



Constrained K-means and Genetic Algorithm-based Approaches for Optimal Placement of Wireless Structural Health Monitoring Sensors

Shih-Lin Hung ^{1*}, Ching-Yun Kao ², Jyun-Wei Huang ¹

¹ Department of Civil Engineering, National Yang Ming Chiao Tung University, Hsinchu 300, Taiwan.

² Department of Public Safety and Fire Science, Chia Nan University of Pharmacy & Science, Tainan 71710, Taiwan.

Received 09 September 2022; Revised 11 November 2022; Accepted 21 November 2022; Published 01 December 2022

Abstract

Optimal placement of wireless structural health monitoring (SHM) sensors has to consider modal identification accuracy and power efficiency. In this study, two-tier wireless sensor network (WSN)-based SHM systems with clusters of sensors are investigated to overcome this difficulty. Each cluster contains a number of sensor nodes and a cluster head (CH). The lower tier is composed of sensors communicating with their associated CHs, and the upper tier is composed of the network of CHs. The first step is the optimal placement of sensors in the lower tier via the effective independence method by considering the modal identification accuracy. The second step is the optimal placement of CHs in the upper tier by considering power efficiency. The sensors in the lower tier are partitioned into clusters before determining the optimal locations of CHs in the upper tier. Two approaches, a constrained K-means clustering approach and a genetic algorithm (GA)-based clustering approach, are proposed in this study to cluster sensors in the lower tier by considering two constraints: (1) the maximum data transmission distance of each sensor; (2) the maximum number of sensors in each cluster. Given that each CH can only manage a limited number of sensors, these constraints should be considered in practice to avoid overload of CHs. The CHs in the upper tier are located at the centers of the clusters determined after clustering sensors in the lower tier. The two proposed approaches aim to construct a balanced size of clusters by minimizing the number of clusters (or CHs) and the total sum of the squared distance between each sensor and its associated CH under the two constraints. Accordingly, the energy consumption in each cluster is decreased and balanced, and the network lifetime is extended. A numerical example is studied to demonstrate the feasibility of using the two proposed clustering approaches for sensor clustering in WSN-based SHM systems. In this example, the performances of the two proposed clustering approaches and the K-means clustering method are also compared. The two proposed clustering approaches outperform the K-means clustering method in terms of constructing balanced size of clusters for a small number of clusters.

Keywords: Structural Health Monitoring; Wireless Sensing Network; K-Means Clustering Method; Effective Independence Method.

1. Introduction

Considering that measurements of large structures are usually incomplete, detecting structural damage in large complex structures is a challenging task. Therefore, structural health monitoring (SHM) using a limited number of sensors is a critical problem. Optimal sensor placement (OSP) of wired SHM systems detecting changes in modal parameters uses information measured from sensors with satisfactory sensitivity [1]. One of the widely used OSP methods is the effective independence (EI) method [2]. The EI method uses the Fisher information associated with candidate sensor locations to solve the OSP problem.

* Corresponding author: slhung0910@nycu.edu.tw

 <http://dx.doi.org/10.28991/CEJ-2022-08-12-01>



© 2022 by the authors. Licensee C.E.J., Tehran, Iran. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Wireless SHM systems have the advantages of low manufacturing costs, low-power requirements, small size, and simplicity of deployment (lack of cables) compared with traditional wired SHM systems [3-5]. Hence, increasing interest has been paid to applying wireless sensor networks (WSNs) to SHM in the last decade. However, power is at a premium in a WSN because the remote sensors are powered by batteries [6]. Thus, power efficiency is another important issue for the OSP of wireless SHM systems, in addition to information effectiveness [7]. There are several techniques to solve this problem, such as data reduction, sleep/wakeup approaches, and power-efficient routing [8, 9]. This study focuses on the investigation of cluster-based, power-efficient routing.

Hussain et al. [10] indicated that cluster-based approaches are suitable for monitoring applications. Several clustering techniques for WSNs are available in the literature, including the hierarchical approach, the K-means algorithm, and the genetic algorithm (GA). Heinzelman et al. [11] proposed the low-energy adaptive clustering hierarchy (LEACH) protocol. To uniformly scatter the energy load on each sensor, the cluster heads (CHs) in LEACH are rotated randomly. Each sensor transmits data to its CH, and then the CH routes the data to the base station (BS). Inspired by the LEACH and Bluetooth protocols, Kottapalli et al. [6] proposed a two-tier, lower and upper tiers, wireless network architecture for SHM. The lower tier is composed of clusters of sensors. The upper tier is composed of CHs. The lower and upper tiers operate on battery power and regular wall power supply with a battery backup, respectively. Similar to LEACH, each sensor transmits its data to the associated CH, and then the result is transmitted to the BS.

K-means clustering is one of the popular unsupervised machine learning algorithms. It separates a dataset into K clusters and searches for the best representative point in each cluster in a certain mathematical sense. In the study of Sasikumar & Khara [12], centralized and distributed K-means clustering algorithms were employed as network simulators. Results indicate that distributed clustering is more efficient than centralized clustering. To mend the choice procedure for the initial centroid in the K-means method, Ray & De [13] proposed the energy efficient clustering protocol based on K-means (EECPK-means) method using the midpoint algorithm for WSNs. Periyasamy et al. [14] developed a modified K-means clustering algorithm. Each cluster includes three CHs (simultaneously chosen). These CHs are rotated as the active CH by using a load-sharing mechanism to conserve the residual energy of sensors and extend the network lifetime.

A GA is an optimization method guided by the principles of evolution [15, 16]. Unlike traditional optimization methods, a GA searches for a global optimal solution without calculating the gradient of the objective function [17]. Some research has applied GAs to sensor clustering of WSNs. For instance, Jin et al. [18] decided the number and location of CHs by using a GA that minimizes the communication distance in a WSN. Each sensor connects to its nearest CH after CHs are selected. Ferentinos & Tsiligiridis [19] presented a multiobjective GA-based optimization methodology for WSN design and energy management, in which each sensor node also connects to its nearest CH. Peiravi et al. [20] used a two-nested GA for sensor clustering of WSNs to optimize the network lifetime. Nayak & Vathasavai [21] developed a clustering algorithm based on GA to optimize the WSN lifetime by considering distance and energy as parameters of the fitness function. Pal et al. [22] proposed a GA-based clustering method, namely energy efficient weighted clustering (EEWC) method. The fitness function is based on cluster separation, cluster compactness, and number of CHs. Simulation result indicates that EEWC is more effective in improving the network performance than other methods. Bhola et al. [23] proposed an optimized LEACH (O-LEACH) protocol based on the GA. The O-LEACH protocol balances the total number of CHs via fitness function based on residual and threshold energy to select CHs. The O-LEACH outperforms the LEACH in terms of improving energy consumption. Khoshraftar & Heidari [24] used the GA to improve the clustering process of wireless sensor nodes and find an optimum route. The fitness function is a function of energy, the total number of nodes and CHs, and sum of distance of all sensor nodes to the associated CH and BS. The network lifetime and reliability are improved by reducing energy consumption, using fewer number of CHs, and decreasing transmission distance.

The above studies developed different clustering approaches to reduce energy consumption and prolong the network lifetime. However, none of these studies analyzed the performance of clustering approaches in terms of constructing a balanced size of clusters which is a benefit of extending the network lifetime. To solve this problem, Hassan et al. [25] proposed three indices, standard deviation of mean square error, variation for clusters size, and clusters size range, to evaluate performance of clustering approaches for constructing a balanced size of clusters. In their study, the performances of K-means and fuzzy C-means algorithms were investigated. It is noteworthy that K-means and fuzzy C-means algorithms without using any constraint, the maximum data transmission distance of each sensor for example. Hence, practical applications of the two algorithms are not easy.

This study investigates the OSP of wireless SHM systems by using the two-tier network architecture proposed by Kottapalli et al. [6]. The first step is the optimal placement of sensors in the lower tier via the effective independence (EI) method by considering modal identification accuracy. The second step is the optimal placement of CHs in the upper tier by considering power efficiency. The sensors in the lower tier are partitioned into clusters before determining the optimal locations of CHs in the upper tier. Two approaches, a constrained K-means clustering approach and a genetic algorithm (GA)-based clustering approach, are proposed in this study to cluster sensors in the lower tier by considering

two constraints: (1) the maximum data transmission distance of each sensor; (2) the maximum number of sensors in each cluster. Given that each CH can only manage a limited number of sensors, these constraints should be considered in practice to avoid overload of CHs. The CHs in the upper tier are located at the centers of the clusters determined after clustering the sensors in the lower tier. The two proposed approaches aim to construct a balanced size of clusters by minimize the number of clusters (or CHs) and the total sum of the squared distance between each sensor and its associated CH under the two constraints. Accordingly, the energy consumption in each cluster is decreased and balanced, and the network lifetime is extended. To demonstrate the feasibility of using the two proposed clustering approaches for sensor clustering of WSN-based SHM systems, a numerical example is studied. The performances of the two proposed clustering approaches and K-means clustering method for constructing a balanced size of clusters are compared in this example by using the three indices proposed by Hassan et al. [25].

2. WSN Architecture

In this study, a two-tier WSN architecture proposed by Kottapalli et al. [6] is adopted. The sensors of a WSN are divided into several clusters. There is a CH in each cluster. The lower tier is composed of sensors communicating with their associated CHs, and the upper tier is composed of the network of CHs. Sensors transmit their data to their individual CHs, and then the CHs route the data to the BS. This study assumes that the CHs themselves do not measure data. The optimal placement of sensors (including the sensors in the lower tier and the CHs in the upper tier) considers modal identification accuracy and power efficiency. The steps are shown in Figure 1. The first step is the optimal placement of sensors in the lower tier by considering modal identification accuracy. In this study, the optimal locations of sensors in the lower tier are determined through EI method to estimate the structural modal parameters accurately by using the minimum number of sensors. The second step is the optimal placement of CHs in the upper tier by considering power efficiency. The sensors in the lower tier are clustered in advance to determine the optimal locations of CHs in the upper tier. Two approaches, a constrained K-means clustering approach and a GA-based clustering approach, are proposed in this study to cluster sensors in the lower tier. The CHs in the upper tier are located at the centers of the clusters determined after clustering the sensors in the lower tier. These steps are detailed in later sections.

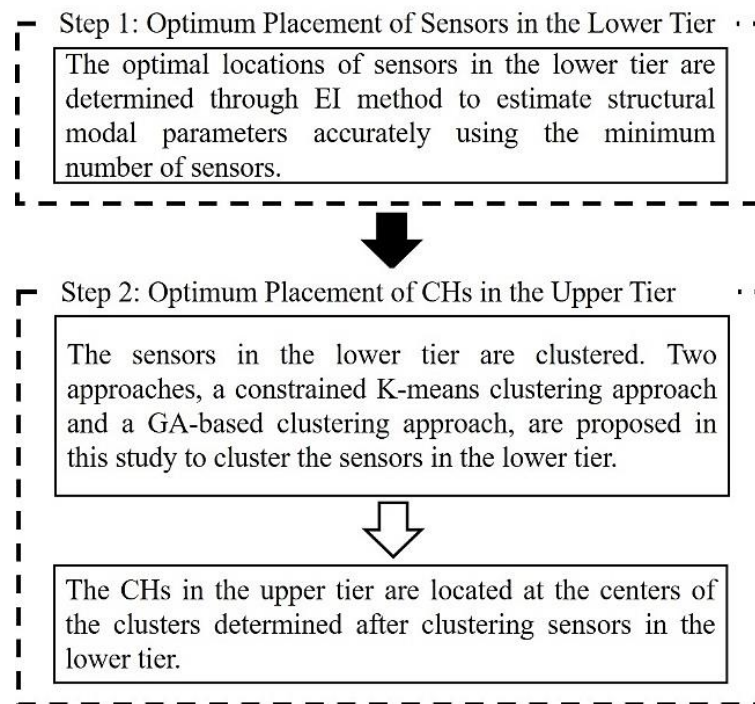


Figure 1. Schematic diagram of the OSP for wireless SHM systems

3. OSP Using the EI Method

The OSP problem for wired SHM systems aims to estimate structural modal parameters accurately by using the minimum number of sensors. In this study, the EI method [2], a widely used OSP technique, is adopted to locate the sensors in the lower tier optimally. The EI method is introduced in the following.

3.1. EI Method

Assuming there are n candidate sensor locations and m target modes, the structural response, y , is estimated as follows:

$$y = \Phi q + w \quad (1)$$

where $\Phi \in R^{n \times m}$ is the target mode shape matrix, q is the modal coordinate, and w is the zero mean white Gaussian noise vector.

The covariance of the error between q and its unbiased estimator \hat{q} , J (defined in Equation 2), is minimized to obtain the best sensor location.

$$J = E[(q - \hat{q})(q - \hat{q})^T] = \left(\frac{1}{\sigma^2} \Phi^T \Phi\right)^{-1} = F^{-1} \quad (2)$$

where σ^2 is the variance of w , the unbiased estimator $\hat{q} = (\Phi^T \Phi)^{-1} \Phi^T$, and F is the Fisher information matrix [26]. Considering that J is equal to the inverse of F , maximizing F will minimize J and achieve the best estimate of q .

The EI method employs the EI distribution vector E_D (defined in Equation 3) to assess the contributions to the independence of the target modes by the candidate sensor locations.

$$E_D = [\Phi \Psi] \otimes [\Phi \Psi] \lambda^{-1} \iota \quad (3)$$

where λ and Ψ are the eigenvalue and eigenvector of the Fisher information matrix F , respectively, ι is an $m \times 1$ vector and each element is equal to one, and the symbol \otimes is the Hadamard product.

Each element of E_D denotes the contribution to the independence of the target modes by the associated sensor location. If the required number of sensors is not reached, the sensor location corresponding to the smallest E_D element is deleted and this process is repeated.

4. Constrained K-Means Clustering Approach

The K-means clustering method is a well-known data-clustering method [27, 28]. Assuming $X = \{x_1 \ x_2 \ \dots \ x_n\}$ is a set of n points to be separated into K clusters, the K-means clustering method finds the cluster centers by minimizing the sum of the squared distance between each point and the associated cluster center $J(C)$.

$$J(C) = \sum_{k=1}^K J(c_k) = \sum_{k=1}^K \sum_{x_i \in c_k} \|x_i - u_k\|^2 \quad (4)$$

where u_k is the mean of cluster c_k , and $J(c_k)$ is the sum of the squared distance between each point of cluster c_k and the cluster center u_k . K-means starts with K initial cluster centers at random positions. Each point belongs to the cluster that has the closest center to it. The K cluster centers need to be recalculated after all points are assigned to their clusters. This process is repeated until all cluster centers no longer move.

In this study, two constraints are considered for the power efficiency problem of WSNs. The first one is the maximum data transmission distance of each sensor (the maximum distance between each sensor and its associated CH). The second one is the maximum number of sensors in each cluster. Given that each CH only manages a limited number of sensors, these constraints should be considered in practice to avoid overload of CHs. Figure 2 presents the flowchart of the constrained K-means clustering approach. Two decisions are made after using the K-means clustering method: the first one is a decision on the limitation of the maximum data transmission distance of each sensor, and the second one is a decision on the limitation of the maximum number of sensors in each cluster. If the two constraints are conformed to, then the clustering process is done; else, one cluster is added, and re-clustering is performed. The process of the proposed constrained K-means clustering approach is detailed below (Figure 2):

Step 1: The cluster number (K), the maximum number of sensors in each cluster (N_{max}), and the maximum data transmission distance of each sensor (d_{max}) are set.

Step 2: The sensors are separated into K clusters by taking K cluster centers initially at random positions.

Step 3: The distances between each sensor and all cluster centers are calculated. Then, each sensor is assigned to the cluster with the closest center to it to form K initial clusters.

Step 4: The positions of cluster centers are recalculated and checked if any cluster center is changed.

Step 5: If any cluster center is changed, then Steps 3 to 5 are repeated; else, Step 6 is performed.

Step 6: If there is any data transmission distance of sensors larger than d_{max} , then one cluster is added ($K=K+1$), and Steps 2 to 6 are repeated; else, Step 7 is performed.

Step 7: If the number of sensors in any cluster is larger than N_{max} , then one cluster is added ($K=K+1$), and Steps 2 to 7 are repeated; else, the clustering process is done.

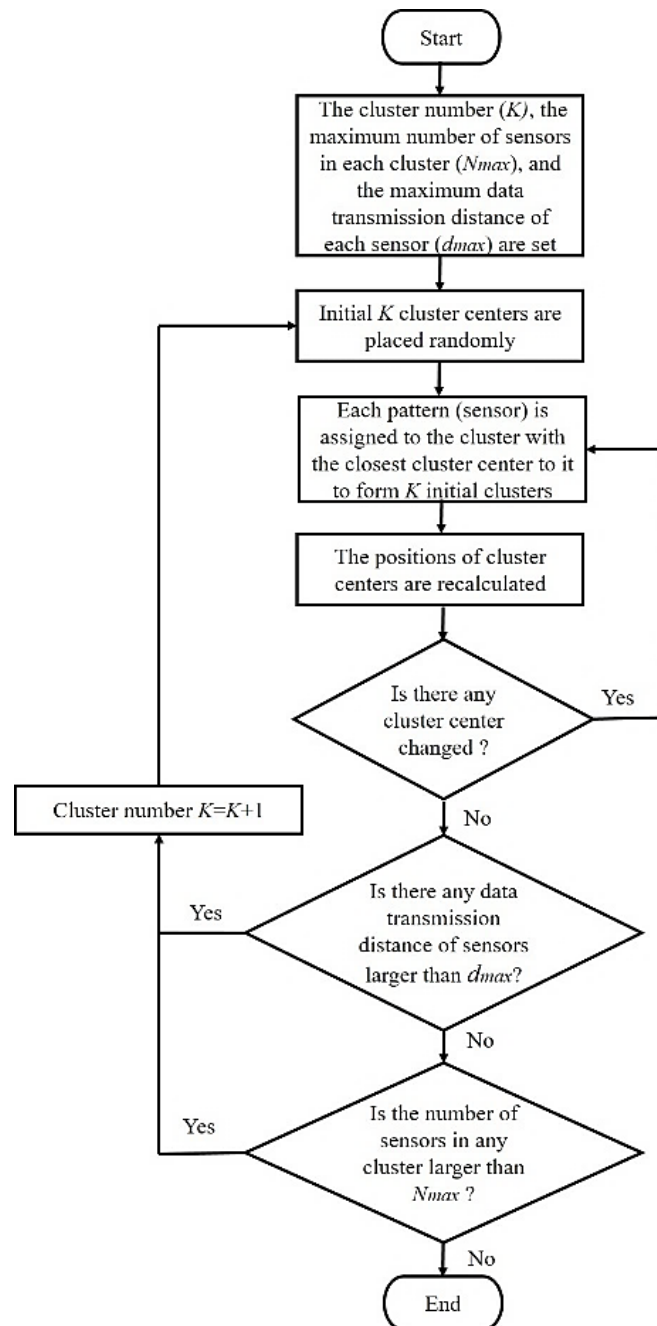


Figure 2. Flowchart of the constrained K-means clustering approach

5. GA-based Clustering Approach

A GA is an optimization method guided by the principles of evolution [15, 16]. Unlike traditional optimization methods, a GA searches for a global optimal solution without calculating the gradient of the objective function.

5.1. Steps of the GA-based Clustering Approach

Herein, a clustering approach using GA for sensor clustering of WSNs is proposed, and Figure 3 displays the flowchart. The process of this approach is detailed in the following:

Step 1: The maximum number of sensors in each cluster (N_{max}), the maximum data transmission distance of each sensor (d_{max}), the maximum number of generations, the mutation rate, and the population size are set. The cluster number (K) is equal to N/N_{max} if the total number of sensors (N) is divisible by N_{max} ; otherwise, K is equal to $N/N_{max} + 1$. Assuming that N has to be divisible by N_{max} in the GA-based clustering approach, some pseudo-sensors are added to N if N is not divisible by N_{max} . The pseudo-sensors are removed after the clustering process is done.

Step 2: Individuals (chromosomes) of the first generation are randomly generated.

Step 3: The fitness value of each chromosome is assessed.

Step 4: Selection: parents for crossover are selected.

Step 5: Crossover: attributes of the parents are mixed to generate better offspring.

Step 6: Mutation: the genetic information is disturbed randomly to have the possibility of leading to a better chromosome.

Step 7: Elite: the elite (the better chromosomes in the current generation and offspring) automatically survives to the next generation.

Step 8: If there is any data transmission distance of sensors larger than d_{max} , then Steps 4 to 8 are repeated; else, Step 9 is performed.

Step 9: Steps 3 to 9 are repeated until the maximum number of generations is reached.

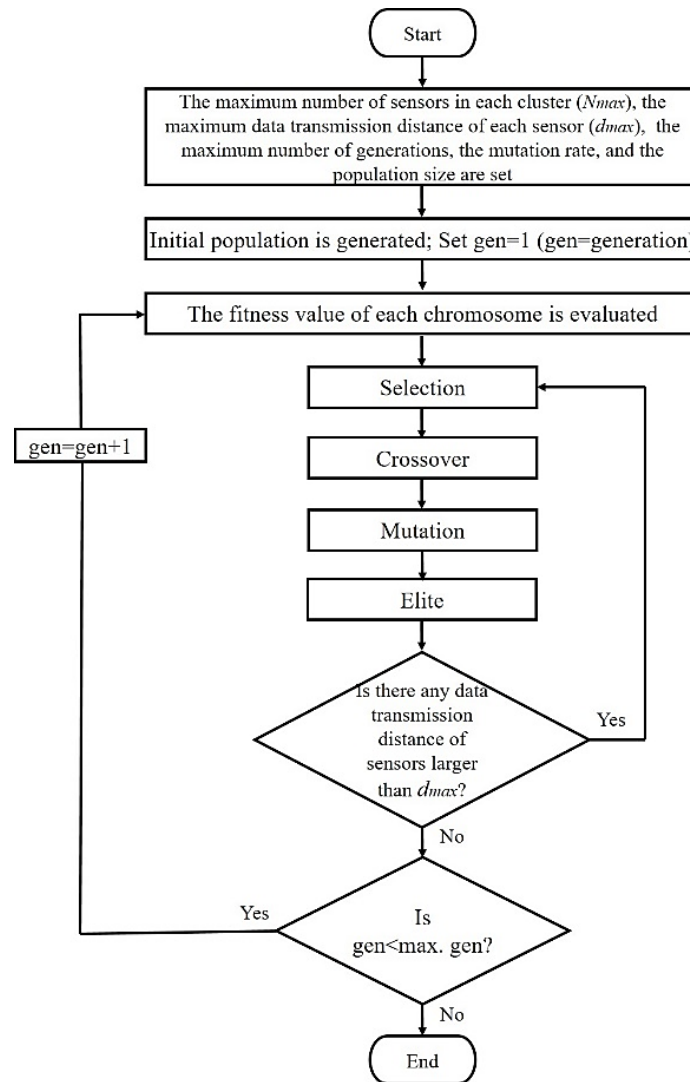


Figure 3. Flowchart of the GA-based clustering approach

5.2. Fundamental Parts of the GA-based Clustering Approach

The fundamental parts of this method are explained in the following.

5.2.1. Population Initialization

In a simple GA [15, 16], a chromosome (or string) is composed of substrings coded by k -bit binary integers, and each substring represents the value of a parameter. However, this is very difficult for sensor clustering of WSNs. Herein, an integer-coded string is designed. All sensors are regarded as a sample string, and the integers in the genes of a string represent the numbering of sensors; that is, the integers in the genes of a string are not duplicated. Figure 4 shows an example of a string representation scheme in the GA-based clustering approach with 10 sensors separated into 5 clusters (each cluster has 2 sensors). The population of the first generation is produced randomly and evolved in an iterative way.

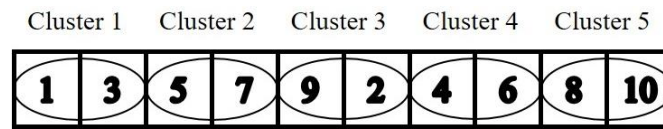


Figure 4. Example of a string representation scheme in the GA-based clustering approach

5.2.2. Fitness Value

The objective function of the proposed GA-based clustering approach is $J(C)$, as defined in Equation 4. The energy dissipation of a sensor to transmit a message, e , can be estimated as follows [11, 29, 30]:

$$e = kd^c \quad (5)$$

where d is the distance between a receiver and a sensor, and k and c are constants (usually $2 < c < 4$). For simplicity, $k=1$ and $c=2$ are considered in this study. Thus the objective function of a string represents its qualification based on energy consumption minimization. Considering that the optimization of the GA maximizes the fitness function and minimizes the objective function, the fitness function of the GA can be defined as a constant minus the objective function.

5.2.3. Selection

Selection is an operator to select parents from the strings in the current generation for crossover based on fitness during each successive generation. There are several selection methods commonly used in GAs, and the roulette-wheel strategy is chosen in this study. The roulette-wheel strategy randomly selects a string from the current generation in accordance with the following probability:

$$P(s_i) = \frac{F(s_i)}{\sum_{j=1}^{NP} F(s_j)} \quad (6)$$

where $F(s_i)$ is the fitness of the string s_i , and NP is the size of the population (NP is an even integer in this study).

5.2.4. Crossover

Crossover is an operator to mix attributes of the parents to generate better offspring. There are several crossover methods commonly used in GAs, such as one-point crossover, two point crossover, and uniform crossover. A portion of genes of the parents is exchanged with each other to create better offspring and to continue the inheritance of superior parents. However, the above crossover operator will generate some offspring with duplicated genes, and this problem is difficult to solve. Accordingly, this study utilizes a crossover method, as shown in Figure 5. First, the first gene of one of the parents is exchanged for that of the other. Second, the duplicated gene is changed to the original first gene of the parent. This study assumes that the parents have produced a total of NP offspring by mating; that is, the crossover rate is assumed to be one.

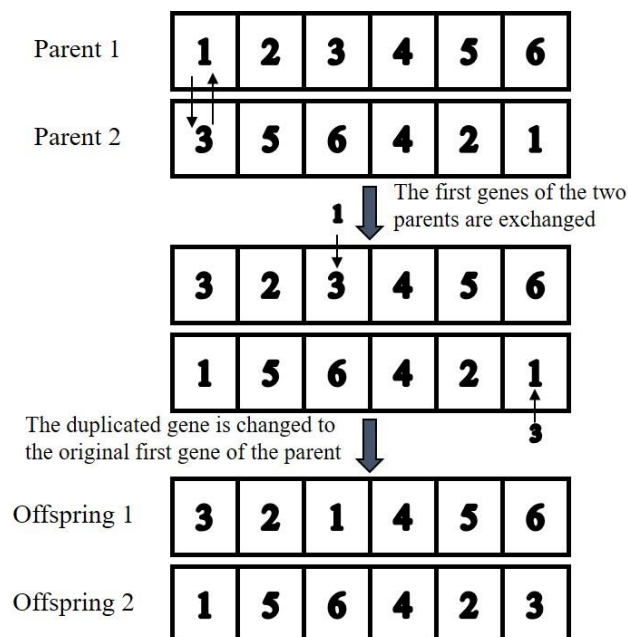


Figure 5. Example of the crossover scheme in the GA-based clustering approach

5.2.5. Mutation

Mutation is an operator to disturb genetic information randomly with a fixed probability p_m to lead to a better individual. For binary integer-coded strings, a gene is mutated by inverting its value (one becomes zero and zero becomes one). In this study, we use the following mutation for integer representation. A string is divided into two parts with an equal number of genes. The mutation probability, p_m , represents the probability to exchange the genes in each position of the two parts. The mutation scheme in the GA-based clustering approach is shown in Figure 6.

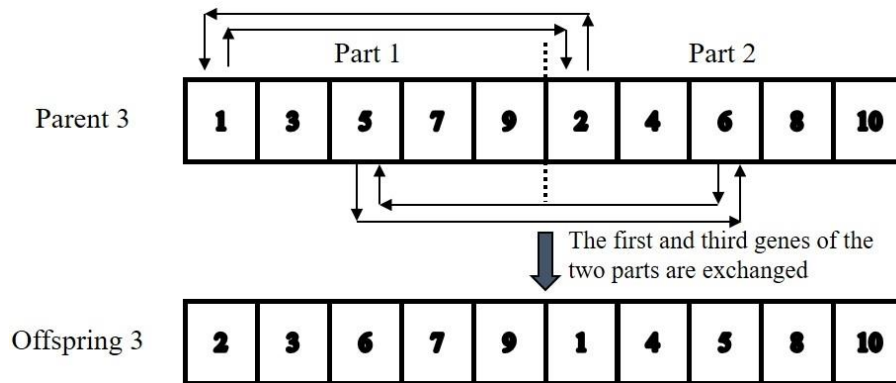


Figure 6. Example of the mutation scheme in the GA-based clustering approach

5.2.6. Elite

After the mutations take place, the elite (the better individuals in the current generation and offspring) automatically survives to the next generation. A better individual is defined as the individual with a smaller sum of the first 80% shortest cluster-based distances in this study. The cluster-based distance is $J(c_k)$ in Equation 4. The reason for using the sum of the first 80% shortest cluster-based distances instead of the sum of all cluster-based distances is that although the individual with a smaller sum of the first 80% shortest cluster-based distances may not be better, some genes of an individual associated with certain cluster-based distances may be the best. Therefore, using an elite based on a smaller sum of the first 80% shortest cluster-based distances may lead to faster convergence.

5.2.7. Locations of CHs

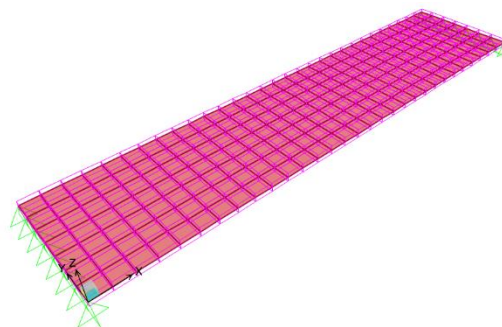
After clustering of sensors, the mean of cluster c_k , u_k , is computed. The locations of sensors and CHs are restricted to nodes with horizontal and vertical integer coordinates in this study. The CH of cluster c_k is located at u_k if u_k is a node with horizontal and vertical integer coordinates, and there is no sensor located at u_k ; otherwise, the node with horizontal and vertical integer coordinates closest to u_k is chosen as the location of the CH for cluster c_k .

6. Numerical Example

To confirm the feasibility of the proposed constrained K-means and GA-based clustering approaches, a numerical example is studied. SAP2000 is adopted for structural analysis in this example.

6.1. Description of the Structural Model

A simply supported deck is selected to illustrate the feasibility of the proposed constrained K-means and GA-based clustering approaches. The deck has dimensions of 15 m in length, 3.5 m in width, and 0.25m in thickness. The elastic modulus $E=24.86$ GPa, and the shear modulus $G=10.35$ GPa. Figure 7 shows the finite element 3D model and joint numbering of the deck. Figures 8 and 9 show the first five 2D and 3D mode shapes of the deck, respectively.

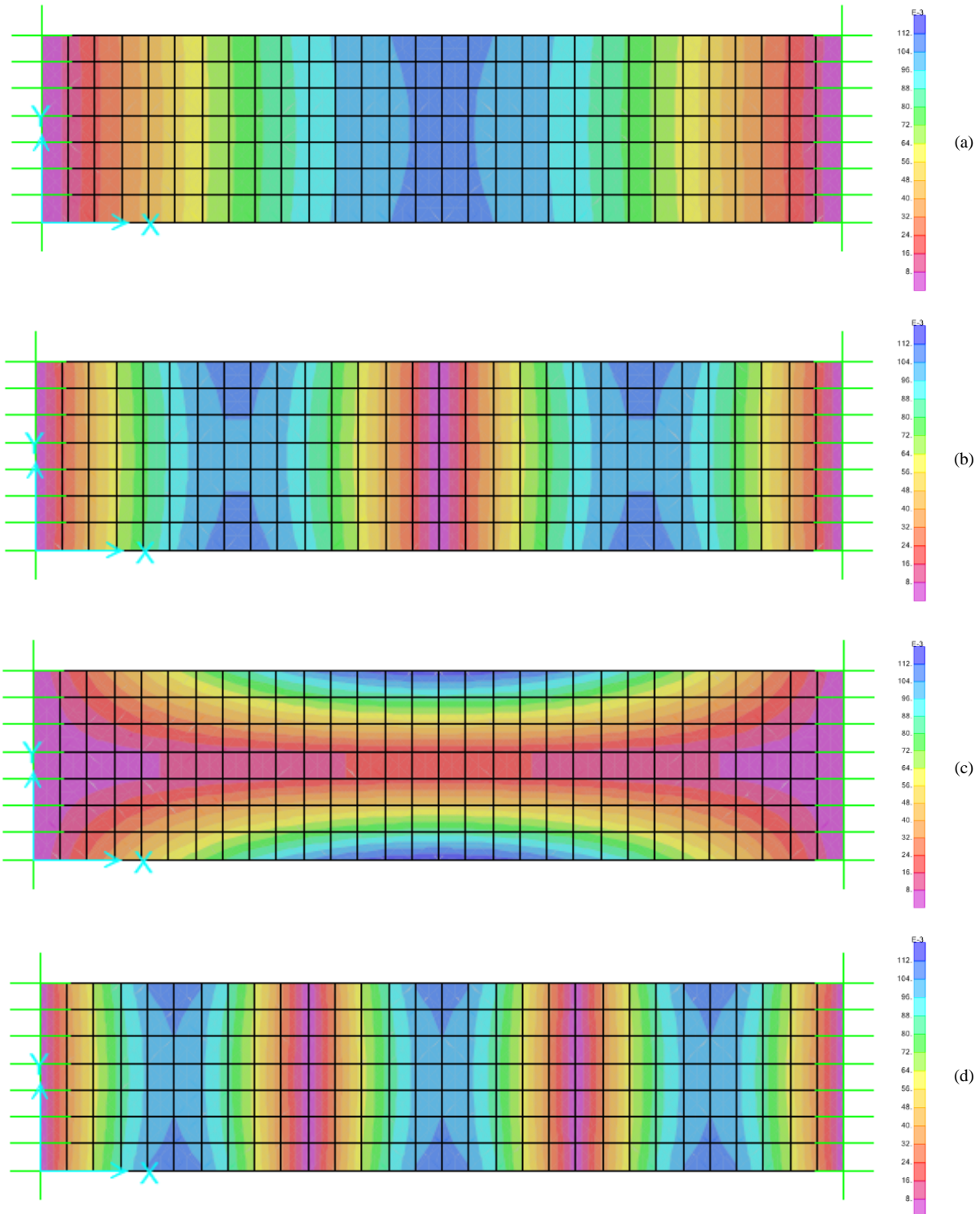


(a)

8	16	24	32	40	48	56	64	72	80	88	96	104	112	120	128	136	144	152	160	168	176	184	192	200	208	216	224	232	240	248
7	15	23	31	39	47	55	63	71	79	87	95	103	111	119	127	135	143	151	159	167	175	183	191	199	207	215	223	231	239	247
6	14	22	30	38	46	54	62	70	78	86	94	102	110	118	126	134	142	150	158	166	174	182	190	198	206	214	222	230	238	246
5	13	21	29	37	45	53	61	69	77	85	93	101	109	117	125	133	141	149	157	165	173	181	189	197	205	213	221	229	237	245
4	12	20	28	36	44	52	60	68	76	84	92	100	108	116	124	132	140	148	156	164	172	180	188	196	204	212	220	228	236	244
3	11	19	27	35	43	51	59	67	75	83	91	99	107	115	123	131	139	147	155	163	171	179	187	195	203	211	219	227	235	243
2	10	18	26	34	42	50	58	66	74	82	90	98	106	114	122	130	138	146	154	162	170	178	186	194	202	210	218	226	234	242
1	9	17	25	33	41	49	57	65	73	81	89	97	105	113	121	129	137	145	153	161	169	177	185	193	201	209	217	225	233	241

(b)

Figure 7. The deck model (a) finite element 3D model (b) joint numbering



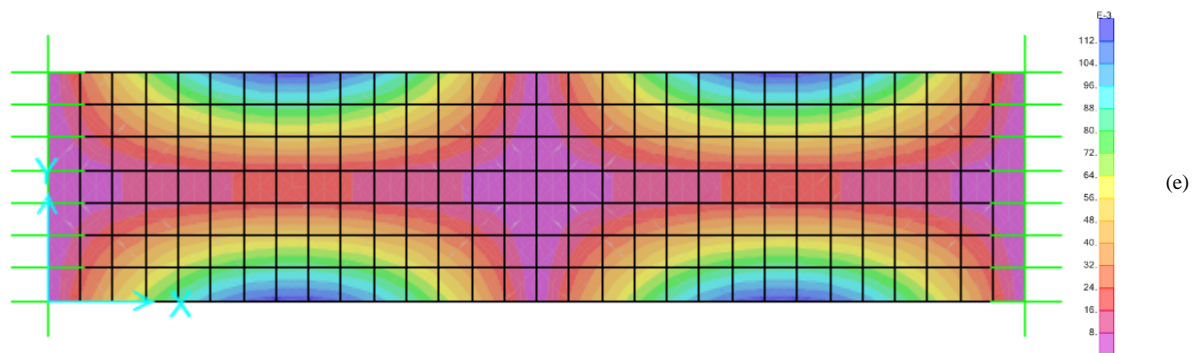
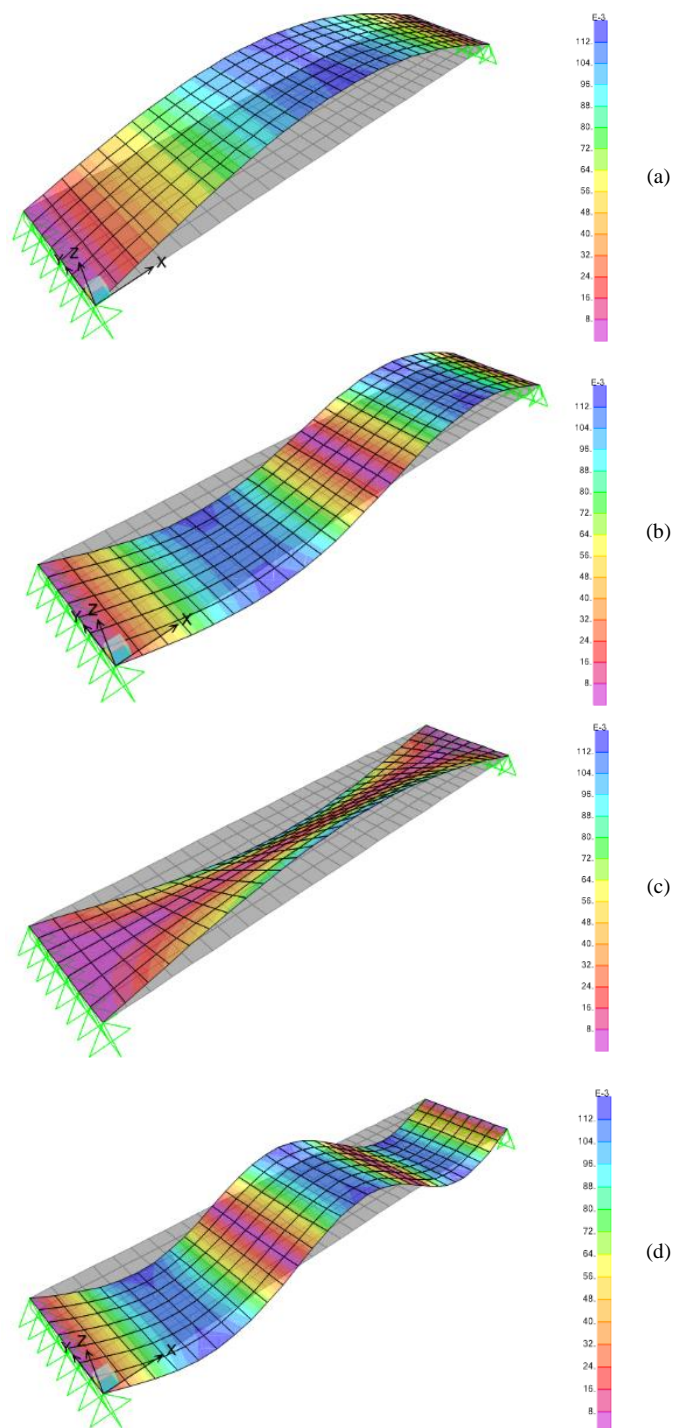


Figure 8. The first five 2D mode shapes of the deck (a) the 1st mode shape (b) the 2nd mode shape (c) the 3rd mode shape (d) the 4th mode shape (e) the 5th mode shape



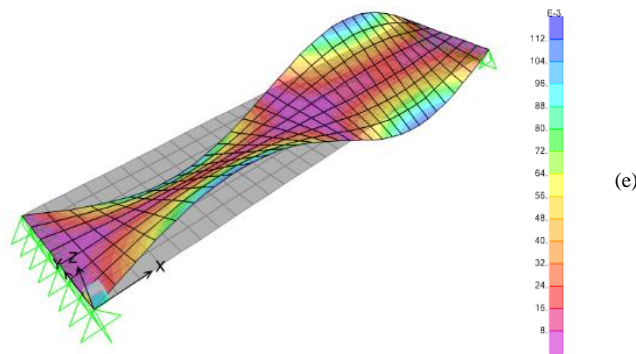


Figure 9. The first five 3D mode shapes of the deck (a) the 1st mode shape (b) the 2nd mode shape (c) the 3rd mode shape (d) the 4th mode shape (e) the 5th mode shape

6.2. Results and Discussion

6.2.1. Main Findings of the Present Study

6.2.1.1. Sensor Location Using the EI Method

The first five modal properties analyzed by SAP2000 are used for searching the optimal locations of sensors in the lower tier in this study. The optimal locations of 10, 15, and 30 vertical wireless sensors in the lower tier are determined using the EI method, and Figure 10 shows the results. Figures 8 and 9 show that along the x direction of the deck, the deformations of both sides are larger than that of the middle part for the five mode shapes. The sensor location with larger modal deformation contributes more to the independence of the target modes. Figure 10 shows that the locations of fewer sensors are within those of more sensors. That is, the locations of more sensors spread from those of fewer sensors to the two simply supported ends of the deck.

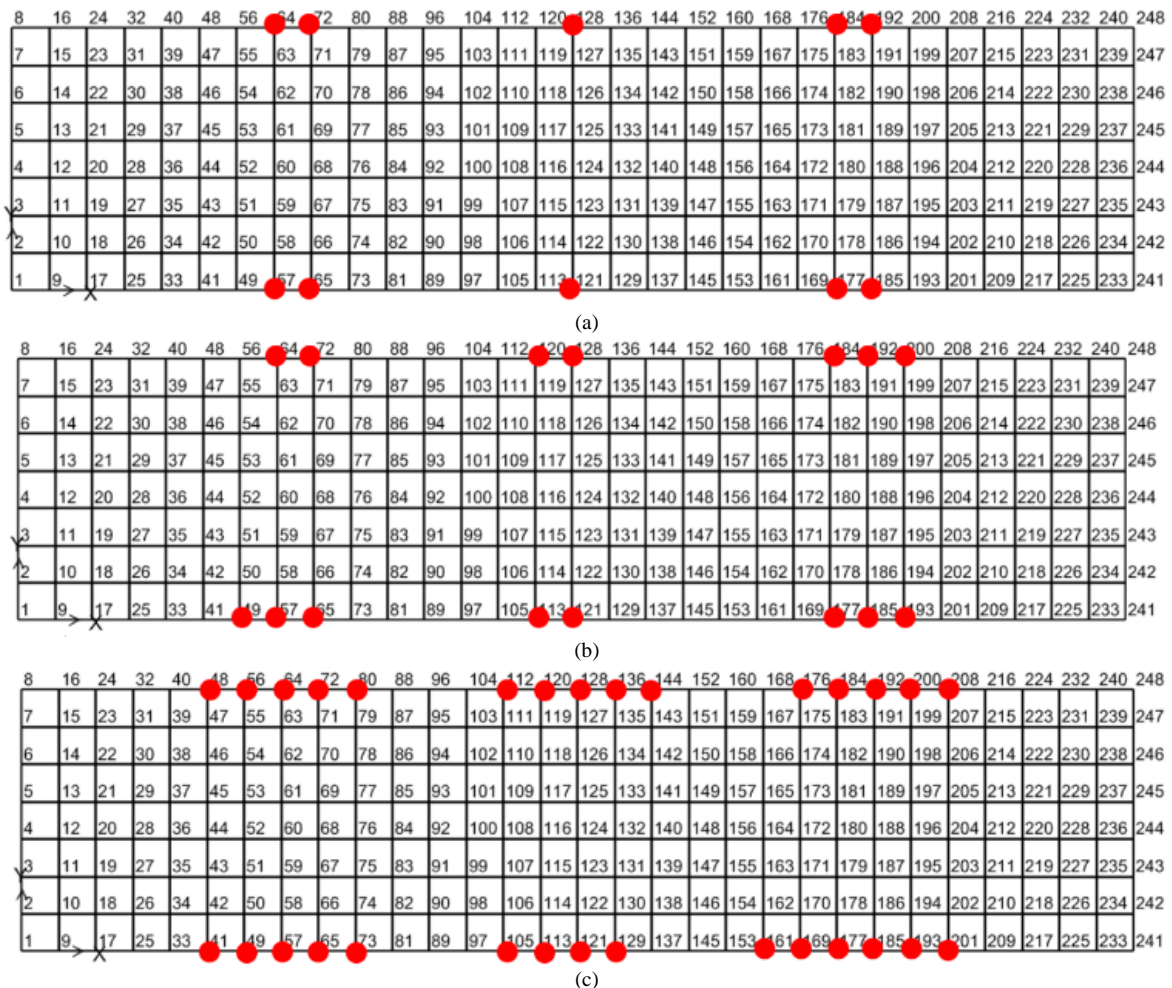


Figure 10. Optimum sensor locations of the deck with (a) 10 sensors (b) 15 sensors (c) 30 sensors

6.2.1.2. Clustering Sensors Using the Constrained K-Means and GA-based Clustering Approaches

Herein, the result of 30 sensors (Figure 10-c) is implemented by the proposed constrained K-means clustering approach. The maximum number of sensors in each cluster (N_{max}) is assumed to be 5, and the maximum data transmission distance of each sensor (d_{max}) is 8 m. The implementation results are shown in Figure 11. The 30 sensors are separated into 11 clusters. The number of sensors in each cluster changes from 2 to 5. Owing to the large variation in the number of sensors in each cluster, constructing a balanced size of clusters by using the constrained K-means clustering approach is not easy.

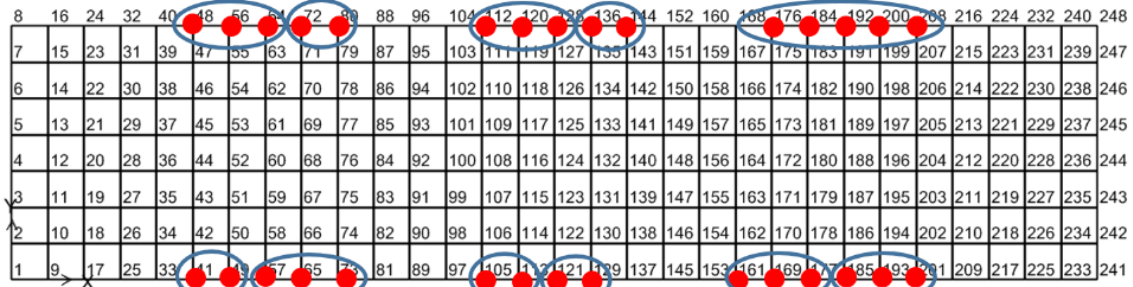
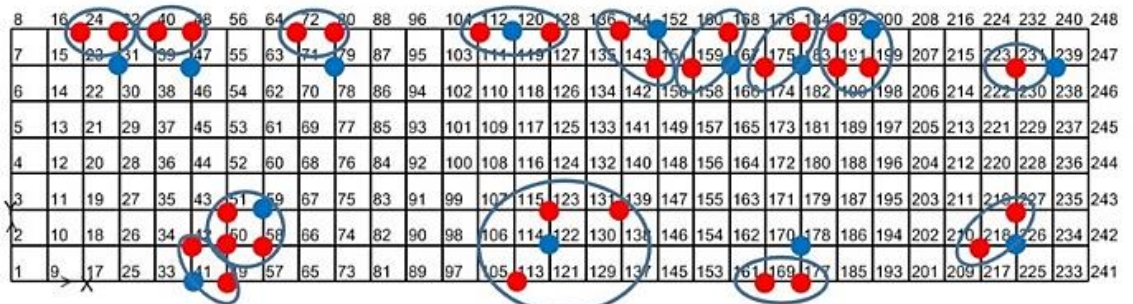
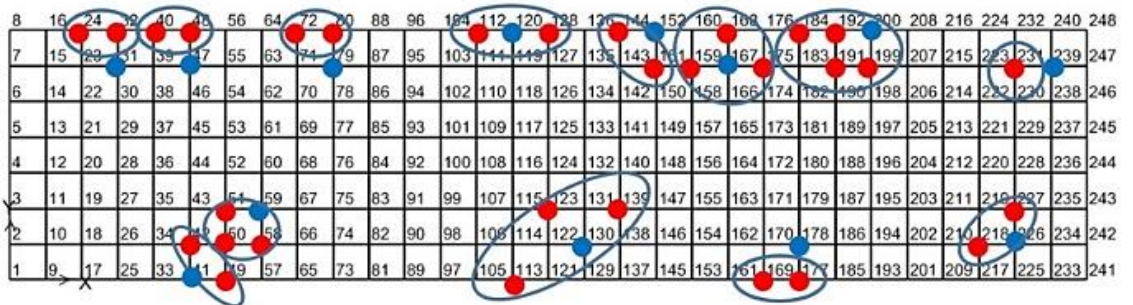


Figure 11. Clustering 30 sensors (deployed by EI method) implemented by using the constrained K-means clustering approach with $N_{max}=5$

Figure 10-c shows that the optimum locations of the 30 sensors solved using the EI method are uniformly distributed on both sides along the x direction. To compare the performances of the two proposed approaches, the 30 sensors are deployed randomly, then the randomly placed 30 sensors are clustered using the two proposed approaches. The maximum data transmission distance of each sensor (d_{max}) is set to 8 m. Figures 12-a, 12-b, and 12-c present the clustering results using the constrained K-means clustering approach for N_{max} equal to 3, 4, and 5, respectively. The number of clusters is 14, 13, and 9 for N_{max} equal to 3, 4, and 5, respectively. The number of sensors for each cluster changes from 1 to 3, 1 to 4, and 2 to 5 for N_{max} equal to 3, 4, and 5, respectively. After clustering sensors, the locations of CHs are determined and shown in Figure 12.



(a)



(b)

● : cluster head
● : sensor unit

● : cluster head
● : sensor unit

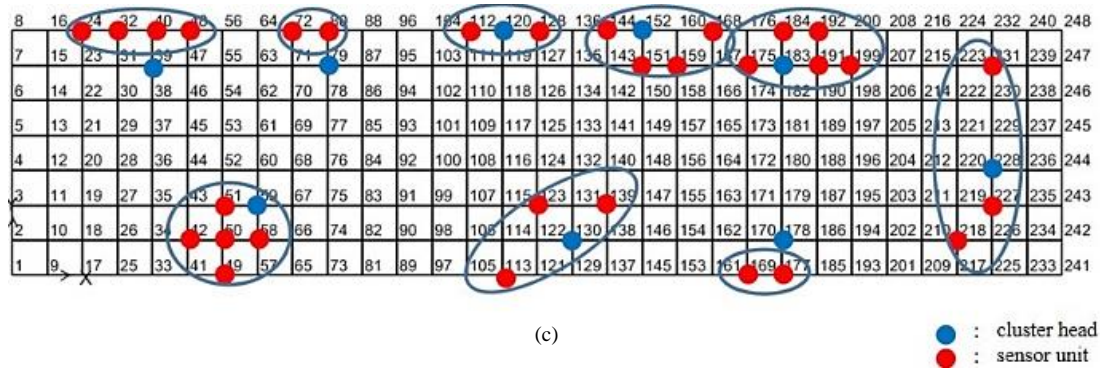


Figure 12. Clustering the randomly placed 30 sensors and deploying cluster heads implemented by using the constrained K-means clustering approach with (a) $N_{max}=3$ (b) $N_{max}=4$ (c) $N_{max}=5$

Next, the GA-based clustering approach is conducted with the following parameters: the population size NP and mutation probability p_m are set to 300 and 0.05, respectively. Figures 13-a, 13-b, and 13-c display the clustering results using the GA-based clustering approach for N_{max} equal to 3, 4, and 5, respectively. The number of clusters is 10, 8, and 6 for N_{max} equal to 3, 4, and 5, respectively. The number of sensors in all clusters is the same for N_{max} equivalent to 3 and 5. The number of sensors in each cluster changes from 2 to 4 for N_{max} equivalent to 4. In fact, the total number of sensors (30) is not divisible by N_{max} (4) for this case. Seven clusters have four sensors, and only one cluster has two sensors

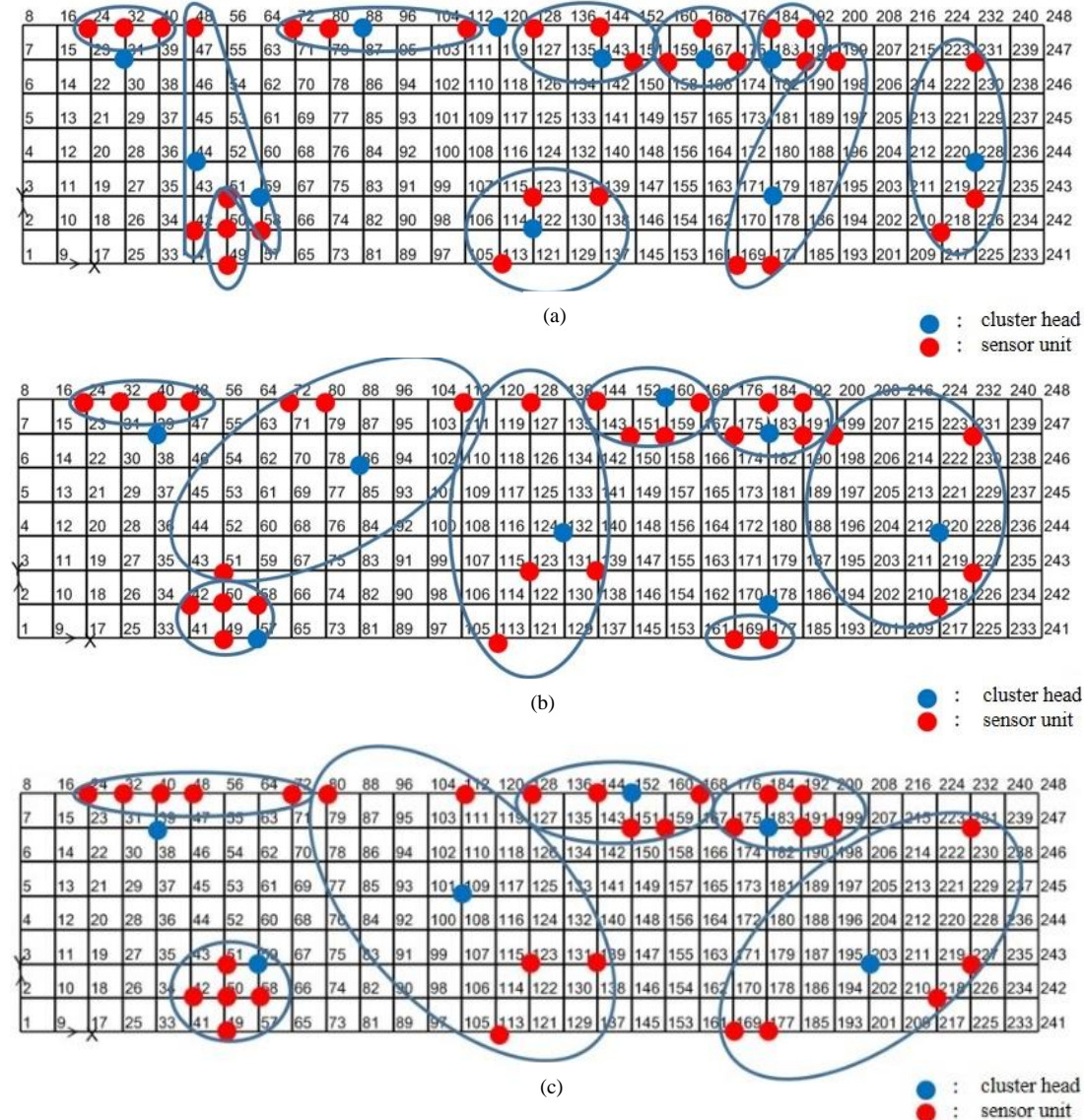


Figure 13. Clustering the randomly placed 30 sensors and deploying cluster heads implemented by using the GA-based clustering approach with (a) $N_{max}=3$ (b) $N_{max}=4$ (c) $N_{max}=5$

Three indices proposed by Hassan et al. [25] are used herein to evaluate the performance of the constrained K-means and GA-based clustering approaches in the formation of a balanced size of clusters. The three indices are the standard deviation of mean squared error ($STD(MSE)$), the variation in cluster size (V), and the cluster size range (CSR), which are introduced as follows:

- Standard deviation of mean squared error $STD(MSE)$: $STD(MSE)$, as shown in Equation 7, measures the difference in homogeneity of the average of intra-distance for each cluster.

$$STD(MSE) = \sqrt{\frac{\sum_{j=1}^K |MSE_j - \mu_{MSE}|^2}{K}} \quad (7)$$

$$MSE_j = \frac{1}{N_j} \sum_{i=1}^{N_j} (x_{ij} - u_j)^2 \quad (8)$$

$$\mu_{MSE} = \frac{\sum_{j=1}^K MSE_j}{K} \quad (9)$$

where MSE_j is the averaged square intra-distances of sensor nodes to the cluster's center for the j th cluster; μ_{MSE} is the averaged mean squared error for distances; K is the number of clusters; N_j is the number of sensor nodes in the j th cluster; x_{ij} and u_j are the sensor node i and cluster center for the j th cluster, respectively. A smaller $STD(MSE)$ means a higher uniformity of the intra-distances for clusters.

- Variation in cluster size V : V , as shown in Equation 10, measures the dissimilarity of the density of the sensor nodes in the clusters (number of sensor nodes in each cluster).

$$V = \frac{\sum_{j=1}^K |S_j - \mu_{S_j}|}{K} \quad (10)$$

$$\mu_{S_j} = \frac{\sum_{j=1}^K S_j}{K} \quad (11)$$

where S_j and μ_{S_j} refer to the cluster size and the mean of cluster size for the j th cluster, respectively. A smaller V means a higher balance in cluster size.

- Cluster size range CSR : CSR , as shown in Equation 12, measures the ratio of the minimum cluster size to the maximum cluster size.

$$CSR = \frac{CS_{min}}{CS_{max}} \quad (12)$$

where CS_{min} and CS_{max} are the minimum and maximum cluster sizes in the network, respectively. CSR takes a value between 0 and 1. A narrower range (CSR close to 1) means a smaller difference in size between the minimum and maximum cluster sizes, and it is better.

Table 1 lists the three indices of the constrained K-means and GA-based clustering approaches for N_{max} equal to 3, 4, and 5. The values of $STD(MSE)$ of the constrained K-means clustering approach are smaller than those of the GA-based clustering approach for N_{max} equal to 3, 4, and 5. Nevertheless, this finding does not imply that the constrained K-means clustering approach outperforms the GA-based clustering approach because $STD(MSE)$ is meaningful for comparing different approaches only when the number of clusters is the same. The reason why the value of $STD(MSE)$ of the constrained K-means clustering approach is smaller than that of the GA-based clustering approach is that the number of clusters of the former is larger than that of the latter under the same d_{max} and N_{max} constraints. A smaller number of clusters implies a smaller number of CHs. Given that CHs consume more power than other sensors, the power consumption implemented by the GA-based clustering approach seems smaller than that by the constrained K-means clustering approach. The values of V of the constrained K-means clustering approach are larger than those of the GA-based clustering approach for N_{max} equal to 3, 4, and 5. Meanwhile, the values of CSR of the constrained K-means clustering approach are smaller than those of the GA-based clustering approach for N_{max} equal to 3, 4, and 5. Based on the results of V and CSR , the GA-based clustering approach is better than the constrained K-means clustering approach for constructing a balanced cluster size. After clustering sensors, the locations of CHs are determined and shown in Figure 13.

6.2.2. Comparison the Constrained K-Means and GA-Based Clustering Approaches with the K-Means Clustering Method

In fact, there is no clustering approach proposed in other studies considering the two constraints, the maximum data transmission distance of each sensor (d_{max}) and the maximum number of sensors in each cluster (N_{max}). Therefore, it is difficult to compare the performances of the constrained K-means and GA-based clustering approaches with other clustering approach. Nevertheless, the performances of the two proposed clustering approaches are still compared with that of the K-means clustering method herein. Table 1 also lists the three indices of the K-means clustering method for the number of clusters equal to 6, 8, 9, 10, 13, and 14.

Table 1. $STD(MSE)$, V , and CSR of the constrained K-means Clustering approach, the GA-based clustering approach, and the K-means Clustering method

Approach	d_{max}	N_{max}	Number of clusters	$STD(MSE)$	V	CSR
Constrained K-means clustering approach	8 m	3	14	0.566	0.209	0.333
	8 m	4	13	0.592	0.521	0.250
	8 m	5	9	1.394	0.988	0.400
GA-based clustering approach	8 m	3	10	3.245	0	1
	8 m	4	8	4.029	0.438	0.5
	8 m	5	6	4.429	0	1
K-means clustering method	/	/	6	5.661	7	0.2
	/	/	8	1.579	1.438	0.333
	/	/	9	1.635	1.333	0.333
	/	/	10	0.696	1.8	0.2
	/	/	13	0.592	0.521	0.25
	/	/	14	0.490	1.265	0.2

Note: “/”: no data available

Comparing the constrained K-means clustering approach for $d_{max}=8$ m and $N_{max}=3$ (the number of clusters is equal to 14) with the K-means clustering method for the number of clusters equal to 14, the K-means clustering method conforms to the constraint of $d_{max}=8$ m, but does not conform to the constraint of $N_{max}=3$ (the largest cluster has 5 sensor). According to $STD(MSE)$, the K-means clustering method outperforms the constrained K-means clustering approach (the value of $STD(MSE)$ of the K-means clustering method is smaller than that of the constrained K-means clustering approach). However, the constrained K-means clustering approach outperforms the K-means clustering method according to V and CSR (the value of V of the constrained K-means clustering approach is smaller than that of the K-means clustering method, and the value of CSR of the constrained K-means clustering approach is larger than that of the K-means clustering method).

Comparing the constrained K-means clustering approach for $d_{max}=8$ m and $N_{max}=4$ (the number of clusters is equal to 13) with the K-means clustering method for the number of clusters equal to 13. The K-means clustering method conforms to the constraints of $d_{max}=8$ m and $N_{max}=4$ (the largest cluster has 4 sensor). The constrained K-means clustering approach and the K-means clustering method have nine identical clusters and only two different clusters. The values of $STD(MSE)$, V , and CSR of the K-means clustering method are all the same with those of the constrained K-means clustering approach; that is the performances of the K-means clustering method and the constrained K-means clustering approach are the same.

Comparing the constrained K-means clustering approach for $d_{max}=8$ m and $N_{max}=5$ (the number of clusters is equal to 9) with the K-means clustering method for the number of clusters equal to 9, the K-means clustering method conforms to the constraint of $d_{max}=8$ m, but does not conform to the constraint of $N_{max}=5$ (the largest cluster has 6 sensor). The constrained K-means clustering approach outperforms the K-means clustering method according to $STD(MSE)$, V , and CSR (the values of $STD(MSE)$ and V of the constrained K-means clustering approach are smaller than those of the K-means clustering method, and the value of CSR of the constrained K-means clustering approach is larger than that of the K-means clustering method).

Comparing the GA-based clustering approach for $d_{max}=8$ m and $N_{max}=3$ (the number of clusters is equal to 10) with the K-means clustering method for the number of clusters equal to 10, the K-means clustering method conforms to the constraint of $d_{max}=8$ m, but does not conform to the constraint of $N_{max}=3$ (the largest cluster has 5 sensor). According to $STD(MSE)$, the K-means clustering method outperforms the GA-based clustering approach (the value of $STD(MSE)$ of the K-means clustering method is smaller than that of the GA-based clustering approach). However, the GA-based clustering approach outperforms the K-means clustering method according to V and CSR (the value of V of the GA-based clustering approach is smaller than that of the K-means clustering method, and the value of CSR of the GA-based clustering approach is larger than that of the K-means clustering method).

Comparing the GA-based clustering approach for $d_{max}=8$ m and $N_{max}=4$ (the number of clusters is equal to 8) with the K-means clustering method for the number of clusters equal to 8, the K-means clustering method conforms to the constraint of $d_{max}=8$ m, but does not conform to the constraint of $N_{max}=4$ (the largest cluster has 6 sensor). According to $STD(MSE)$, the K-means clustering method outperforms the GA-based clustering approach (the value of $STD(MSE)$ of the K-means clustering method is smaller than that of the GA-based clustering approach). However, the GA-based clustering approach outperforms the K-means clustering method according to V and CSR (the value of V of the GA-based clustering approach is smaller than that of the K-means clustering method, and the value of CSR of the GA-based clustering approach is larger than that of the K-means clustering method).

Comparing the GA-based clustering approach for $d_{max}=8$ m and $N_{max}=5$ (the number of clusters is equal to 6) with the K-means clustering method for the number of clusters equal to 6, the K-means clustering method conforms to the constraint of $d_{max}=8$ m, but does not conform to the constraint of $N_{max}=5$ (the largest cluster has 10 sensor). The GA-based clustering approach outperforms the K-means clustering method according to $STD(MSE)$, V , and CSR (the values of $STD(MSE)$ and V of the GA-based clustering approach are smaller than those of the K-means clustering method, and the value of CSR of the GA-based clustering approach is larger than that of the K-means clustering method).

6.2.3. Implication and Explanation of Findings

The constrained K-means and GA-based clustering approaches construct a balanced size of clusters by minimizing the number of clusters (or CHs) and the total sum of the squared distance between each sensor and its associated CH under the two constraints. Accordingly, the energy consumption in each cluster is decreased and balanced, and the network lifetime is extended.

The GA-based clustering approach is better than the constrained K-means clustering approach for constructing a balanced cluster size based on the results of V and CSR . Hence, the consumption of energy in each cluster implemented by the GA-based clustering approach is more balanced than that by the constrained K-means clustering approach. A balanced consumption of energy in clusters will prolong the network lifetime. Nevertheless, the GA-based clustering approach is slow in execution compared with the constrained K-means clustering approach. Although the constrained K-means clustering approach may not have the best solution, it is still acceptable for clustering numerous sensors.

Generally speaking, the K-means clustering method without any constraints, and its result is not easy to conform to practical constraints. In terms of constructing a balanced cluster size, the constrained K-means clustering approach and the GA-based clustering approach outperform the K-means clustering method for a small number of clusters based on $STD(MSE)$, V , and CSR (the number of clusters equal to 9 and 6, for example). The fewer clusters, the fewer CHs. Since CHs consume more energy than other sensors, using fewer CHs will save more energy and extend the network lifetime.

7. Conclusions

This study investigates the OSP of two-tier WSN-based SHM systems with clusters of sensors. The lower tier is composed of sensors communicating with their associated CHs, and the upper tier is composed of the network of CHs. The first step is the optimal placement of sensors in the lower tier via the EI method by considering modal identification accuracy. The second step is the optimal placement of CHs in the upper tier by considering power efficiency. The sensors in the lower tier are clustered before determining the optimal locations of CHs in the upper tier. Two approaches, a constrained K-means clustering approach and a GA-based clustering approach, are proposed to cluster sensors in the lower tier. The CHs in the upper tier are located at the centers of the clusters determined after clustering the sensors in the lower tier. The two proposed clustering approaches construct a balanced size of clusters by minimizing the number of clusters (or CHs) and the total sum of the squared distance between each sensor and its associated CH under the two constraints. Accordingly, the energy consumption in each cluster is decreased and balanced, and the network lifetime is extended. To illustrate the feasibility of the two proposed approaches, a numerical example is studied. In this example, the performances of these clustering approaches are also compared. Important conclusions are summarized in the following:

- The optimal locations of different numbers of sensors solved by using the EI method have an inheritable characteristic; that is, the locations of fewer sensors are within those of more sensors.
- The performance for the constrained K-means and the GA-based clustering approaches in the formation of a balanced size of clusters cannot be evaluated by $STD(MSE)$ because $STD(MSE)$ is meaningful for comparing different approaches only when the number of clusters are the same.
- The number of clusters implemented by the constrained K-means clustering approach is larger than that by the GA-based clustering approach under the same d_{max} and N_{max} constraints. A larger number of clusters implies a larger number of CHs. Given that CHs consume more power than other sensors, the power consumption implemented by the GA-based clustering approach seems smaller than that by the constrained K-means clustering approach.
- According to V and CSR , the GA-based clustering approach is better than the constrained K-means clustering approach for constructing a balanced cluster size. In other words, the consumption of energy in each cluster implemented by the GA-based clustering approach is more balanced than that by the constrained K-means clustering approach. A balanced consumption of energy in clusters will extend the network lifetime.
- The GA-based clustering approach is slow in execution compared with the constrained K-means clustering approach. Although the constrained K-means clustering approach may not have the best solution, it is still acceptable for clustering numerous sensors.

- Since the K-means clustering method without any constraint, its results is not easy to conform to practical constraints. In terms of constructing balanced cluster size, the constrained K-means and the GA-based clustering approaches outperform the K-means clustering method for a small number of clusters based on $STD(MSE)$, V , and CSR . The fewer clusters, the fewer CHs. Since CHs consume more energy than other sensors, using fewer CHs will save more energy.
- How to apply the proposed approaches to complicated real structures would be explored on the basis of this research in the future.

8. Declarations

8.1. Author Contributions

Conceptualization, S.L.H.; methodology, S.L.H.; software, J.W.H.; validation, J.W.H., C.Y.K. and S.L.H.; formal analysis, J.W.H.; investigation, J.W.H.; resources, S.L.H.; data curation, J.W.H. and C.Y.K.; writing—original draft preparation, J.W.H.; writing—review and editing, C.Y.K.; visualization, J.W.H. and C.Y.K.; supervision, S.L.H.; project administration, S.L.H.; funding acquisition, S.L.H. All authors have read and agreed to the published version of the manuscript.

8.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

8.3. Funding

This work was supported by the Ministry of Science and Technology of the Republic of China under research grants MOST 106-2221-E-009 -052 -MY3 and MOST 103-2625-M-009-003.

8.4. Acknowledgements

The authors would like to thank the Ministry of Science and Technology of the Republic of China for financially supporting this research.

8.5. Conflicts of Interest

The authors declare no conflict of interest.

9. References

- [1] Cha, Y. J., Agrawal, A. K., Kim, Y., & Raich, A. M. (2012). Multi-objective genetic algorithms for cost-effective distributions of actuators and sensors in large structures. *Expert Systems with Applications*, 39(9), 7822–7833. doi:10.1016/j.eswa.2012.01.070.
- [2] Kammer, D. C. (1990). Sensor Placement for On-Orbit Modal Identification and Correlation of Large Space Structures. 1990 American Control Conference. doi:10.23919/acc.1990.4791265.
- [3] Spencer, B. F., Ruiz-Sandoval, M. E., & Kurata, N. (2004). Smart sensing technology: Opportunities and challenges. *Structural Control and Health Monitoring*, 11(4), 349–368. doi:10.1002/stc.48.
- [4] Lynch, J. P. (2006). A Summary Review of Wireless Sensors and Sensor Networks for Structural Health Monitoring. *The Shock and Vibration Digest*, 38(2), 91–128. doi:10.1177/0583102406061499.
- [5] Lin, T. H., Lu, Y. C., & Hung, S. L. (2014). Locating damage using integrated global-local approach with wireless sensing system and single-chip impedance measurement device. *The Scientific World Journal*, 2014, 729027. doi:10.1155/2014/729027.
- [6] Kottapalli, V. A., Kiremidjian, A. S., Lynch, J. P., Carryer, E., Kenny, T. W., Law, K. H., & Lei, Y. (2003). Two-tiered wireless sensor network architecture for structural health monitoring. *Smart Structures and Materials 2003: Smart Systems and Nondestructive Evaluation for Civil Infrastructures*. doi:10.1117/12.482717.
- [7] Zhou, G.-D., Yi, T.-H., Xie, M.-X., Li, H.-N., & Xu, J.-H. (2021). Optimal Wireless Sensor Placement in Structural Health Monitoring Emphasizing Information Effectiveness and Network Performance. *Journal of Aerospace Engineering*, 34(2), 4020112. doi:10.1061/(asce)as.1943-5525.0001226.
- [8] Abdulkarem, M., Samsudin, K., Rokhani, F. Z., & A Rasid, M. F. (2020). Wireless sensor network for structural health monitoring: A contemporary review of technologies, challenges, and future direction. *Structural Health Monitoring*, 19(3), 693–735. doi:10.1177/1475921719854528.
- [9] Hung, S. L., Ding, J. T., & Lu, Y. C. (2019). Developing an energy-efficient and low-delay wake-up wireless sensor network-based structural health monitoring system using on-site earthquake early warning system and wake-on radio. *Journal of Civil Structural Health Monitoring*, 9(1), 103–115. doi:10.1007/s13349-018-0315-2.

- [10] Hussain, S., Matin, A. W., & Islam, O. (2007). Genetic algorithm for hierarchical wireless sensor networks. *Journal of Networks*, 2(5), 87–97. doi:10.4304/jnw.2.5.87-97.
- [11] Heinzelman, W. R., Chandrakasan, A., & Balakrishnan, H. (2000). Energy-efficient communication protocol for wireless microsensor networks. *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*, Maui, HI, USA. doi:10.1109/hicss.2000.926982.
- [12] Sasikumar, P., & Khara, S. (2012). K-Means Clustering in Wireless Sensor Networks. *2012 Fourth International Conference on Computational Intelligence and Communication Networks*. doi:10.1109/cicn.2012.136.
- [13] Ray, A., & De, D. (2016). Energy efficient clustering protocol based on K-means (EECPK-means)-midpoint algorithm for enhanced network lifetime in wireless sensor network. *IET Wireless Sensor Systems*, 6(6), 181–191. doi:10.1049/iet-wss.2015.0087.
- [14] Periyasamy, S., Khara, S., & Thangavelu, S. (2016). Balanced Cluster Head Selection Based on Modified k -Means in a Distributed Wireless Sensor Network. *International Journal of Distributed Sensor Networks*, 2016(3), 5040475. doi:10.1155/2016/5040475.
- [15] Holland, J. H. (2019). *Adaptation in Natural and Artificial Systems*. The University of Michigan Press, Ann Arbor, United States. doi:10.7551/mitpress/1090.001.0001.
- [16] Goldberg, D. E. (1989). Genetic algorithms in search, optimization, and machine learning. (1989). *Choice Reviews Online*, 27(02), 27-0936-27-0936. doi:10.5860/choice.27-0936.
- [17] Adeli, H., & Hung, S. L. (1994). *Machine learning: neural networks, genetic algorithms, and fuzzy systems*. John Wiley & Sons, Hoboken, United States.
- [18] Jin, S., Zhou, M., & Wu, A. S. (2003). Sensor network optimization using a genetic algorithm. *Proceedings of the 7th world multiconference on systemics, cybernetics and informatics*, 27-30July, 2003, Orlando, United States.
- [19] Ferentinos, K. P., & Tsiligiridis, T. A. (2007). Adaptive design optimization of wireless sensor networks using genetic algorithms. *Computer Networks*, 51(4), 1031–1051. doi:10.1016/j.comnet.2006.06.013.
- [20] Peiravi, A., Mashhadi, H. R., & Hamed Javadi, S. (2013). An optimal energy-efficient clustering method in wireless sensor networks using multi-objective genetic algorithm. *International Journal of Communication Systems*, 26(1), 114–126. doi:10.1002/dac.1336.
- [21] Nayak, P., & Vathasavai, B. (2017). Genetic algorithm based clustering approach for wireless sensor network to optimize routing techniques. *2017 7th International Conference on Cloud Computing, Data Science & Engineering - Confluence*. doi:10.1109/confluence.2017.7943178.
- [22] Pal, R., Yadav, S., Karnwal, R., & Aarti. (2020). EEWC: energy-efficient weighted clustering method based on genetic algorithm for HWSNs. *Complex & Intelligent Systems*, 6(2), 391–400. doi:10.1007/s40747-020-00137-4.
- [23] Bhola, J., Soni, S., & Cheema, G. K. (2020). Genetic algorithm based optimized leach protocol for energy efficient wireless sensor networks. *Journal of Ambient Intelligence and Humanized Computing*, 11(3), 1281–1288. doi:10.1007/s12652-019-01382-3.
- [24] Khoshraftar, K., & Heidari, B. (2020). A Hybrid Method Based on Clustering to Improve the Reliability of the Wireless Sensor Networks. *Wireless Personal Communications*, 113(2), 1029–1049. doi:10.1007/s11277-020-07266-6.
- [25] Hassan, A. A. hussian, Shah, W. M., Othman, M. F. I., & Hassan, H. A. H. (2020). Evaluate the performance of K-Means and the fuzzy C-Means algorithms to formation balanced clusters in wireless sensor networks. *International Journal of Electrical and Computer Engineering*, 10(2), 1515–1523. doi:10.11591/ijece.v10i2.pp1515-1523.
- [26] Middleton, D. (2009). *An Introduction to Statistical Communication Theory*. McGraw-Hill, New York, United States. doi:10.1109/9780470544112.
- [27] Jain, A.K., & Dubes, R.C. (1988). *Algorithms for Clustering Data*. Prentice Hall, New Jersey, United States. doi:10.2307/1268876.
- [28] Forero, P. A., Cano, A., & Giannakis, G. B. (2011). Distributed Clustering Using Wireless Sensor Networks. *IEEE Journal of Selected Topics in Signal Processing*, 5(4), 707–724. doi:10.1109/jstsp.2011.2114324.
- [29] Jae-Hwan Chang, & Tassiulas, L. (n.d.). Energy conserving routing in wireless ad-hoc networks. *Proceedings IEEE INFOCOM 2000. Conference on Computer Communications. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (Cat. No.00CH37064)*. doi:10.1109/infcom.2000.832170.
- [30] Li, Q., Aslam, J., & Rus, D. (2001). Hierarchical power-aware routing in sensor networks. *Proceedings of the DIMACS workshop on pervasive networking*, 10, No. 381677.381687, 21 May, 2001, New Jersey, United States.