Wireless Personal Communications (2023) 128:537–558
https://doi.org/10.1007/s11277-022-09966-7

GA-UCR: Genetic Algorithm Based Unequal Clustering and Routing Protocol for Wireless Sensor Networks

Gunjan[1] · **Ajay K. Sharma**[2] · **Karan Verma**[2]

Accepted: 28 August 2022 / Published online: 15 September 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Wireless Sensor Networks (WSN) is an increasingly growing field, due to its enormous applications. In WSNs, energy conservation is the most important design challenge. In WSNs, unequal clustering can be classified as the best data transmission method that saves energy, where the size of the cluster changes in proportion to the cluster head's (CH's) distance from the base station (BS), so as to prevent energy holes/hot-spots from being formed. We have developed GA-UCR in this paper, a "Genetic Algorithm based Unequal Clustering and Routing Protocol for Wireless Sensor Networks". For CH election, genetic algorithm (GA) has been utilized with three fitness functions- remaining energy of CH nodes, distance between CH and BS/sink, and inter-cluster separation. For inter-cluster multi-hopping, to route the data towards BS, again GA is utilized due to the NP-Hard nature of the problem, with three fitness functions-residual/remaining energy of next hop nodes, CH to next hop node distance and number of hops. Simulation outcomes and analysis show that with reference to energy consumption, network lifetime and scalability, the proposed algorithm exceeds the existing algorithms such as Direct propagation, LEACH, TL-LEACH, GCA, EAERP and GAECH.

Keywords Wireless sensor networks · Unequal clustering protocols · Genetic Algorithm · Hot-spot problem · Energy holes

✉ Gunjan
gunjan.rehani@srmuniversity.ac.in

Ajay K. Sharma
director@nitdelhi.ac.in

Karan Verma
karanverma@nitdelhi.ac.in

1 SRM University, Delhi-NCR, Sonipat, India

2 NIT Delhi, Delhi, India

1 Introduction

Sensor development has become cheaper and simpler in recent years as a result of technological advancements [1]. Wireless sensors nodes are low-power devices that are equipped with one or more sensors, such as humidity, temperature, and pressure sensors, and the measurements acquired are then translated into signals that are transmitted to a data sink [2]. A sensing unit (sensor), a processing unit, a computing unit, and a power unit make up a sensor node/mote (GPS, mobility unit may also exist). A huge number of these devices make up the Wireless Sensor Network (WSN), which work together to communicate the collected data to the data sink/BS. By dropping them from the sky, these sensors can be placed randomly in the field or by deliberately positioning them in a particular area, if it is reachable, these can be uniformly deployed. Environmental control, forestry, health care, military, home automation, and many other WSN applications are available [3]. WSNs offers us the benefit of being able to work in environments where traditional computer devices are unable to function, such as fog, sleet, high temperatures, high humidity, etc [4]. Because of the working conditions, the sensor batteries are not replaceable often. Energy conservation is a major issue for WSNs.

A wireless sensor node's tasks include transmission of data, reception of data, and data aggregation. Of all these functions, the most energy-consuming process is data transmission, as per the radio communication model used in [5]. Proper routing algorithms should be developed for WSNs to save data transmission energy.

1.1 Motivation

Based on data transmission, WSN routing algorithms are divided into two forms, i.e. Flat-based routing (single-hop) and the Hierarchical (multi-hop) routing [6]. All nodes are handled equally in flat-based routing and have the same functionality i.e. no leader/head node concept exists. Example of flat-based protocols are Direct propagation [7], Directed diffusion [8], SPIN [9]. Some of the nodes in Hierarchical routing are leader/head nodes. As leader/head nodes, higher energy nodes are chosen and perform more energy-intensive tasks such as data propagation and aggregation of data, whereas low-energy nodes are used for data sensing and local communication. It is possible to further split hierarchical protocols into two categories: chain-based protocols and cluster-based protocols. A leader node is selected for chain-based protocols, and chain creation begins from the node farthest from the leader node. At any receiving hop, data aggregation occurs before data reaches the leader, which further communicates it to the BS. While chain-based protocols are helpful in conserving resources, they are not ideal for large-scale WSNs. One reason is the great delay due to chain construction in the transmission of data. Another explanation is that a single failure of the node will lead to the whole chain being rebuilt. An example of chain based hierarchical routing protocols is PEGASIS [10]. The network is structured into a series of clusters inside Cluster-based routing algorithms. From each cluster, a cluster head is selected. Nodes relay their information to the head of it's cluster, which further transmits it towards the BS. Example of cluster-based hierarchical routing protocols are LEACH [7], LEACH-DT [11], TEEN [12].

The cost of transmitting data is determined by the distance between transmitter and receiver. As shown in equation 1, a sensor's data transmission energy is defined as the energy needed to send a message of *m* bits over a *l* units distance.

$$E_{Tx}(m, l) = \begin{cases} m * E_{elec} + m * E_{fs} * l^2 l < do \\ m * E_{elec} + m * E_{mp} * l^4 l >= do \end{cases} \tag{1}$$

where threshold distance,

$$d_0 = \sqrt{\frac{E_{fs}}{E_{mp}}} \tag{2}$$

E_{elec} is given as the amount of electronic energy used by the hardware circuitry. $E_{fs} * l^2$ or $E_{mp} * l^4$ is the energy expanded in amplification and, it is proportional to the sender-to-receiver distance.

The free space model is used if the distance between the sender and the recipient is less than the d_0 threshold, else the multi-path model is used [7, 14, 15] and comparatively more energy is expanded than the free space model.

The energy expanded in reception for receiving a *m* bit message is given in equation 3 as :

$$E_{Rx}(m) = m * E_{elec} \tag{3}$$

Based on equation 1, we can conclude that flat routing algorithms perform better when the BS/ sink is situated close to the sensing field. When BS is away, the energy expended is proportional to the fourth power of distance, so flat-based routing algorithms are insufficient. Hierarchical algorithms work better in such situations compared to flat routing algorithms, in which nodes communicate their data to the local BS/leader/head node that further communicates it to the BS, thus reducing the number of long-distance communications. Hierarchical protocols are often used when the sensing area is wide. Hierarchical protocols for clustering can be further subdivided into algorithms for single hop and multi-hop clustering. The CH nodes in single-hop algorithms send the data directly to sink node after receiving it from the respective cluster members. Due to long-distance communications, the head/leader nodes that are away from the sink node exhaust their energy quickly in single hop algorithms. It produces energy-holes. Thus, there is uneven distribution of energy in such networks. Hierarchical multi-hop clustering algorithms are used to solve this problem [16, 17]. In such algorithms, the far away nodes do not explicitly send their information to the sink/BS. Instead, CHs use multi-hop communication, in which data is sent from one CH to the next before it reaches the BS. The network lifetime is largely extended by multi-hop clustering. But the hotpot issue continues to occur. Near the sink node, hot-spots or energy-holes are generated, because these nodes are also forwarding far-away node data.

To avoid the hot-spot problem, multiple solutions are designed by various experts [18–20]. It is one such approach to deploy more nodes near the hotspots. But this may not always be possible since the deployment is largely random due to operating environments that are difficult to reach. The solution is expensive. The base station/sink can also be mobile. In that case, these regions are going to continue to change. In the area where energy holes are formed, another option is to use larger initial energy nodes. Again the solution is expensive and needs location information. An alternate solution in such scenarios include unequal clustering protocols. All the clusters do not have the same size in unequal clustering protocols. In order to handle the load of relaying the data from far away cluster heads, the clusters near the sink are smaller, while the farther clusters are comparatively larger in size.

In several ways, inter cluster multi-hop routing can be accomplished. There are four categories of hierarchical routing techniques: chain-based, tree-based, grid-based and area-based routing techniques. One or more chains are created in chain-based routing and a leader is selected in each chain, which acts as a local BS, aggregates and sends the data received to the base station. At each receiving node, data aggregation is carried out. Data transmission starts from the chain's farthest node and is sent at each hop to the nearest node. Data transmission in tree-based routing starts from the leaf node and is delivered to the parent node until it reaches the sink. At the same time, several leaf nodes transmit and the delay is therefore reduced compared to chain-based routing. Grid-based and area-based routing exploit node location information to communicate their information. Hence, we can conclude that an unequal clustering-based routing protocol may be utilized for large scale WSNs to minimize energy consumption.

1.2 Contribution

An NP-hard issue in WSNs is the optimal cluster head election. Many meta-heuristics for the effective choice of CHs have been investigated. Given n sensor nodes, total clusters possible are $2^n - 1$. This demonstrates that for wide area WSNs, the computing complexity of using a brute force method to identify optimal cluster heads is enormous. To solve this WSN clustering problem, heuristic strategy is therefore necessary. For CH choice in WSN, genetic algorithms have been widely used. Fitness parameters are taken to select the best fit individuals, for example, remaining energy of cluster head nodes, distance from the base station. Optimal route selection for cluster head nodes is also an NP-Hard problem. For routing in WSNs, genetic algorithms were also explored. To select the best fit individuals from the population of individuals, fitness functions such as number of hops to reach the BS, residual energy of the next hop node are taken. Depending on the constraints, the resulting technique may be chain based, grid based or a combination of both. We have proposed a "Genetic Algorithm-inspired Protocol for Unequal Clustering and Routing" (GA-UCR) in this paper. If the base station is threshold or less distance apart the nodes switch to flat based routing, otherwise unequal clustering is performed. Therefore, the benefits of both flat and hierarchical clustering are taken into account. The genetic algorithm has been used for election of CHs. Due to the NP-Hard nature of the issue, genetic algorithms are again used for multi-hop routing for inter-cluster communication by the CHs. A combination of chain based and tree based routing is utilized for multi-hop inter-cluster communication. Furthermore, we do not re-cluster and re-route in every round of protocol execution to reduce energy consumption during advertising message and the overhead in topology setup, every time. Instead, it is carried out every time the energy of a CH falls below a predefined threshold.

The key contributions of the presented paper are listed as follows.

1. For the selection of cluster heads, we used a genetic algorithm-based approach.
2. For inter-cluster communication, the genetic algorithm-based unequal clustering and routing technique is employed.
3. A dynamic routing has been carried out to reduce energy consumption.

2 Related Work

A founding clustering algorithm in the area of WSNs is the "Low Energy Adaptive Clustering Hierarchy" (LEACH). This uses the probabilistic choice of cluster heads and their randomized rotation, so that each node has a fairly equal chance of being chosen as the CH. A multitude of LEACH-inspired algorithms exist. M-LEACH [17], LEACH-DT [11], TL-LEACH [16], HEED [21], TEEN [12] are a few examples to cite. All of these algorithms have equal size clusters and therefore suffer from the problem of the energy hole.

A WSN, in general, is comprised of a large amount of nodes dispersed across the network. Unequal clustering is a powerful method of organizing such a large number of nodes, equally distributing the load between the nodes of the sensor and preventing the creation of energy holes.

Unequal clustering algorithms can be categorized into three categories based on CH selection: Probabilistic, Deterministic and Preset, depending on the criterion of cluster head creation [1]. CHs are randomly or hybridly selected in Probabilistic clustering algorithms by combining random methods with parameters such as remaining energy, number of neighboring nodes, etc. On the contrary, deterministic unequal clustering algorithms use a set of standard metrics for the creation of CH. This deterministic CH creation criterion can be weighted such as in EADUC [22], UCMR [23], Fuzzy based, such as in EAUCF [24], or heuristic based, as in EBUC [25], GAEEP [26], or it may be a combination of these approaches. The location of CHs in Preset clustering algorithms is predefined before deployment, such as in UCS [27].

One of the earliest works in the field of unequal clustering was the Uniform Clustering Size (UCS) model for network organisation [27]. UCS handles both homogeneous scenarios and heterogeneous ones. In UCS, node locations are predefined and fixed. It demonstrates a 10-30 percent improvement in network life. Later, an "energy efficient clustering scheme" (EECS) [28] in WSNs was developed. For communication, EECS utilizes single hop inter-cluster communication between cluster heads, such as LEACH. CHs are elected on the basis of their remaining energy, such as no overlapping CHs are available in the same communication range. Far away clusters are smaller in size due to single hop inter-cluster communication. The CH, whose cost is minimum, is joined by non-cluster members. Cost here refers to the sum of the distance between the head of the node and the head of the cluster to sink. To better manage energy consumption, "Energy Efficient Unequal Clustering" (EEUC) [29] was developed, which is a multi-hop inter cluster routing algorithm. In EEUC, CHs are chosen based on their remaining energy, and competition radius is assigned to them, so that CHs closer to the BS are smaller in size. For inter cluster communication, nearest nodes are chosen and among them the one with maximum residual energy is chosen as the forwarding node. "Multi hop routing protocol with unequal clustering" (MRPUC) was proposed [30]. This differs from EEUC in such a way that in MRPUC there are no isolated nodes, i.e. each node belongs to the cluster. It also assumes that, unlike EEUC, the BS is situated in the sensing area's centre, and provides better performance, then when far away. [29]. A weighted function of distance to the CH and its remaining energy are taken to pick a node as the CH. To relay data, the weighted distance function of the BS and the remaining energy of the relaying CH are considered to create an inter-cluster tree rooted at the BS. For WSNs, an "energy-efficient unequal chain length clustering" (EEUCLC) was recently proposed [31]. It uses BS distance and a node's residual energy as CH election parameters. The distance to the BS is considered for cluster radius estimation. For both inter-cluster and intra-cluster communication, multi-hop clustering is used. Multiple chains are designed for inter-cluster communication and one of them is chosen at random in each round. EEUCLC has been found to outperform LEACH, enhanced LEACH and EEUC protocols. Very recently, EA-CRP : "Energy-aware Clustering and Routing Protocol" [32] was proposed. The sensing area is split into layers, where layers are shorter in size near the BS than those far away. The choice of cluster head is deterministic (weighted), with remaining energy and distance from the BS as parameters of the choice of cluster head. For routing, distance is the primary parameter, while energy is the secondary parameter.

Due to the NP-hard nature of the problem, heuristic based solutions were proposed for clustering and routing in WSNs. Provided n sensor nodes, $2^n - 1$ is the number of potential clusters. If p is the number of CHs selected and each CH has neighbors on average of q, the number of possible routing schemes is p^q. Meta-heuristic algorithms such as GAs are extremely useful for such large-scale sensor network problems [33, 34].

Another works in the domain of genetic algorithm inspired solutions for clustering is LEACH-GA [35]. In this work, using genetic algorithms, the optimum threshold likelihood (p) for cluster formation was calculated, leading to energy efficient clustering. Clustering and routing based on the Genetic Algorithm (GACR) was suggested in [36]. The clustering fitness function is chosen to reduce the average distance between nodes and their respective CHs as well as the normal cluster load variance. The load was adjusted using the nodes' remaining energy, and the algorithm works with both equal and unequal loads. The fitness function for routing was based on distance and hop count.

Wang et al. proposed GECR [37], an energy-efficient clustering and routing technique based on genetic algorithms. Clustering and routing methods are both combined into a single chromosome in GECR. The fitness function is chosen to minimise the overall energy consumed in clustering and routing, as well as the standard deviation of the cluster load based on residual energy. Table 1 shows the comparison of related clustering and routing protocols with our proposed work.

3 Assumptions and Mathematical Models used

The simulation assumptions, mathematical models used in the description of WSNs, and problem specifications are defined in this section.

3.1 Assumptions

The below assumptions have been introduced for the simulation of a WSN.

1. All of the nodes are stationary.
2. In the network, the deployment of SNs is uniform and random.
3. The batteries in the sensor nodes are non-replaceable, while the BS has an endless supply of energy.
4. All SNs have same initial energy.
5. Every node has the ability to aggregate data.
6. The nodes are not included with any GPS module.
7. The nodes will determine the distance from the signal source and change the transmitting power based on the signal intensity.
8. The MAC layer is responsible for error management.

Table 1 Comparison of proposed work with related protocols

Related work	Features	Drawbacks	How proposed work is different
LEACH [7] ("Low Energy Adaptive Clustering Hierarchy")	Probabilistic election of cluster heads with every node getting an equal fair chance to be elected as CH	Energy hole problem, single hop inter-cluster communication, CHs may be chosen from low-energy nodes	Eliminates energy holes, multi-hop inter-cluster communication
HEED [21] ("Hybrid energy efficient distributed clustering")	CH selection using remaining energy as the primary criterion and node proximity or node degree as the secondary criterion	Energy holes may be created, single-hop inter cluster communication	Eliminates energy holes, multi-hop inter cluster communication
EEUC [29] ("Energy Efficient Unequal Clustering")	Probabilistic CH election, unequal cluster radius to reduce energy hole problem, multi-hop inter cluster communication	Lower energy nodes may be chosen as CHs	Includes residual energy as CH election parameter
GCA [38] ("Genetic Clustering Algorithm")	Uses Genetic Algorithm (GA) for cluster creation with total transmission distance and number of cluster heads as fitness parameters	Energy hole problem, single hop intra-cluster communication, CHs may be chosen from low-energy nodes	Eliminates energy holes, multi-hop inter-cluster communication, includes residual energy as CH election parameter
MR-LEACH [39] ("Multi-hop Routing with LEACH")	Deterministic CH election on the basis of residual energy, equal sized clusters for load balancing, multi-hop inter-cluster communication	Energy hole problem	Eliminates energy hole problem by unequal cluster radius
LEACH-GA [40]	Uses Genetic Algorithm to calculate optimum probability of CHs, thus better energy utilization and larger lifetime compared to LEACH	Energy hole problem, single hop intra-cluster communication, lower energy nodes may be chosen as CHs	Eliminates energy holes, multi-hop inter-cluster communication, includes residual energy as CH election parameter
GABEEC [41] ("Genetic Algorithm Based Energy Efficient Clusters")	GA is used to extend the network's life. The fitness function takes three parameters into account: the round in which the first node dies, the round in which the last node dies, and the cluster distance	Lower energy nodes can be chosen as CHs, suffer from the energy hole problem and only have single hop inter-cluster connection	Eliminates energy holes, multi-hop inter-cluster communication, includes residual energy and coverage as additional CH election parameter

Table 1 (continued)

Related work	Features	Drawbacks	How proposed work is different
GAECH [32] ("Genetic Algorithm Based Energy Efficient Clustering Hierarchy")	Uses four fitness-related GA features, including total energy consumption for a single data collection round, Standard energy consumption variation between clusters, CH dispersion, and energy consumption of CHs	Single hop inter cluster communication, lower energy nodes may be chosen as CHs	Multi-hop inter-cluster communication, includes residual energy as additional CH election parameter
EEUCLC ("Energy Efficient Unequal Chain Length Clustering") [31]	Multi hop inter cluster coordination, unequal cluster radius, probabilistic CH selection based on remaining energy and distance from base station	May include long distance transmissions from CH to CH	No long distance transmissions from CH to CH, better energy management and CH election by utilizing GA

3.2 Radio Energy Model

First order radio energy model has been used in this paper. Energy expanded in data transmitting and receiving is given by equation 1 and 2. Each node is able to conduct data aggregation. Nodes give the CHs their respective data. Cluster heads compile and forward a single aggregated packet with the data collected. Therefore, less packets need to be transmitted, causing a large reduction in energy consumed. To make the comparisons easier with current algorithms, the first order radio energy model and parameters have been selected.

3.3 Data Aggregation Model

Every node has the capacity of data aggregation. Intra-cluster data is believed to be strongly co-related, so the data is aggregated into a single fixed-length packet by the cluster leader. While inter-cluster data is less related, it does not merge those packets together as in [30]. Relay members (RM) of our algorithm, before routing, also combine the data into a single packet, since this data is also spatially correlated.

3.4 Genetic Algorithm (GA) Overview

GA is a prominent meta-heuristic method to search for an optimal solution according to the principle of genetics and natural evolution [33]. The flowchart of the steps involved in genetic algorithm are given in Fig. 1. With each solution being a string of numbers/characters, the initial population is generated randomly in the first step. Each solution of the initial population is termed as a chromosome, and each unit (character/string) of the chromosome is termed as a gene. After the initial population is created, individual selection is carried out on the basis of fitness value, with a preference for higher fitness solutions. Crossover operator is then introduced, where two alternatives are randomly selected from the parent/initial population and two new solutions are generated by crossover. Mutation is applied on these solutions to produce the children population. After mutation, we have the new children. Now we have the combined population, with parent population and the generated child population after mutation. Half the population with better fitness is chosen for the next generation, while the other half is dismissed. This method continues until the condition for termination is not fulfilled.

4 GA-UCR Algorithm

The algorithm's working is split in rounds, and every round comprises of two sub-phases-the Setup Phase, and Data Transmission Phase. Setup phase comprises of Cluster Creation phase and Route construction phase. The cluster creation phase and the route construction phase is executed for the first round of operation and is repeated every time the re-clustering parameter (*rp*) is set to 1. *rp* is set to one, whenever a CH's energy becomes less than a predefined threshold CH_{et} (energy threshold of cluster head). The idea behind this is to save delay, because of the utilization of genetic algorithm in cluster formation and route construction phase. Thus using *rp*, we get a trade-off between time and efficiency. Nodes send their related data towards forwarding nodes in the data transfer phase before it reaches the BS. For any round of service, the data transmitting phase is repeated.

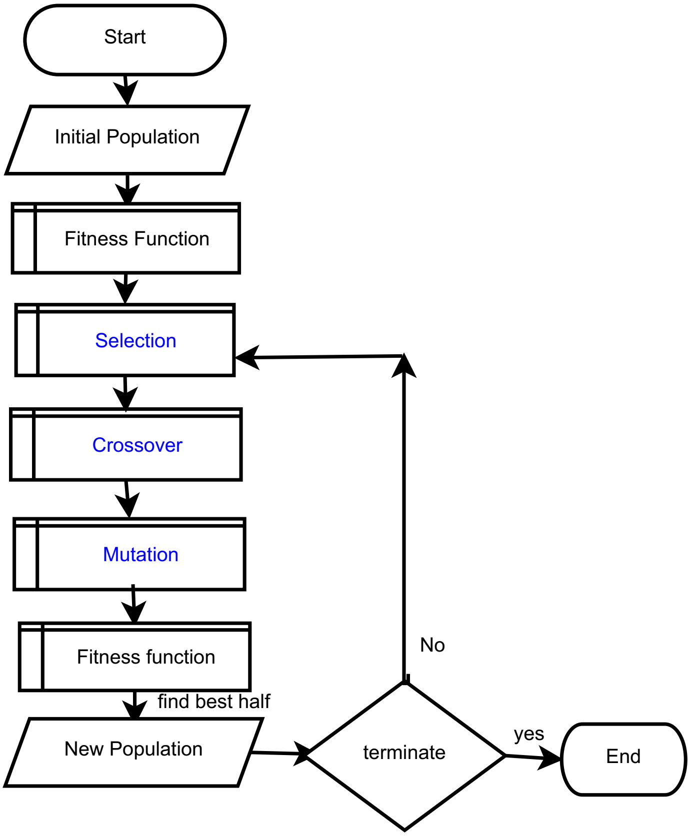

Fig. 1 Flowchart of steps involved in genetic algorithm

4.1 Setup Phase

4.1.1 GA inspired Cluster formation

In this phase, CH creation is carried out for every node, which is more than d_0 distance apart from the BS. The BS conducts CH selection, and it is not performed for each round of service. Instead it is carried out every time a CH's energy falls below a predefined threshold CH_{et} . CHs are selected on the grounds of number of alive nodes in each layer at the time of clustering. A competition radius is given after a CH is elected, based on its distance to the BS and its remaining energy. A lower competition radius is allocated to CH near the BS and with low residual energy. Once all the clusters have been formed and their corresponding cluster heads have been joined by the nodes, we might have left isolated nodes that are not in the radius of any CH competition. A join request is sent to the closest cluster member by all those nodes. As it also performs additional tasks of relaying the data, the selected cluster member updates its status to the relay member.

For cluster head creation, we use genetic algorithm. The methodologies for population representation, fitness evaluation, and for the genetic operators i.e. selection, crossover and mutation, are described below.

Table 2 Representation of a chromosome

Sensor node (SN) number	1	2	3	4	5
Sensor node type	0	1	0	1	0

Table 3 Generating initial population

SN number	1	2	3	4	5
P_1	0	0	0	1	0
P_2	0	0	1	0	0
	.	.	.	.	.
P_N	0	1	0	0	0

- *Representation of chromosomes* We use the binary technique to represent individual members of the population as strings. Each chromosome in the population is specified as a binary string of 0s and 1s, where a 0 indicates that the associated node is a member node/non-CH and a 1 indicates that the corresponding node is a CH. In addition, the overall length of a chromosome shows the total number of sensor nodes in the area of interest. In Table 2, we illustrate the depiction of a chromosome by example.
- *Generating initial population* By producing N chromosomes, we generate the initial population, with each chromosome is a string of 0s and 1s randomly generated, and its length is equal to the number of sensor nodes. An example is shown in Table 3. We use fitness function to calculate the fitness of these individuals after the initial population is produced. There are three fitness functions that are used : cluster head remaining energy, distance between CH and BS, separation between CHs, as given in the next section by Eqs. (4)–(6).
- *Fitness function for clustering Cluster head residual energy* Energy consumption by a CH includes the following:
  - Energy expended in data reception from the cluster members.
  - Energy expended in data aggregation, after receiving from cluster members.
  - Energy expended in forwarding he received data to the base station.

  Those nodes which have high value of remaining energy, should be chosen as CHs, so that they are able to operate for a large amount of time. A chromosome which is having a greater sum of residual energy for the CHs is considered a better solution within a set of solutions. The sum of residual energy of CHs in a chromosome, is denoted as :

$$res_{ch} = \sum_{i=1}^{n} energy(node(i)) \tag{4}$$

where $node(i)$ represents a CH and $energy(node(i))$ represents the residual/remaining energy of $node(i)$. *Distance between CH and BS* Once the cluster heads have received data from the neighbors/cluster members, they aggregate this data and forward it to the BS. The power consumption in transferring this data is related to the distance between the sink and the cluster head, and therefore it is critical to keep this distance as short as possible. In each chromosome, the distance of each CH to BS is calculated and then summed up, and is denoted as :

$$dss_{ch} = \sum_{i=1}^{n} dist(node(i)) \tag{5}$$

where $node(i)$ represents a CH and $dist(node(i))$ represents the distance between $node(i)$ and the BS. *Cluster head separation* The distance between any two CHs should be as large as possible, so that we do not have overlapping clusters and also that the coverage is maximum. It is calculated as :

$$sep = max(dist(node(i), node(j))) \tag{6}$$

where $node(i)$ and $node(j)$ are any two CHs. Of the three fitness functions discussed above, it is necessary to maximize the residual energy of cluster head nodes given by equation 4 and the separation between cluster heads given by equation 6, while it is necessary to minimize the distance of cluster heads to BS, shown by equation 5. Once the fitness of each chromosome is calculated, genetic algorithm operators i.e. selection, crossover and mutation are applied on this initial population (P_0)), to produce the off-spring population (Q_0).

– *Genetic operators Selection operator* This operator selects those chromosomes which have higher fitness values, to be included in the next generation. Because of its wide-spread usage and acceptability, we used the Binary Tournament Selection operator [43] in our work. After the completion of Selection process, next the crossover operator is applied as explained below. *Crossover operator* Through sharing information between existing solutions, the crossover operator produces new solutions from existing solutions. Two chromosomes are randomly selected from the population, and every part of one string is swapped with the other string. The crossover point is therefore chosen at random. A crossover probability is often added to allow an independent solution string the freedom to decide if the solution will go to crossover or not [32]. In our research work, we employed a single-point crossover with a crossover rate of 0.5. Based on the defined problem space, this crossover rate will lead to a faster convergence to the solution. The crossover operator is illustrated in the following example. Consider the case of a WSN with 5 sensor nodes. In this case, 2 is used as the crossover point. The remaining string is split between the two chromosomes after this point. Figure 2 demonstrates this. The next step is mutation. *Mutation operator* The mutation operator is available to preserve the population's diversity. It gives the population new characteristics. The mutation operator chooses a node at random from the string and changes its form, i.e. if the value is 0, it is modified to 1, and vice versa (non cluster head changes to a cluster head and vice versa). A generally low value of mutation probability is chosen for steady convergence. The mutation operator used in this paper is flip bit operator, and the mutation rate is kept as 0.05. We show the mutation operator with the below

Chromosome 1	1	0	0	1	0
Chromosome 2	0	1	0	0	0
Offspring 1	1	0	0	0	0
Offspring 2	0	1	0	1	0

Crossover Point

Fig. 2 Single point crossover

Table 4 flip bit mutation

Offspring1	0	1	0	1	0
Offspring2	0	0	1	0	1
Mutated 1	1	1	0	1	0
Mutated 2	0	0	1	0	0

Table 5 Representation of chromosome

Cluster head number	1	3	5
Next-hop node	3	5	*bs*

example, considering a WSN of 5 nodes in Table 4. In this example, randomly bit 3 is mutated in Offspring 1 to produce mutated 1, and bit 5 is mutated in Offspring 2 to produce mutated 2.

4.1.2 GA Inspired Route Formation

The routing process in GA-UCR is guided, to reduce the search space. If base station is in d_0 range of cluster head, then *bs* is added to the set next-hop-node, else each CH keeps a set of all CHs in d_0 range, who are nearer to the sink than itself, i.e. $d_{j,sink} < d_{i,sink}$. If no such CH is found (there is no forwarding CH), then we put all cluster members or direct members nearer to sink in the next-hop-node set to act as relays. For route construction, we use genetic algorithm. The procedure for representation of population, evaluation of fitness, applying the genetic operators i.e. selection, crossover and mutation, are described in the subsections.

– *Representation of chromosomes* We represent the individual members of the population as a string, using the binary method. Each chromosome in the population is mentioned as a binary string consisting of 0s and 1s, with a 0 indicating a member node/non-CH and a 1 indicating a CH. A chromosome's total length also reflects the total number of sensor nodes in the region. The corresponding depiction of the chromosome can be seen in Table 5, where the nodes numbered 1, 3, and 5 are CH nodes, while the rest are non-CH nodes. Also, the gene value 3 indicated that CH with id 3 is the possible next hop node of CH with id 1.
– *Generating initial population* The initial population is made up of *N* chromosomes, with each chromosome's length equal to the number of cluster heads. Perceive a wireless sensor network (WSN) made up of 5 sensor nodes, with node 1,3,5 as CHs as shown in table 6. Once we generate the initial population, the fitness value of each individual is calculated. The three fitness functions used for routing are remaining energy of cluster heads, distance between CH and BS, delay/maximum number of hops, as given in the next section.
– *Fitness function for routing Residual energy of next hop nodes* For reliability, we should chose high energy nodes as next hop nodes. Thus, for a given set of solutions, the chro-

Table 6 Generating initial population

Cluster head number	1	3	5
P_1	3	5	*bs*
P_2	*bs*	*bs*	1
	.	.	.
P_N	5	*bs*	1

Table 7 Chromosome representing cluster heads next hop

Cluster head number	1	3	5	7	10
Next-hop node	3	5	*bs*	10	*bs*

mosomes in which sum of remaining energy of next hop nodes is higher, will be a better solution. This is denoted as :

$$res_{nh} = \sum_{i=1}^{n} energy(node(i)) \tag{7}$$

if $node(i)$ symbolizes a next hop node and $energy(node(i))$ symbolizes the remaining energy of $node(i)$ *Distance between CH and next hop node* The CHs forward the data gathered to the next hop nodes. This transmission energy is proportional to the distance between these two nodes, hence this distance should be minimum. We calculate the sum of distances of all cluster heads to their next hop node as :

$$d_{nh} = \sum_{i=1}^{n} dist(node(i)) \tag{8}$$

if $node(i)$ symbolizes a cluster head and $dist(node(i))$ symbolizes the distance between $node(i)$ and its next hop node. *Number of hops* It represents the length of the largest chain to reach the base station. This value should be minimum to minimize delay in routing. The process for calculating the hop count is described as follows: Each cluster head calculates its hop count. The hop count is taken as the the largest value of hop count among these.

Algorithm 1 maxhopcount

procedure MAXHOPCOUNT($setofch[], nexthopch[]$)
  Input:$setofch[], nexthopch[]$
  Output:$maxhopcount$
  $maxhopcount \leftarrow 0$
  for each cluster head i **do**
    $length(i) \leftarrow 1$
    while nexthop(i)!='bs' **do**
      $length(i) \leftarrow length(i) + 1$
  $maxhopcount \leftarrow max(length)$

Consider a WSN comprising of 10 sensor nodes. Consider the set of cluster heads and their corresponding next hop nodes to be that given in Table 7. The two chains that exist in this topology are- $1->3->5->bs$ and $7->10->bs$. Each node

Table 8 Parameters of radio energy model used

Parameter	Value
Sensor node's initial energy (E_{init})	0.02 J
Free space energy consumption (E_{fs})	10 pJ/bit/m^2
Multi-path energy consumption (E_{mp})	0.0013 pJ/bit/m^4
Threshold distance (d_0)	87.7 m
Length of data packet	2000 bits
Energy consumption for data aggregation	5 nJ/bit/signal
Energy consumption of transmitter circuitry	50 nJ/bit
Energy consumption of receiver circuitry	50 nJ/bit

calculates its chain length as given in algorithm 1 above. length(1) = 3 length(3) = 2 length(5) = 1 length(7) = 2 length(10) = 1 Thus *maxhopcount* = 3, which is the length of the largest chain among the two chains in the example. Among the above three fitness functions for routing, residual energy of next hop nodes [given by Eq. (7)] needs to be maximized, while the rest two, i.e. distance of cluster heads to next hop [given by Eq. (8)] and number of hops (given by Algorithm 1) needs to be minimized. Genetic operators are applied to the initial population (P_0) after evaluation of fitness values, to create the offspring population (Q_0) as mentioned in the preceding section. To assure that the mutation operator always produces a valid chromosome, a randomly selected gene value is replaced with a valid gene value from the set of possible next hop nodes. We repeat this process up-to the required number of generations.

After the route for cluster heads is created, the next step is to construct the route for isolated/orphan nodes, i.e nodes which are not in communication radius of any cluster head. Such isolated nodes, if any, find a nearest cluster member, whose energy is above a specified threshold as the relay member to forward its data to cluster head, along with its own data. This is done to ensure that there are no coverage holes in the network.

5 Simulation Model and Parameters

5.1 Parameters for Simulation

We assume the allocation of sensor nodes as Uniform in the region of interest. The assumed region of interest is two-dimensional (200×200 unit^2) as shown in Fig. 3, for all our experiments. First order radio energy model parameters are assumed for simulation [5] as shown in Table 8.

For the simulation of the proposed algorithm: GA-UCR, the parameters taken are as given in Table 9. The parameters are chosen after carrying out multiple experiments and finding the best possible values for the scenario under consideration.

In terms of network lifetime: first node dead (FND), half node dead (HND), last node dead (LND), count of alive nodes in each round, and cumulative energy consumption by the nodes during each round, the proposed GA-UCR algorithm is compared to existing

LEACH, TL-LEACH, and GAECH algorithms. All the algorithms are simulated in MATLAB R2017b.

6 Results and Discussion

6.1 Comparison of Network lifetime

In Fig. 4, the number of alive nodes in each round is plotted against the number of rounds. The value of parameter p taken in case of Leach, TL-LEACH, GAECH, GA-UCR (proposed) during the simulation is $p = 0.05$. Compared to GAECH, TL-LEACH and traditional LEACH protocols, GA-UCR improves the network lifetime.

6.1.1 Network Lifetime Percentage Gain Comparison

Comparison of first node dead (FND), half node dead (HND), full/complete node dead (CND), in case of Direct Propagation, Leach, TL-LEACH, EEUCLC and LEBUCR (proposed) is shown in Fig. 5 for Scenario 1. It is noted that in terms of all three parameters, LEBUCR performs better. To evaluate the percentage gain of GA-UCR (proposed algorithm) against the related algorithms, we consider the data obtained from the Fig. 4.

Table 10 shows the percentage gain in network lifetime in relation to all the three parameters i.e FND, HND, CND of GA-UCR with the compared algorithms.

6.2 Relative Energy Consumption Comparison

The energy consumption against the number of rounds, for the algorithms LEACH, TL-LEACH, GAECH and GA-UCR is shown in the Fig. 6. In all five instances, the energy level of each node is 0.5 J initially. The sensing field's number of nodes is calculated as 150. As a result, $[0.5 \times 150 = 75\,\text{J}]$ is the total energy available. When comparing GAECH, TL-LEACH, and conventional LEACH protocols, it has been discovered that GA-UCR nodes consume less energy.

Table 9 GA parameters

Parameter	Value
Initial population size	50
Maximum generations	100
Initial population	Random generation (for clustering), guided (for routing)
Operator for selection	Binary tournament
Operator for crossover	Single point
Rate of crossover	0.6
Operator for mutation	Flip bit
Rate of mutation	0.05

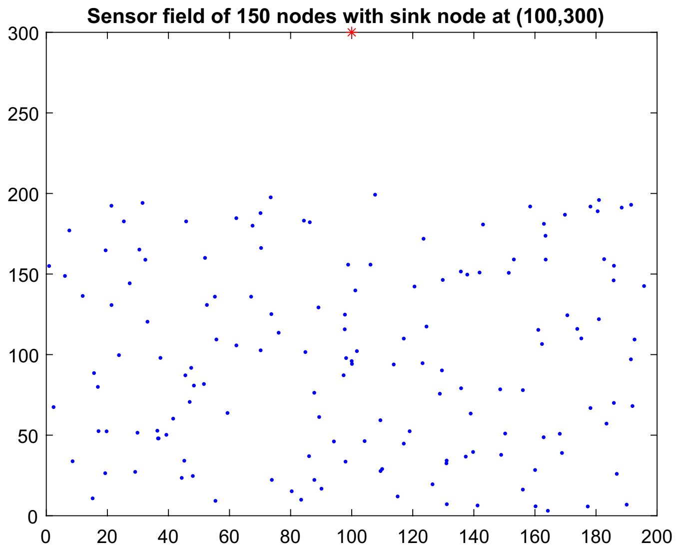

Fig. 3 Sensing region with 150 sensor nodes

6.3 Relative Comparison of Network Lifetime

The comparison of network lifetime of GA-UCR is done with a few related algorithms in Table 11. The proposed protocol GA-UCR was implemented for the similar scenarios

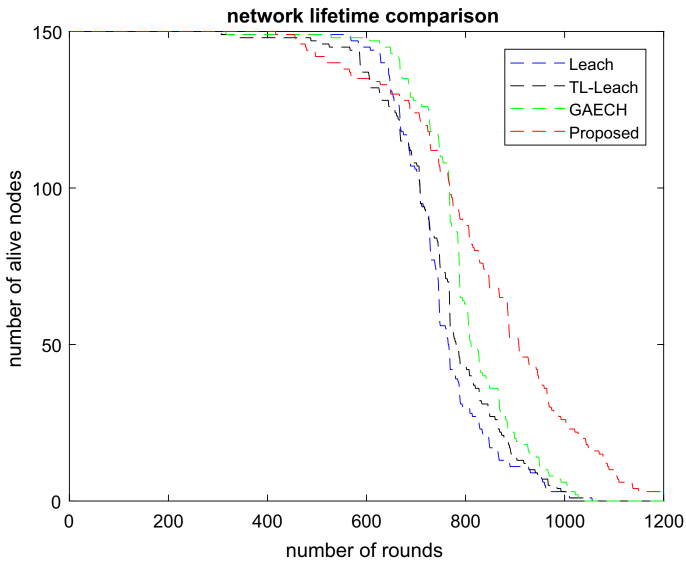

Fig. 4 Network lifetime comparison

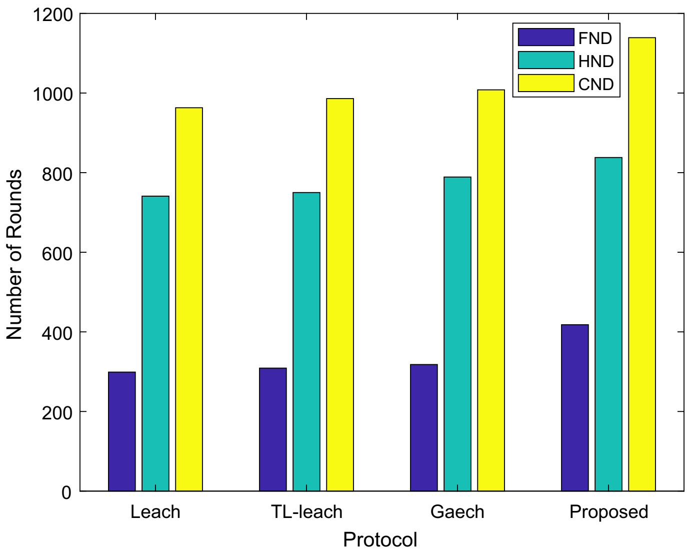

Fig. 5 Comparison of FND, HND, CND Statistics

Table 10 Comparison of percentage gain of GA-UCR

Protocol	FND	HND	CND
LEACH	39.80	13.09	–
TL-LEACH	35.28	11.73	15.52
GAECH	31.45	6.21	13.00

as in the compared protocols and network lifetime metrics were taken. It has been found that GA-UCR outperforms all these algorithms.

7 Conclusion and Future Scope

We have established a GA-based joint clustering and routing protocol to manage energy more efficiently in a WSN, in this paper. For cluster head election, genetic algorithm has been utilized with three fitness functions- residual/remaining energy of CH nodes, CH to sink distance, and inter CH separation. For inter cluster multi-hop routing among CHs, again genetic algorithm is utilized due to the NP-Hard nature of the problem with three fitness functions-residual energy of next hop nodes, distance of cluster head to next hop nodes and number of hops. Our proposed algorithm shows 30–35% improvement in FND (First Node Dead) statistics as compared to GAECH, and hence greater stability.

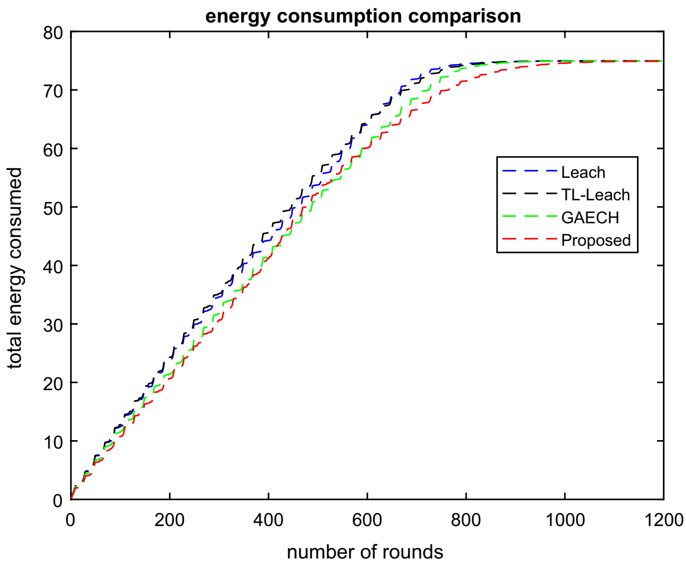

Fig. 6 Energy consumption comparison

Table 11 Relative comparison of network lifetime

Protocol	Metrics for network lifetime			Proposed		
	FND	HND	CND	FND	HND	CND
LEACH [7]	299	741	963	418	838	1139
TL-LEACH [16]	309	750	986	418	838	1139
GCA [38]	–	2018	2081	2893	2430	2655
EAERP [13]	–	2129	2184	2893	2430	2655
GAECH [32]	318	789	1008	418	838	1139

The proposed algorithm can be improved further by taking mobility of nodes and sink into account.

Funding Not applicable.

Availability of data and materials Not applicable.

Code availability Not applicable.

Declarations

Conflict of interest Not applicable.

References

1. Arampatzis, T., Lygeros, J., & Manesis, S. (2005). A survey of applications of wireless sensors and wireless sensor networks. In *Intelligent control, 2005. Proceedings of the 2005 IEEE international symposium on, mediterrean conference on control and automation*. IEEE.
2. Al-Karaki, J. N., & Kamal, A. E. (2004). Routing techniques in wireless sensor networks: A survey. *IEEE Wireless Communications, 11*(6), 6–28.
3. Puccinelli, D., & Haenggi, M. (2005). Wireless sensor networks: Applications and challenges of ubiquitous sensing. *IEEE Circuits and Systems Magazine, 5*(3), 19–31.
4. Yick, J., Mukherjee, B., & Ghosal, D. (2008). Wireless sensor network survey. *Computer Networks, 52*(12), 2292–2330.
5. Heinzelman, W. B., Chandrakasan, A. P., & Balakrishnan, H. (2002). An application-specific protocol architecture for wireless microsensor networks. *IEEE Transactions on Wireless Communications, 1*(4), 660–670.
6. Younis, M., & Akkaya, K. (2008). Strategies and techniques for node placement in wireless sensor networks: A survey. *Ad Hoc Networks, 6*(4), 621–655.
7. Heinzelman, W. R., Chandrakasan, A., & Balakrishnan, H. (2000). Energy-efficient communication protocol for wireless microsensor networks. In *System sciences, 2000. Proceedings of the 33rd annual Hawaii international conference on*. IEEE.
8. Intanagonwiwat, C., Govindan, R., & Estrin, D. (2000). Directed diffusion: A scalable and robust communication paradigm for sensor networks. In *Proceedings of the 6th annual international conference on Mobile computing and networking*. ACM.
9. Heinzelman, W. R., Kulik, J., & Balakrishnan, H. Adaptive protocols for information dissemination in wireless sensor networks. In *Proceedings of the 5th annual ACM/IEEE international conference on Mobile computing and networking*. ACM.
10. Lindsey, S., & Raghavendra, C. S. (2002). PEGASIS: Power-efficient gathering in sensor information systems. In *Proceedings, IEEE aerospace conference* (Vol. 3). IEEE.
11. Kang, S. H., & Nguyen, T. (2012). Distance based thresholds for cluster head selection in wireless sensor networks. *IEEE Communications Letters, 16*(9), 1396–1399.
12. Manjeshwar, A., & Agrawal, D. P. (2001). TEEN: A routing protocol for enhanced efficiency in wireless sensor networks. In *Null*. IEEE.
13. Khalil, E. A., & Bara'a, A. A. (2011). Energy-aware evolutionary routing protocol for dynamic clustering of wireless sensor networks. *Swarm and Evolutionary Computation, 1*(4), 195–203.
14. Tripathi, R. K., Singh, Y. N., & Verma, N. K. (2012). N-leach, a balanced cost cluster-heads selection algorithm for wireless sensor network. In *Communications (NCC), 2012 national conference on*. IEEE.
15. Heinzelman, W. B., Chandrakasan, A. P., & Balakrishnan, H. (2002). An application-specific protocol architecture for wireless microsensor networks. *IEEE Transactions on Wireless Communications, 1*(4), 660–670.
16. Loscri, V., Morabito, G., & Marano, S. (2005). A two-levels hierarchy for low-energy adaptive clustering hierarchy (TL-LEACH). In *Vehicular technology conference, 2005. VTC-2005-Fall. 2005 IEEE 62nd* (Vol. 3). IEEE.
17. Neto, J. H. B., Rego, A., Cardoso, A. R., & Celestino, J. (2014). MH-LEACH: A distributed algorithm for multi-hop communication in wireless sensor networks. *ICN, 2014*, 55–61.
18. Perillo, M., Cheng, Z., & Heinzelman, W. (2005). Strategies for mitigating the sensor network hot spot problem. In *Proceedings of MobiQuitous*.
19. Perillo, M., Cheng, Z., Heinzelman, W. (2004). On the problem of unbalanced load distribution in wireless sensor networks. In *IEEE Global Telecommunications Conference Workshops, 2004. GlobeCom Workshops 2004*. IEEE.
20. Jaichandran, R., & Irudhayaraj, A. A. (2010). Effective strategies and optimal solutions for hot spot problem in wireless sensor networks (WSN). In *10th international conference on information science, signal processing and their applications (ISSPA 2010)*. IEEE.
21. Younis, O., & Fahmy, S. (2004). HEED: A hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks. *IEEE Transactions on Mobile Computing, 3*(4), 366–379.
22. Yu, J., Qi, Y., Wang, G., Guo, Q., & Gu, X. (2011). An energy-aware distributed unequal clustering protocol for wireless sensor networks. *International Journal of Distributed Sensor Networks, 7*(1), 202145.
23. Gupta, V., & Pandey, R. (2016). An improved energy aware distributed unequal clustering protocol for heterogeneous wireless sensor networks. *Engineering Science and Technology, an International Journal, 19*(2), 1050–1058.

24. Bagci, H., & Yazici, A. (2013). An energy aware fuzzy approach to unequal clustering in wireless sensor networks. *Applied Soft Computing, 13*(4), 1741–1749.
25. Jiang, C.-J., Shi, W.-R., & Tang, X.-L. (2010). Energy-balanced unequal clustering protocol for wireless sensor networks. *The Journal of China Universities of Posts and Telecommunications, 17*(4), 94–99.
26. Abo-Zahhad, M., Ahmed, S. M., Sabor, N., & Sasaki, S. (2014). A new energy-efficient adaptive clustering protocol based on genetic algorithm for improving the lifetime and the stable period of wireless sensor networks. *International Journal of Energy, Information and Communications, 5*(3), 47–72.
27. Soro, S., & Heinzelman, W. B. (2005). Prolonging the lifetime of wireless sensor networks via unequal clustering. In *19th IEEE international parallel and distributed processing symposium*. IEEE.
28. Ye, M., Li, C., Chen, G., & Wu, J. (2005). EECS: An energy efficient clustering scheme in wireless sensor networks. In *PCCC 2005. 24th IEEE international performance, computing, and communications conference, 2005*. IEEE.
29. Li, C., Ye, M., Chen, G., & Wu, J. (2005). An energy-efficient unequal clustering mechanism for wireless sensor networks. In *IEEE international conference on mobile adhoc and sensor systems conference*. IEEE.
30. Gong, B., Li, L., Wang, S., & Zhou, X. (2008). Multihop routing protocol with unequal clustering for wireless sensor networks. In *2008 ISECS international colloquium on computing, communication, control, and management* (Vol. 2). IEEE.
31. Baniata, M., & Hong, J. (2017). Energy-efficient unequal chain length clustering for wireless sensor networks in smart cities. *Wireless Communications and Mobile Computing*, 2017.
32. Baranidharan, B., & Santhi, B. (2015). GAECH: genetic algorithm based energy efficient clustering hierarchy in wireless sensor networks. *Journal of Sensors*, 2015.
33. Gen, M., & Lin, L. (2007). Genetic algorithms. *Wiley Encyclopedia of Computer Science and Engineering*, 1-15.
34. Gunjan. (2022). A Review on Multi-objective Optimization in Wireless Sensor Networks Using Nature Inspired Meta-heuristic Algorithms. *NEURAL PROCESSING LETTERS*.
35. Liu, J.-L., & Ravishankar, C. V. (2011). LEACH-GA: Genetic algorithm-based energy-efficient adaptive clustering protocol for wireless sensor networks. *International Journal of Machine Learning and Computing, 1*(1), 79.
36. Gupta, S. K., & Jana, P. K. (2015). Energy efficient clustering and routing algorithms for wireless sensor networks: GA based approach. *Wireless Personal Communications, 83*(3), 2403–2423.
37. Wang, T., Zhang, G., Yang, X., & Vajdi, A. (2018). Genetic algorithm for energy-efficient clustering and routing in wireless sensor networks. *Journal of Systems and Software, 146*, 196–214.
38. Mudundi, S., & Ali, H. H. (2007). A new robust genetic algorithm for dynamic cluster formation in wireless sensor networks. In *Proceedings of Wireless and Optical Communications, Montreal*.
39. Farooq, M. O., Dogar, A. B., & Shah, G. A. (2010). MR-LEACH: Multi-hop routing with low energy adaptive clustering hierarchy. In *2010 4th international conference on sensor technologies and applications*. IEEE.
40. Liu, J.-L., & Ravishankar, C. V. (2011). LEACH-GA: Genetic algorithm-based energy-efficient adaptive clustering protocol for wireless sensor networks. *International Journal of Machine Learning and Computing, 1*(1), 79.
41. Bayraklı, S., & Erdogan, S. Z. (2012). Genetic algorithm based energy efficient clusters (GABEEC) in wireless sensor networks. *Procedia Computer Science, 10,* 247–254.
42. Gajjar, S., Sarkar, M., & Dasgupta, K. (2016). FAMACROW: Fuzzy and ant colony optimization based combined mac, routing, and unequal clustering cross-layer protocol for wireless sensor networks. *Applied Soft Computing, 43*, 235–247.
43. Whitley, D. (1994). A genetic algorithm tutorial. *Statistics and Computing, 4*(2), 65–85.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Gunjan is working as Assistant Professor in Computer Science and EngineeringDepartment at SRM University, Delhi-NCR Sonipat. Prior to this, she has served asAssistant Professor at Bennett University, Greater Noida. She has completed herPh.D. in Computer Science and Engineering Department from the National Instituteof Technology, Delhi, India. She received her Bachelor's in Technology in 2012 fromMaharishi Dayanand University, Haryana, India; Master's in technology fromNational Institute of Technology, Jalandhar, Punjab, India in 2014. Her researchinterest lies in the domain of energy-efficient techniques in wireless sensornetworks, evolutionary algorithms.

Prof. Ajay K. Sharma is working as Director, NIT Delhi. Prior to this, he has served as Vice-Chancellor IKG Punjab Technical University, Jalandhar, Punjab, India. He received his B.E. in Electronics and Electrical Communication Engineering from Punjab University Chandigarh, India in 1986, M.S. in Electronics and Control from Birla Institute of Technology (BITS), Pilani in the year 1994 and Ph.D. in Electronics Communication and Computer Engineering in the year 1999. His Ph.D. thesis was on "Studies on Broadband Optical Communication Systems and Networks". After serving various organizations from 1986 to 1995, he has joined the National Institute of Technology (Erstwhile Regional Engineering College) Jalandhar as Assistant Professor in the Department of Electronics and Communication Engineering in the year 1996. Since November 2001, he has worked as Professor in the ECE Department and after that, he has worked as Professor in Computer Science and Engineering from 2007 to 2013 in the same institute. His major areas of interest are broadband optical wireless communication systems and networks, dispersion compensation, fiber nonlinearities, optical soliton transmission, WDM systems and networks, radio-over-fiber (RoF) and wireless sensor networks and computer communication.

Dr. Karan Verma is working as Assistant Professor in Computer Science and Engineering Department, NIT Delhi. He received his B.E and M. Tech degrees in Information Technology and Computer Science & Engineering, respectively, from the University of Rajasthan in 2008, and the Indian Institute of Technology, Roorkee in 2010. He received his Ph.D. degree from the Department of Computer and Information Science at the Universiti Teknologi PETRONAS, Malaysia, and Post DoC (PDF) in Computer Science & Engineering from the University of Californian, Davis, CA, USA. His research interests include information security, wireless sensor networks, and VANET.