deviation σ = 0.01 m/s^2 is added to each sensor. In the average the noise level is approximately 1% of the signal. The sampling frequency is 571 Hz and in each measurement the number of samples is 2859. The first 50 measurements are from the undamaged structure. A slow environmental variability is assumed, justifying a constant environment during each measurement. As a conclusion, the environmental or operational variability thus introduced originate from (1) a variable spring with a non-linear relationship between temperature and the spring constant, (2) three regions with independently varying Young's moduli, and (3) random load distribution at three points.

Damage is a reduction in the beam depth at the spring support in two elements along a total length of 19.4 mm. This type of damage could represent local corrosion around the spring joint. Sensor 21 is located in the middle of the damaged region (Fig. 7). The height of the damaged beam varies in five different levels: 4.5, 4, 3.5, 3, and 2.5 mm. Each damage level is monitored with 10 measurements at variable unknown environmental conditions.

The first 25 measurements are the training data. The test data includes 75 measurements, 25 from the undamaged structure (26–50) and 50 from the damaged structure (51–100).

6.4 A beam with a non-linear breathing crack

The second beam structure is a simple beam with the following dimensions: length 5 m, height 0.5 m, and width 0.01 m (Kullaa 2014). It was modelled with 4-node linear 2D elements with reduced integration, and Abaqus Explicit finite element code was used for the simulations. The beam ends were supported on the neutral axis of the beam. In addition, the beam end nodes were forced to follow the theory of the Euler Bernoulli beam by assuming that the planes remained planes at the beam ends. Rayleigh damping (Clough and Penzien 1993) was used, because a full damping matrix was needed. The parameters were chosen to produce low damping in the frequency range of interest. In the tangential direction, the contact of the crack surfaces was assumed to be frictionless. The Abaqus surface - to - surface option has been used in the normal direction.

Fig. 8 shows a cracked beam model in a displaced configuration. A uniform transverse random load history, different in each case, was applied to the beam's top surface. The load histories were low-pass filtered below 1000 Hz, resulting in the structure's five active dynamic modes. The measurement period included 4001 samples for two seconds. There was no crack in the undamaged case. Damage was a single vertical crack with various crack lengths of 10, 20, 30, 50, 100, and 150 mm at the bottom of the midspan. The total number of sensors was 30. They measured accelerations at the beam's top or bottom edge in the transverse (vertical) direction. In Fig. 8, the top sensors are shown. The bottom sensors were on the opposite side of the beam at the same longitudinal positions. The midspan bottom sensor was located at the crack edge, which may result in too an idealistic case. Noise was added to the acceleration records obtained from the finite element analyses. The signal -to-noise ratio (SNR) for vibration measurement systems was 30 dB, which is a typical value. The SNR value had a major impact on the minimum size of the crack that can be detected (Kullaa *et al*. 2013).

Measures 1–6 were the training data for damage detection using proposed method. The test data included six measurements, four measurements (7–10) from the undamaged structure and two measurements (11–12) from the damaged structure.

6.5 Moving loads on a beam, modelling traffic on a bridge

The second beam structure is a modification of Tedesco

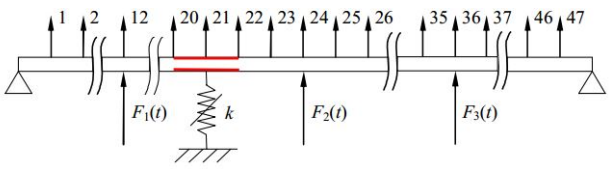


Fig. 7 Simply-supported beam with a variable spring (Kullaa 2011)

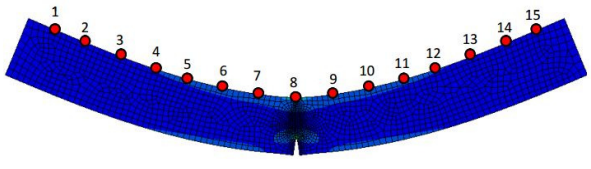


Fig. 8 Monitoring of a cracked beam with 15 accelerometers (Kullaa 2014)

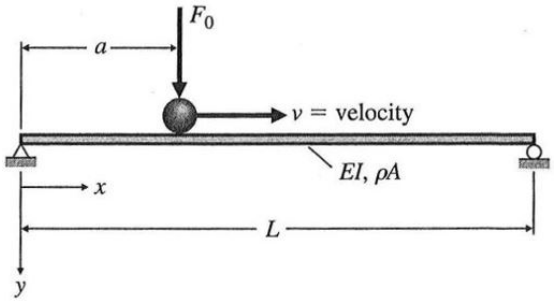


Fig. 9 A moving load on a beam modelling a vehicle on a bridge (Kullaa 2014)

bridge model (Tedesco *et al*. 1999). A simply supported uniform bridge girder is subjected to a moving load having constant velocity v (Fig. 9) (Kullaa 2014). This simulates a vehicle crossing the bridge. The bridge is analyzed as a continuous system. The transverse acceleration of the beam $\ddot{y}(x,t)$ is calculated with assuming zero initial conditions. The increased mass or the dynamics of the vehicle are not considered (Kullaa 2014).

During the measurement period, several vehicles are permitted to cross the one-way bridge. The mass of Vehicle, velocity, and the entering time are random variables. Linear superposition is used to compute the bridge response. The vehicle mass varies uniformly between 500 and 20 000 kg, the velocity varies uniformly between 50 and 120 km/h, and the entering time varies according to a oisson process with a mean of 5 s between events.

With a sampling frequency of 100 Hz, the measuring period is 30 s. Acceleration is measured along the bridge at 98 equidistant points. The number of modes in the analysis is 10. It is assumed that the measuring system includes a low-pass filter that removes the higher modes' contribution. The following material and structural parameters are used: span length L = 100 m, flexural rigidity EI = 754110 kNm^2, and mass/length ρA= 262 kg/m. Damping is zero. Damage is eterioration of the support, resulting in an increase of span length by 50, 100, and 200 mm. The first 20 measurements are from the undamaged structure and from each level of damage two measurements are acquired. Gaussian noise is added to each sensor with a standard deviation of $\sigma = 0.1$ m/s^2.

Measurements 1–10 are the training data for damage identification. The test data includes 16 measurements, 10 from the undamaged structure (11–20) and 6 from the damaged structure (21–26).

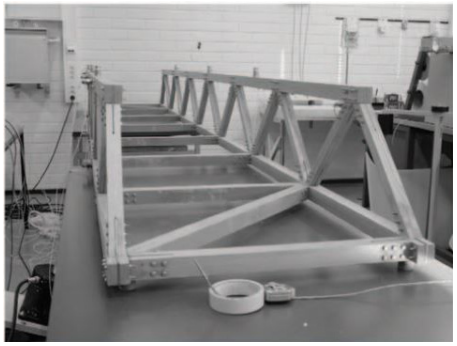
6.6 Wooden bridge model

An experimental research was carried out using a monitoring system built in the laboratory of Helsinki Polytechnic Stadia (Kullaa 2011, 2009). The structure was model of a wooden bridge shown in Fig. 10. In order to excite the lowest modes, random excitation was applied to the structure. The response was measured at three different longitudinal positions by 15 accelerometers. The frequency of sampling was 256 Hz and the measurement period was 32 s. The data were filtered below 64 Hz and re - sampled for sufficient redundancy. During several days, the measurements were made and it was noticed that the structure's dynamic properties varied due to changes in the environment. The main influences on the wooden structure were assumed to be variations in temperature and humidity. Damage was then introduced to the structure by adding point masses. The mass sizes were 23.5, 47.0, 70.5, 123.2 and 193.7 in terms of grams. The point masses were attached on the top flange, 600 mm left from the midspan (Fig. 10). The last measurements were again from a healthy structure. Compared to the total weight (36 kg), the added mass was very small, even the highest mass increase was only half a percent.

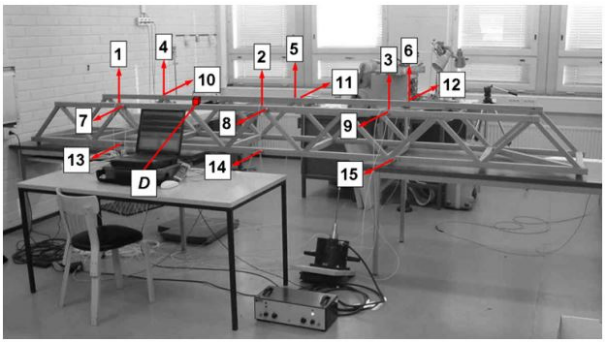
For a sensor fault, precision degradation fault was introduced by adding random noise to the sensor 3 during 10 measurements with a standard deviation of $\sigma = 0.01$. During the sensor fault, no additional masses were present. The first 175 measurements were the training data. The test data consisted of both healthy and abnormal systems measurements.

7. Experimental studies

In this section the accuracy and effectiveness of proposed scheme for feature extraction/selection in SHM domain will be evaluated based on data sets, described in section 6.



(a) Wooden bridge model



(b) Wooden bridge with the locations of sensors and damage (D) are indicated (Kullaa 2011)

Fig. 10 Wooden bridge

Furthermore, to validate the results obtained by the proposed IBASHS algorithm, it is compared with Binary Particle Swarm Optimization (BPSO) (Chuang *et al*. 2011), Binary Genetic Algorithm (BGA) (Oh *et al*. 2004), Improved Binary Gravitational Search (IBGSA) (Rashedi and Nezamabadi-pour 2014), Advanced Binary Ant Colony Optimization (ABACO) (Kashef and Nezamabadi-pour 2015), results of which are reported in Subsection 7.3. It is worth to note that, tuning parameters of these algorithms is set to best values that reported in paper accordingly.

The IBASHS parameters are: HMS = 50, HMCR = 0.9, PAR_{min} = 0.2, PAR_{max} = 0.5, $HMCR_{min}$ = 0.7 and $HMCR_{max}$ = 0.9. Population size for all algorithms is 50 and the maximum iterations is set to 500. The weighting factor W in fitness function is varied from 0.6 to 0.9 to get the different sets of features. Results are averaged over 20 independent runs in each data set and by every algorithm.

Table 4 Comparing the performance of data extraction methods

Feature extraction method	Evaluation index	GMDH	WWLS -SVM	ELM
Statistical feature	CA	83.12	81.1	74.31
	Precision	82.2	81.3	3.4
	Recall	81.3	80.3	71.4
	F-score	83.51	82.1	74.52
	Training time (s)	31	22	41
WPRE	CA	87	89.91	82.9
	Precision	86.1	88.2	81.5
	Recall	86.8	89	82.2
	F-score	88.2	90.02	82.5
	Training time (s)	35	27	49
AR model	CA	85	87.7	80.9
	Precision	84.8	86	80.7
	Recall	85.1	85.9	81
	F-score	86	87.8	81.1
	Training time (s)	37	33	48

7.1 Feature extraction and metamodeling step

For damage detection by the proposed algorithm IBASHS, firstly it is required to select best algorithm for Part A (Feature extraction) and part C (metamodeling) of framework. To reach to this goal, ASCE SHM benchmark dataset will be used as representative example. After choosing best algorithm for these two parts, effect of various metaheuristic algorithms on feature selection step of all data sets will be studied in Subsection 7.2.

Signals are preprocessed based on Section 4, Table 4 shows the effect of selecting WPRE, AR parameters or statistical features as data extraction methods on the damage detection accuracy of classification algorithms. Ratio of correctly detected damage cases to all test data (61 cases) is defined as classification accuracy (CA) same as Eq. (22).

The results show that the accuracy of damage detection based on WPRE is higher than that based on statistical features and AR model, under the same conditions. This could be due to the fact that WPT is a powerful mathematical tool for capturing changes of structural characteristics/properties induced by damage. As compared with statistical features, WPRE provides an effective feature extraction procedure for compressing the data measured and obtaining useful information for damage assessment. In other words, the features that are acquired by WPRE from vibrational signal have higher sensitivity to the damage of the structure in comparison with the features acquired by statistical features. Therefore, the use of WPRE as the data extraction method for other studies has been credit in this paper.

Furthermore, in order to confirm the efficiency of the proposed surrogate models, the performances of WWLS-SVM, ELM and GMDH neural networks are obtained based on evaluation function defined in Subsection 5.4, the results of which are shown in Table 4. Generally, Table 4 shows that WWLS-SVM has a high accuracy and performance for damage detection of the structure with any kind of feature extraction method. In addition, performance of GMDH neural network is increased with selecting WPRE as main feature extraction method meanwhile best classification accuracy of these surrogate model is 87%, therefore WWLS-SVM is chosen for metamodeling in future subsections.

Table 5 Classification accuracy of each desirability index on the tested data sets

Data set	Method 1: correlation coefficient	Method 2: Max-relevance min-redundancy	Method 3: Distance evaluation	Method 4: F-Score
	Mean of CA (±Std)	Mean of CA (±Std)	Mean of CA (±Std)	Mean of CA (±Std)
Four-story structure	0.955 (±0.0007)	0.961 (±0.0006)	0.986 (±0.0003)	0.997 (±0.0002)
Three-story frame structure	0.918 (±0.0032)	0.931 (±0.0007)	0.931 (±0.0007)	0.968 (±0.0143)
Wooden bridge	0.801 (±0.0093)	0.811 (±0.0096)	0.870 (±0.0113)	0.862 (±0.0202)
Beam with spring	0.922 (±0.0009)	0.951 (±0.0011)	0.951 (±0.0006)	0.991 (±0.0004)
Beam with a nonlinear breathing crack	0.742 (±0.0422)	0.790 (±0.0096)	0.828 (±0.0113)	0.832 (±0.0151)
Beam with moving load	0.701 (±0.0421)	0.749 (±0.0366)	0.736 (±0.0456)	0.798 (±0.0156)
Average	0.823	0.848	0.868	0.892

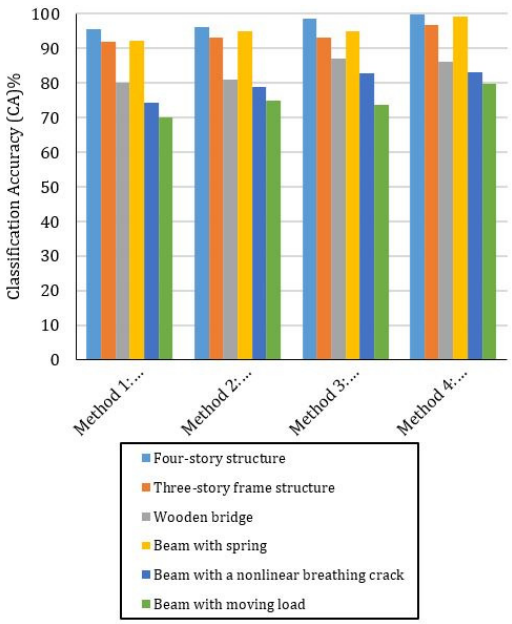


Fig. 11 Classification accuracy of each desirability index on the tested data sets

7.2 Classification accuracy of desirability indices

To specify the best method for measuring the desirability of features, the average classification accuracies of the proposed algorithm over 20 independent runs on the tested datasets are given in Table 5. The numbers in the last row of the table, show the average classification accuracy over the datasets. Furthermore, Figs. 11 and 12 illustrates the performance of the four approaches. According to these results, method 4 could achieve the best results among other proposed methods. Therefore, this method is chosen for measuring the desirability of features.

7.3 Classification accuracy of metaheuristic optimization algorithms

To show the utility of the proposed IBASHS algorithm, we compare the algorithm with IBGSA , BGA, BPSO, and ABACO, which are reported to be very strong algorithms in FS (Kashef and Nezamabadi-pour 2015). Table 6 and Fig. 13 shows the mean of classification accuracy (CA) results of every algorithm for each dataset. The average of results calculated over 20 independent runs. The number in

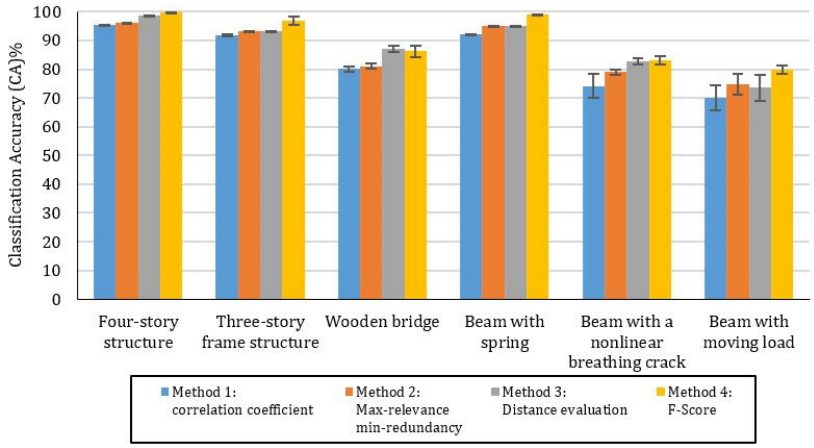


Fig. 12 Classification accuracy of the tested data sets for each desirability index. Error bars show the standard deviations

Table 6 Classification accuracy of each algorithm for the tested datasets

Data set	IBASHS	BGA	BPSO	IBGSA	ABACO
	Mean of CA (Rank)	Mean of CA (Rank)	Mean of CA (Rank)	Mean of CA (Rank)	Mean of CA (Rank)
Four-story structure	0.997 (2)	0.981 (4)	0.961 (5)	0.979 (3)	**0.998** (1)
Three-story frame structure	**0.968** (1)	0.930 (3)	0.922 (4)	0.941 (2)	**0.968** (1)
Wooden bridge	**0.862** (1)	0.821 (4)	0.801 (5)	0.856 (3)	0.859 (2)
Beam with spring	**0.991** (1)	0.987 (2)	0.981 (3)	**0.991** (1)	**0.991** (1)
Beam with a nonlinear breathing crack	**0.832** (1)	0.801 (4)	0.789 (5)	0.821 (3)	0.826 (2)
Beam with moving load	**0.798** (1)	0.728 (4)	0.713 (5)	0.740 (3)	0.743 (2)

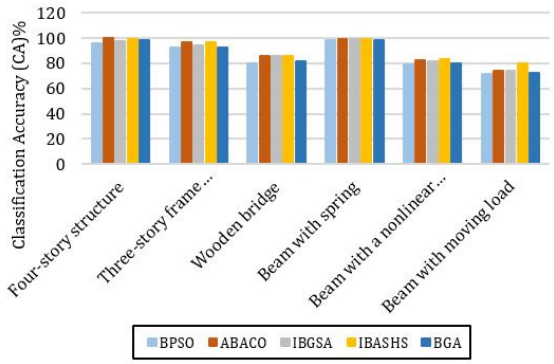


Fig. 13 Classification accuracy results of every algorithm for each dataset

brackets in each table slot shows the ranking of each. Comparison of the average precision, recall and the amount of F_r for the competing algorithms on the datasets are displayed in Table 7. It can conclude from these tables that the proposed IBASHS algorithm can obtain, in most of cases, better classification accuracy using a smaller feature set, compared to other algorithms.

In Table 8, another comparison has been made according to the sum of the ranks available in Table 6 for each algorithm. The lower sum of ranks for an algorithm shows the better average results in total cases against the others. Although this quantity is of lower accuracy degree for reporting results in some cases, it is common in Nonparametric Statistics. As can be seen in this table, IBASHS, gets the first rank. ABACO achieve the second rank and IBGSA ranked third. BGA and BPSO get the fourth to fifth rankings, respectively.

Furthermore, for better and more intuitively understanding the feature selection method and showing individual detection results of the proposed method on real

Table 8 The sum of the relative obtained ranks on the 6 number of data sets for each of the algorithms

IBASHS	BGA	BPSO	IBGSA	ABACO
7(1)	21(4)	27(5)	15(3)	9(2)

structure, extended result for ASCE benchmark data are shown in Figs. 14-17.

Further observations regarding the number of features that have been selected during the search procedure by each algorithm can be found in Figs. 14 and 15.

One may admit that IBASHS not only finds smaller feature subsets than the other algorithms on large-scale problems, but the number of selected features also decreases much faster.

From and Figs. 16 and 17, one can conclude that IBASHS performs a higher degree of exploration than the other algorithms, which enables it to explore the search

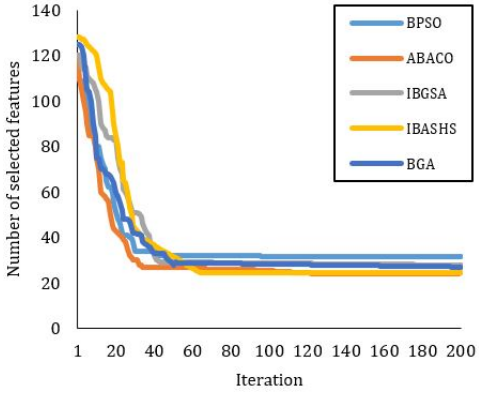


Fig. 14 Average number of selected features of each optimization algorithms over generations (ASCE Benchmark)

Table 7 Comparison of performance (precision, recall and Fr) of the algorithms on 6 data sets

	Metrics	Four-story structure	Three-story frame structure	Wooden bridge	Beam with spring	Beam with a nonlinear breathing crack	Beam with moving load	Sum
IBASHS	Precision	0.995	0.964	0.861	0.987	0.823	0.781	**5.411**
	Recall	0.988	0.962	0.857	0.971	0.819	0.779	**5.376**
	Fr	0.798	0.519	0.418	0.621	0.454	0.478	**3.288**
BGA	Precision	0.979	0.939	0.828	0.964	0.810	0.734	5.254
	Recall	0.976	0.940	0.829	0.962	0.804	0.714	5.225
	Fr	0.781	0.481	0.391	0.532	0.415	0.452	3.052
BPSO	Precision	0.963	0.918	0.808	0.942	0.781	0.701	5.113
	Recall	0.964	0.92	0.81	0.95	0.775	0.712	5.131
	Fr	0.750	0.484	0.383	0.531	0.402	0.441	2.991
IBGSA	Precision	0.977	0.942	0.849	0.982	0.813	0.725	5.288
	Recall	0.981	0.935	0.851	0.987	0.803	0.731	5.288
	Fr	0.787	0.497	0.421	0.595	0.451	0.413	3.164
ABACO	Precision	0.981	0.972	0.859	0.989	0.822	0.746	5.369
	Recall	0.975	0.975	0.86	0.981	0.813	0.743	5.347
	Fr	0.801	0.498	0.415	0.628	0.44	0.428	3.21

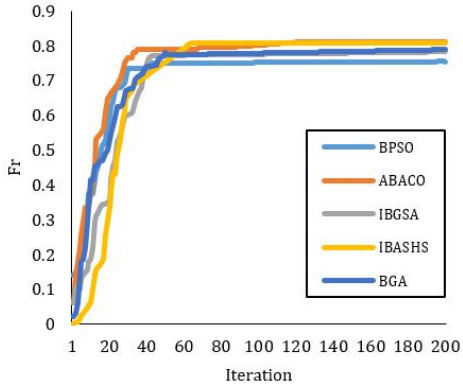


Fig. 15 Average feature reduction index of each optimization algorithms over generation (ASCE Benchmark)

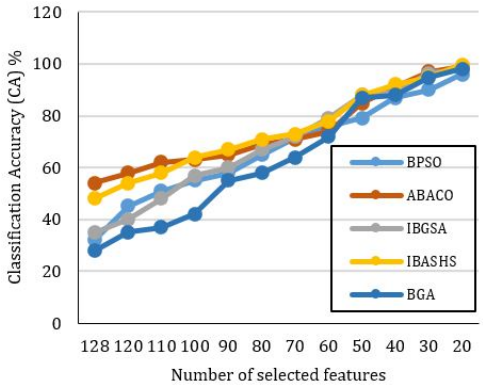


Fig. 16 Classification accuracy of each optimization algorithm (ASCE Benchmark)1

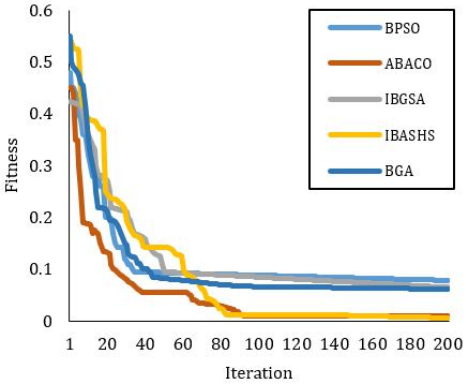


Fig. 17 Average fitness of each optimization algorithms with respect to each generation (ASCE Benchmark)

space to find a solution that selects a smaller number of features and better performance.

It is worth to note that, the FS method proposed in this study is a supervised wrapper-based feature selection method. Generally, in comparison with the filter model, the wrapper model could achieve a higher classification accuracy and tend to have a smaller subset size; however, it has high time complexity.

Finally, according to the results shown, adding desirability index and modifying PAR and HMCR parameters of BASHS, increases the exploration of search and guide algorithm to more salient features.

8. Conclusions

Feature selection is an important task which can significantly affect the performance of classification and recognition. In this paper, a new feature selection technique was presented based on IBASHS by modifying tuning parameter of BASHS and adding desirability index of feature to objective function. The proposed algorithm has a strong search capability in the problem space and can effectively find the minimal feature subset. This algorithm is compared with some powerful algorithms, including IBGSA, BGA, ABACO and BPSO. The roles of different feature extraction methods and different surrogate model were also investigated. In order to assess the performance of the proposed method on structural damage detection, six well known benchmark datasets in SHM domain were considered. Based on the numerical results, the following conclusions accomplished:

(1) The feature selection can remove the irrelevant and the redundant information by choosing useful features as input of surrogate model. Proposed FS approach based on IBASHS optimization algorithm reaches a better feature set in terms of classification accuracy and number of selected features. Generally, the IBASHS has capability of 41.8% to 79.8 % data reduction.

(2) There is no need to predefine the number of features to be selected. This task is thus assigned to the algorithm to select feature subsets with arbitrary numbers.

(3) Adding desirability index of feature to feature selection function has a great impact on number of selected features and guide algorithm to more salient features.

(4) Although performance of all surrogate models is improved by using WPRE as feature extraction method, for most cases considered, the classification accuracy by WWLS-SVM is better than that of GMDH and ELM.

(5) WPRE is capable of capturing changes of structural characteristics/properties induced by damage. It provides an effective feature extraction procedure for compressing the data measured and obtains useful information for damage assessment in comparison with AR model parameter and statistical feature.

References

Afkhami, S., Ma, O.R. and Soleimani, A. (2013), "A binary harmony search algorithm for solving the maximum clique problem", *Int. J. Comput. Appl.*, **69**.

An, D., Kim, N.H. and Choi, J. (2015), "Practical options for selecting data-driven or physics-based prognostics algorithms with reviews", *Reliab. Eng. Syst. Saf.*, **133**, 223-236. https://doi.org/10.1016/j.ress.2014.09.014

Beheshti, H., Mahmoud, N., Mohammad, M. and Ghasemi, R. (2017), "New neural network-based response surface method for reliability analysis of structures", *Neural Comput. Appl.* https://doi.org/10.1007/s00521-017-3109-2

Bishop, C.M. (1995), *Neural Networks for Pattern Recognition*, Oxford University Press.

Boller, C., Chang, F.-K. and Fujino, Y. (2009), *Encyclopedia of Structural Health Monitoring*, John Wiley & Sons.

Cai, J., Luo, J., Wang, S. and Yang, S. (2018), "Feature selection in machine learning: A new perspective", *Neurocomputing*, **300**, 70-79. https://doi.org/https://doi.org/10.1016/j.neucom.2017.11.077

Chan, T.H.T., Wong, K., Li, Z. and Ni, Y. (2011), "Structural health monitoring for long span bridges-Hong Kong experience and continuing into Australia", In: Chan, Tommy H T, Thambiratnam, D.P. (Eds.), *Structural Health Monitoring in Australia*, NOVA Science Publishers, Inc., New York.

Chuang, L.-Y., Yang, C.-H. and Li, J.-C. (2011), "Chaotic maps based on binary particle swarm optimization for feature selection", *Appl. Soft Comput.*, **11**, 239-248. https://doi.org/10.1016/j.asoc.2009.11.014

Clough, R.W. and Penzien, J. (1993), *Dynamics of Structures*, 2nd ed., New York McGraw-Hill.

Coifman, R.R. and Wickerhauser, M.V. (1992), "Entropy-based algorithms for best basis selection", *Inf. Theory, IEEE Trans.*, **38**, 713-718. https://doi.org/10.1109/18.119732

Cortes, C. and Vapnik, V. (1995), "Support-vector networks", *Mach. Learn.*, **20**, 273-297. https://doi.org/10.1007/BF00994018

Cremona, C. and Santos, J. (2018), "Structural health monitoring as a big-data problem", *Struct. Eng. Int.*, **28**, 243-254. https://doi.org/10.1080/10168664.2018.1461536

Das, S. and Saha, P. (2018), "Structural health monitoring techniques implemented on IASC–ASCE benchmark problem: a review", *J. Civ. Struct. Heal. Monit.*, **8**, 689-718. https://doi.org/10.1007/s13349-018-0292-5

Daubechies, I. (1992), Ten lectures on wavelets. Siam.

Figueiredo, E., Park, G., Figueiras, J., Farrar, C. and Worden, K. (2009), "Structural health monitoring algorithm comparisons using standard data sets", Los Alamos National Laboratory (LANL), Los Alamos, NM, USA.

Geem, Z.W. (2005), "Harmony search in water pump switching problem", *Proceedings of International Conference on Natural Computation*, Springer, pp. 751-760.

Geem, Z.W., Kim, J.H. and Loganathan, G.V. (2001), "A new heuristic optimization algorithm: harmony search", *Simulation*, **76**, 60-68. https://doi.org/10.1177/003754970107600201

Ghiasi, R. and Ghasemi, M.R. (2018a), "Optimization-based method for structural damage detection with consideration of uncertainties-a comparative study", Smart Struct. Syst. 22, 561–574. https://doi.org/10.1007/s00366-018-0636-0

Ghiasi, R. and Ghasemi, M.R. (2018b), "An intelligent health monitoring method for processing data collected from the sensor network of structure", *Steel Compos. Struct., Int. J.*, **29**(6), 703-716. https://doi.org/10.12989/scs.2018.29.6.703

Ghiasi, R., Torkzadeh, P. and Noori, M. (2016), "A machine-learning approach for structural damage detection using least square support vector machine based on a new combinational kernel function", *Struct. Heal. Monit.*, **15**, 302-316. https://doi.org/10.1177/1475921716639587

Ghiasi, R., Ghasemi, M.R. and Noori, M. (2018), "Comparative studies of metamodeling and AI-Based techniques in damage detection of structures", *Adv. Eng. Softw.*, **125**, 101-112. https://doi.org/10.1016/j.advengsoft.2018.02.006

Ghiasi, R., Fathnejat, H. and Torkzadeh, P. (2019), "A three-stage damage detection method for large-scale space structures using forward substructuring approach and enhanced bat optimization algorithm", *Eng. Comput.*, **35**, 857-874. https://doi.org/10.1007/s00366-018-0636-0

Gomes, G.F., Mendez, Y.A.D., Alexandrino, P. da S.L., da Cunha, S.S. and Ancelotti, A.C. (2018), "A review of vibration based inverse methods for damage detection and identification in mechanical structures using optimization algorithms and ANN", *Arch. Comput. Methods Eng.*, **26**(4), 883-897. https://doi.org/10.1007/s11831-018-9273-4

Gu, S., Cheng, R. and Jin, Y. (2018), "Feature selection for high-dimensional classification using a competitive swarm optimizer", *Soft Comput.*, **22**, 811-822. https://doi.org/10.1007/s00500-016-2385-6

Gui, G., Pan, H., Lin, Z., Li, Y. and Yuan, Z. (2017), "Data-driven support vector machine with optimization techniques for structural health monitoring and damage detection", *KSCE J. Civ. Eng.*, **21**, 523-534. https://doi.org/10.1007/s12205-017-1518-5

Guyon, I. and Elisseeff, A. (2003), "An introduction to variable and feature selection", *J. Mach. Learn. Res.*, **3**, 1157-1182.

Han, J.-G., Ren, W.-X. and Sun, Z.-S. (2005), "Wavelet packet based damage identification of beam structures", *Int. J. Solids Struct.*, **42**, 6610-6627. https://doi.org/10.1016/j.ijsolstr.2005.04.031

Huang, C.-L. (2009), "ACO-based hybrid classification system with feature subset selection and model parameters optimization", *Neurocomputing*, **73**, 438-448. https://doi.org/10.1016/j.neucom.2009.07.014

Huang, G.-B., Zhu, Q.-Y. and Siew, C.-K. (2006), "Extreme learning machine: theory and applications", *Neurocomputing*, **70**, 489-501. https://doi.org/10.1016/j.neucom.2005.12.126

Ivakhnenko, A.G. and Ivakhnenko, G.A. (1995), "The review of problems solvable by algorithms of the group method of data handling (GMDH)", *Pattern Recognit. Image Anal. C/C Raspoznavaniye Obraz. I Anal. Izobr.*, **5**, 527-535.

Jensen, R. (2005), "Combining rough and fuzzy sets for feature selection", Ph.D. Dissertation; University of Edinburgh, UK.

Johnson, E.A., Lam, H.F., Katafygiotis, L.S. and Beck, J.L. (2003), "Phase I IASC-ASCE structural health monitoring benchmark problem using simulated data", *J. Eng. Mech.*, **130**, 3-15. https://doi.org/10.1061/(ASCE)0733-9399(2004)130:1(3)

Kabir, M.M., Shahjahan, M. and Murase, K. (2011), "A new local search based hybrid genetic algorithm for feature selection", *Neurocomputing*, **74**, 2914-2928. https://doi.org/10.1016/j.neucom.2011.03.034

Kashef, S. and Nezamabadi-pour, H. (2015), "An advanced ACO algorithm for feature subset selection", *Neurocomputing*, **147**, 271-279. https://doi.org/10.1016/j.neucom.2014.06.067

Kashef, S., Nezamabadi-pour, H. and Nikpour, B. (2018), "Multilabel feature selection: A comprehensive review and guiding experiments", *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, **8**, e1240. https://doi.org/10.1002/widm.1240

Khatibinia, M., Javad Fadaee, M., Salajegheh, J. and Salajegheh, E. (2013), "Seismic reliability assessment of RC structures including soil–structure interaction using wavelet weighted least squares support vector machine", *Reliab. Eng. Syst. Saf.*, **110**, 22-33. https://doi.org/10.1016/j.ress.2012.09.006

Kullaa, J. (2009), "Eliminating environmental or operational influences in structural health monitoring using the missing data analysis", *J. Intell. Mater. Syst. Struct.*, **20**, 1381-1390. https://doi.org/10.1177/1045389X08096050

Kullaa, J. (2011), "Distinguishing between sensor fault, structural damage, and environmental or operational effects in structural health monitoring", *Mech. Syst. Signal Process.*, **25**, 2976-2989.

Kullaa, J. (2014), "Benchmark data for structural health monitoring", In: *EWSHM-7th European Workshop on Structural Health Monitoring*.

Kullaa, J., Santaoja, K. and Eymery, A. (2013), "Vibration-based structural health monitoring of a simulated beam with a breathing crack", In: *Key Engineering Materials*, Trans Tech Publ, pp. 1093-1100. https://doi.org/10.4028/www.scientific.net/KEM.569-570.1093

Liu, H. and Yu, L. (2005), "Toward integrating feature selection algorithms for classification and clustering", *IEEE Trans. Knowl. Data Eng.*, **17**(4), 491-502. https://doi.org/10.1109/TKDE.2005.66

Liu, Y., Ju, Y., Duan, C. and Zhao, X. (2011), "Structure damage diagnosis using neural network and feature fusion", *Eng. Appl. Artif. Intell.*, **24**, 87-92. https://doi.org/10.1016/j.engappai.2010.08.011

Mahdavi, M., Fesanghary, M. and Damangir, E. (2007), "An improved harmony search algorithm for solving optimization problems", *Appl. Math. Comput.*, **188**, 1567-1579. https://doi.org/10.1016/j.amc.2006.11.033

Mallat, S.G. (1989), "A theory for multiresolution signal decomposition: the wavelet representation", *Pattern Anal. Mach. Intell. IEEE Trans.*, **11**, 674-693. https://doi.org/10.1109/34.192463

Manjarres, D., Landa-torres, I., Gil-lopez, S., Ser, J. Del, Bilbao, M.N., Salcedo-sanz, S. and Geem, Z.W. (2013), "A survey on applications of the harmony search algorithm", *Eng. Appl. Artif. Intell.*, 1-14. https://doi.org/10.1016/j.engappai.2013.05.008

Moh'd Alia, O. and Mandava, R. (2011), "The variants of the harmony search algorithm: an overview", *Artif. Intell. Rev.*, **36**, 49-68. https://doi.org/10.1007/s10462-010-9201-y

Monavari, B., Chan, T.H.T., Nguyen, A. and Thambiratnam, D.P. (2018), "Structural deterioration detection using enhanced autoregressive residuals", *Int. J. Struct. Stab. Dyn.*, **18**, 1850160. https://doi.org/10.1142/S0219455418501602

Moyo, P. and Brownjohn, J.M.W. (2002), "Application of Box-Jenkins models for assessing the effect of unusual events recorded by structural health monitoring systems", *Struct. Heal. Monit.*, **1**, 149-160. https://doi.org/10.1177/1475921702001002003

Nguyen, N.-T., Lee, H.-H. and Kwon, J.-M. (2008), "Optimal feature selection using genetic algorithm for mechanical fault detection of induction motor", *J. Mech. Sci. Technol.*, **22**, 490-496. https://doi.org/10.1007/s12206-007-1036-3

Oh, I.-S., Lee, J.-S. and Moon, B.-R. (2004), "Hybrid genetic algorithms for feature selection", *IEEE Trans. Pattern Anal. Mach. Intell.*, **26**, 1424-1437. https://doi.org/10.1109/TPAMI.2004.105

Onwuegbuzie, A.J., Daniel, L. and Leech, N.L. (2007), "Pearson product-moment correlation coefficient", *Encycl. Meas. Stat.*, **2**, 751-756.

Peng, H., Long, F. and Ding, C. (2005), "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy", *IEEE Trans. Pattern Anal. Mach. Intell.*, 1226-1238. https://doi.org/10.1109/TPAMI.2005.159

Powers, D.M. (2011), "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation", J. *Mach. Learn. Technol.*, **2**(1), 37-63.

Rashedi, E. and Nezamabadi-pour, H. (2014), "Feature subset selection using improved binary gravitational search algorithm", *J. Intell. Fuzzy Syst.*, **26**, 1211-1221. https://doi.org/10.3233/IFS-130807

Ravanfar, S.A. (2017), "Vibration-based structural damage detection and system identification using wavelet multiresolution analysis", PhD. Dissertation, University of Malaya, Kuala Lumpur, Malaysia.

Salehi, H. and Burgueno, R. (2018), "Emerging artificial intelligence methods in structural engineering", *Eng. Struct.*, **171**, 170-189. https://doi.org/10.1016/j.engstruct.2018.05.084

Santos, A., Figueiredo, E., Silva, M.F.M., Sales, C.S. and Costa, J.C.W.A. (2015), "Machine learning algorithms for damage detection : Kernel-based approaches", *J. Sound Vib.*, **363**, 584-599. https://doi.org/10.1016/j.jsv.2015.11.008

Suykens, J.A.K. and Vandewalle, J. (1999), "Least squares support vector machine classifiers", *Neural Process. Lett.*, **9**, 293-300. https://doi.org/10.1023/A:1018628609742

Tedesco, J.W., McDougal, W.G. and Ross, C.A. (1999), *Structural Dynamics: Theory and Application*, Addison-Wesley, Menlo Park, CA, USA.

Wang, Z. and Ong, K.C.G. (2009), "Structural damage detection using autoregressive-model-incorporating multivariate exponentially weighted moving average control chart", *Eng. Struct.*, **31**, 1265-1275. https://doi.org/10.1016/j.engstruct.2009.01.023

Wang, L., Zhou, P., Fang, J. and Niu, Q. (2011), "A hybrid binary harmony search algorithm inspired by ant system", *Proceedings of the 2011 IEEE 5th International Conference on Cybernetics and Intelligent Systems* (*CIS*), IEEE, pp. 153-158.

Widodo, A., Yang, B.-S. and Han, T. (2007), "Combination of independent component analysis and support vector machines for intelligent faults diagnosis of induction motors", *Expert Syst. Appl.*, **32**, 299-312. https://doi.org/10.1016/j.eswa.2005.11.031

Wu, R.-T. and Jahanshahi, M.R. (2018), "Data fusion approaches for structural health monitoring and system identification: Past, present, and future", *Struct. Heal. Monit.*, **19**(2), 552-586. https://doi.org/10.1177/1475921718798769

Zhong, L., Song, H. and Han, B. (2006), "Extracting structural damage features: Comparison between PCA and ICA", In: *Intelligent Computing in Signal Processing and Pattern Recognition*, Springer, pp. 840-845.

FC