A Genetic Algorithm Inspired Optimized Cluster Head Selection Method in Wireless Sensor Networks

Biswa Mohan Sahoo[a,c], Hari Mohan Pandey[b], Tarachand Amgoth[c]

[a]School of Computing and IT, Manipal University Jaipur, India
[b]Data Science and Artificial Intelligence Department, Bournemouth University, UK
[c]Indian Institute of Technology (Indian School of Mines), Dhanbad
biswamohans@gmail.com, hpandey@bournemouth.ac.uk, tarachand@iitism.ac.in

Abstract: In this paper, an optimized cluster head (CH) selection method based on genetic algorithm (NCOGA) is proposed which uses the adaptive crossover and binary tournament selection methods to prolong the lifetime of a heterogeneous wireless sensor network (WSN). The novelty of the proposed algorithms is the integration of multiple parameters for the CH selection in a heterogeneous WSN. NCOGA formulates fitness parameters by integrating multiple parameters like the residual energy, initial energy, distance to the sink, number of neighbors surrounded by a node, load balancing factor, and communicating mode decider (CMD). The parameters for load balancing and CMD are utilized to discover out the best candidate to be selected as a relay CH and for deciding the mode of communication (single or multi-hop) of CH. Further, these parameters are useful in avoiding hot-spot problem in the network. The working of the NCOGA starts based on the criteria "consider only those nodes which have energy higher than the pre-defined threshold energy". This criterion of nodes selection makes the NCOGA more efficient and quickly convergent. Extensive computer simulations are conducted to determine the effectiveness of the NCOGA. Simulation results reveal that the proposed NCOGA outperforms the state-of-the-art optimization algorithms based on GA in terms of several performance metrics, specifically, stability period, residual energy, network lifetime, and throughput.

Keywords: GA-based CH selection; Clustering; Load balancing; Communicating mode decider; Wireless sensor networks.

Abbreviations

NCOGA	*Novel Optimized CH selection method based on Genetic Algorithm*
CMD	*Communication Mode Decider*
GA	*Genetic Algorithm*

N_{NORM}	*Represent the network Normal Node*
N_{ADVN}	*Represent the network Advance Node*
N_{SUP}	*Represent the network Super Node*
N	*Total number of nodes*
Ʊ	*Represent the super node energy fraction*
Θ	*Represent the advance node energy fraction*
ϖ	*As compared to a normal node, advanced nodes have a higher energy fraction*
ϕ	*As compared to normal nodes, super nodes have a higher energy fraction*
E_O	*Represent the node initial energy*
E_T	*Represent the network total energy*
FP_S	*Represent the Fitness Parameters*
$E_{R(i)}$	*Represent the i^{th} node residual energy*
$E(i)$	*Represent the i^{th} node initial energy*
V_i	*Represent the i^{th} population vector*
CH_P	*The suggested probability of cluster heads*
$D_{(N(i)-S)}$	*Represent the i^{th} node Euclidean distance from the sink*
$D_{(AVG(N(i)-S))}$	*Represent the average distance from i^{th} node to sink*
N_{CL}	*Cluster size (number of nodes)*
$D_{((N(i)-N(j)))}$	*Represent the distance from i^{th} node to j^{th} node*
RCH	*Represent the Relay Cluster Head*
$D_{((CH(i)-CH(j)))}$	*Distance between the CHs and $D_{(N(i)-S)}$*
N_C	*Number of clusters in the network*
$L.B.$	*Load Balancing*
$\varphi, \delta, \gamma, \alpha, \sigma, \beta$	*Represent the Weight coefficients*
P_m	*Mutation rate*
P_c	*Crossover rate*
i_{rank}	*Non-domination rank*
$i_{cr_{dist}}$	*Crowding distance*
$C(V)$	*Represent the current sensed value*
$S(V)$	*Represent the sensed value*
$H(T)$	*Represent the Hard Threshold*
$S(T)$	*Represent the Soft Threshold*
$E_{tx}(z, d)$	*The energy used to transmit z-bit data over distance d*
E_{elc}	*The energy used to activate the transmitter and receiver circuitry*
E_{efs}	*Represent the free space energy model*
E_{amp}	*Represent the energy used for multipath energy model*
d_o	*Represent the Threshold Distance*
E_{rx}	*The energy used to receive z-bit data*
E_{da}	*Represent the energy used in the data aggregation of 1-bit data*
E_{dx}	*Consumption of energy during data aggregation*

1. Introduction

Wireless sensor network (WSN) is a self-organized network that comprises of spatially distributed wireless sensor nodes that sense the target area and transmit the sensed information to the sink [1]. Due to technological advancements made in the Micro-Electro-Mechanical-System (MEMS), it has become possible to develop various low-cost sensor nodes with embedded processing, automated sensing, and non-rechargeable batteries. These sensor nodes are extensively used in the different sectors of applications namely military operations, agriculture, disaster management, health care, habitat monitoring, and industrial purposes [2]. When sensor nodes are deployed in the harsh environment, it becomes quixotic to replace the batteries of sensor nodes, which puts a constraint on the network running of WSN [3] [4]. Consequently, it is of prime importance to conserve the battery resource of sensor nodes by building up the energy efficient communication among nodes such that the network running could be elongated to the maximum possible period.

The classical routing protocols with homogeneous network are unable to save the nodes' energy. For the homogeneous network, all the nodes have similar capacity, computing, and coverage resources, and they have proven to be inefficient. It is observed, however, that the energy of the nodes declines with the incremental progression of data transmission. On the contrary, for heterogeneous network the nodes are divided into various genres based on their original energy stock. The designed algorithm works in a way that the algorithm designed for different categories of sensor nodes works according to the energy profile of these nodes. Depending on the energy expenditure of the nodes they turn from one genre to another. The truth is, however, that the guidelines for all node genres remain the same during the operation of the network. When a node labeled as an advance node reaches the energy level of super nodes as well as super node reaches the energy level of a normal node, it would also have the same requirements for CH selection as super nodes. The main function of the high-energy nodes is to act as a repository for the other nodes on the network. Expected battery resources are the most generally used because of the important position they perform in energy heterogeneity [10].

One of the major objectives of the researchers exploring WSN is the minimization of energy consumption of the sensor nodes [5]. With this perspective, a plethora of research has been carried out by proposing different energy efficient routing strategies that aim to conserve the energy of the sensor nodes so that period of network run can be enhanced [6]. One of the essential attempts in the energy conservation has been the concept of 'clustering' in WSN. In clustering mechanism, the nodes are grouped together to form a cluster and one node in the cluster is appointed as 'Cluster Head (CH)' and rest of the nodes are named as 'cluster members' [7]. The CH gathers information from the cluster members, removes the correlated data (i.e., process is termed as aggregation), and forwards the useful information to the sink in a single hop or via multi-hop transmission. If the multi-hop communication is involved for the data transmission among the different CHs, the intermediate CHs get burdened heavily due to the relaying of data to the sink. Eventually, these intermediate CHs and most likely the CHs closest to the sink will fully drained off their energies. As a result, the data transmission from the victim cluster (a cluster suffering from hotspot) is halted which affects the network reliability and its performance heavily. This problem is termed as the hot-spot problem. A high magnitude of research work has been reported so far that attempts to alleviate the hot-spot problem [8] [9].

It is a noteworthy fact that selection of CH is a non-Deterministic polynomial (NP)-hard optimization problem due to the selection of 'q' number of optimal CHs among 't' i.e., total number of sensor nodes in the network generates $\binom{q}{t}$ possibilities. Different computational optimization techniques have been applied to acquire the optimal solution of such NP problems [10]. Many optimization techniques have been proposed to achieve enhanced network performance by optimizing different attributes namely routing, CH selection, sink placement, etc. [11] [12] [13]. Genetic Algorithm is an effective approach for solving NP-hard problems as it has the characteristics of ease of implementation, solution with better quality, ability to retain the best solution (i.e., also termed as elitism), and low convergence time [14]. The CH selection is done by optimizing the different attributes of nodes taken in the initialization. The genes are represented in the sequence of bit stream of 1's and 0's, where '1' represents the selected CH and '0' is used for the cluster member nodes. It is the fitness function that decides for a chromosome representing a node whether it will be selected as CH or not. The fitness function is integrated by considering different parameters and the chromosome with values closest to the optimal values is selected as CH, i.e., selecting the optimized node as a CH and afterwards it performs clustering operations. In this paper, a multi-objective algorithm is used that tends to optimize the different parameters. It is believed that multi-objective optimization (MOO) algorithms optimize the conflicting goals in WSNs simultaneously. Most of these algorithms optimized the consumption of energy while considering other conflicting goals together. Weights are assigned to the input parameters, which transformed the multi-dimensional optimization problem to single dimension.

Motivated from aforementioned discussion, we suggest a new method that deliberates the NCOGA to select CH in a WSN so as the efficient network performance can be acquired. The main contributions are highlighted as follows:

a) We propose a novel CH selection optimized based on GA (NCOGA) for the heterogeneous WSN. NCOGA is complied with fitness function, which considers six parameters, namely, residual energy, initial energy, distance to the sink, a node having neighbors' nodes, load balancing factor and communicating mode decider (CMD). It is to be noted that it is the first ever attempt with these parameters for the CH selection in heterogeneous WSN.
b) The parameters of load balancing and CMD help in discovering out the best candidate to be selected as relay CH and deciding for the mode of communication of CH i.e., whether it is single or multi hop, respectively. With the inclusion of these parameters, hot-spot problem is avoided in the network.
c) The proposed improved GA considers only those nodes for the initialization process which have energy more than the pre-defined threshold energy. This criterion renders the early convergence to the proposed optimization strategy. Unlike many state-of-art protocols, the proposed NCOGA considers the elitism process that keeps a record of the best solutions i.e., already selected CH so that the frequent selection of a node as CH could be avoided.
d) The performance comparison of the proposed technique is made with the state-of-art optimized routing techniques based on GA i.e., GASONeC [49], BMHGA [60], GAOC [18], GATERP [53] and it is found that the proposed algorithm has comprehensively performed better than other protocols.

The rest outline of this paper is given as follows: **Section 2** summarizes the related work; **Section 3** comprehensively discusses the proposed algorithm; **Section 4** shades light on the computer simulation, results, and analysis; the conclusion and future scope is reported in **Section 5**.

2. Related work

This section presents a comprehensive discussion of related work on different CH selection methods. This section is comprised as follows: (a) a short review of clustering method is given in subsection 2.1; (b) literature related to metaheuristic algorithm-based CH selection are presented in subsection 2.2; and (c) Genetic algorithm inspired cluster head selection is shown in subsection 2.3.

2.1 Clustering algorithm for WSNs

LEACH [17] initiated a distributive clustering algorithm that randomly selects CH and enhances the network performance. It is a simple routing topology in wireless sensor network. The clustering ameliorates the performing of the network from the prior algorithms and quite a few variations to LEACH were introduced. Verma et al. [18] proposed genetic algorithm-based optimized clustering (GAOC) method for CH selection in heterogeneous wireless networks. GAOC was implemented for the network having single and multiple data sink using optimized and non-optimized methods. Tyagi et al. [19] discussed the systematic advancements of the clustering algorithms. A systematic review on the clustering algorithms for heterogeneous WSN was presented in [20].

There are several concepts on heterogeneous routing and two or three levels of heterogeneous energy node. For example, stable election protocol (SEP) which operated on two levels of heterogeneous energy nodes (advance and intermediate) [21]. Despite the high probability on CH selection within the nodes, several considerable parameters (i.e., distance, density of nodes, etc.) were ignored. Later, distributed energy-efficient clustering (DEEC) [22] and DDEEC [23] were introduced where CH was employed by the residual energy of the network. But penalization of high energy nodes was ignored in [21] and [23]. On the other hand, after heterogeneous nodes of level two there were abundance of algorithm which were proposed on the level three heterogeneous energy nodes. For instance, EEHC [24] explained the process of CH selection for level three heterogeneous energy nodes. Furthermore, the performance quality of the network could not be upgraded optimally as penalization of nodes was not inevitable. To mitigate this concern, EDDEEC [25] was introduced for CH selection. For the betterment of heterogeneity levels, BEENISH [26] was proposed to select energy efficient CH and this protocol enhances the stability period of the network.

Akbar et al. [27] suggested improved balanced energy-efficient network-integrated super heterogeneous (IBEENISH), which strengthened BEENISH [26] by avoiding penalization at four-level energy heterogeneous nodes. Paola et al. [28] suggested P-SEP protocol which was designed for single-hop communication and this protocol selected CH at random rather than assigning some preference to the high energy nodes, with the same goal of improving the network lifespan and stability period. However, stable energy efficient clustering protocol (SEECP) [30] and distance-based residual energy-efficient stable election protocol (DRESEP) [29] designed for dual-hop communication and suffered from hot-spot problem due to the burdening of relaying nodes. Here, DRESEP [29] selected CH based on residual energy and distance while SEECP [30] adopted the topology of fixed number of CHs were present in the network. However, the dual hop communication involved in the inter-cluster communication was inefficient.

2.2 Metaheuristic algorithm for cluster head selection in WSNs

It is noted that the advanced technologies have not left much space for improving network efficiency, however the CH selection is observed to be NP-Hard, with optimum network performance being one of the challenging tasks [32]. As a result, there is a need of a metaheuristic approach that can satisfy critical parameters, which are required for CH selection, in an optimized manner [33]. Hence, for optimizing the cluster selection various metaheuristic methods were proposed.

Chandirasekaran and Jayabarathi [34] exploited cat swarm optimization (CSO) for selecting the CH by considering (a) signal strength, (b) residual energy, and (c) intra-cluster distance. Chawra and Gupta [35] proposed a metaheuristic-based clustering method which has utilized salp swarm optimization algorithm to target the energy hole and non-uniform load distribution in the network. John and Rodrigues [36] presented multi-objective taylor crow optimization algorithm (MOTCO) based CH selection by utilizing the several parameters like distance, energy, and traffic density of the nodes and delay in the data transmission. Lee et al. [37] implemented a spider monkey optimization for CH selection with the aim to improve the selection accuracy. Poluru et al. [38] presented an improved fruit fly optimization algorithm (IFFOA) for selecting the CH by integrating the various factors such as: (a) residual energy, (b) total distance to base-station, (c) distance from node to node link, and (d) vicinity. Rambabu et al. [39] proposed a hybrid artificial bee colony and monarchy butterfly optimization algorithm (HABC-MBOA) for CH selection. HABC-MBOA found effective in eliminating the overloading of sensor node to be selected as CH. Vijayalakshmi and Anandan [40] utilized tabu particle swarm optimization to select CH in an effective way. It is comprehended from the deep study of the aforesaid algorithms that the network performance is deprived to be optimum for one or another factor [41].

2.3 Genetic algorithm inspired cluster head selection

Further, we comprise the prospective work stated for the selection of CH by implementing the genetic algorithm (GA). GA is one of the most outstanding optimization methods. GA has been implemented to solve NP hard problems. Literature reveals [31] [42] [43] that GAs was implemented successfully for CH selection. GA based routing strategies are illustrated in Table 1.

Hussain et al. [31] implemented GA to frame clusters for energy effective data processing. This method worked well for CH selection, but residual energy was neglected. Liu et al. [42] suggested LEACH-GA for CH selection. Here, authors have given less importance to the energy factor and, therefore, LEACH-GA was noted ineffective for numerous broad area requirements in WSNs environment. Singh et al. [43] implemented an elitism dependent GA for CH selection. Here, residual energy was considered mostly for the current node. This method suffered mainly because it was not suitable to increase the network lifespan.

Attea and Khalil [10] presented evolutionary based routing protocol (ERP) protocol that focused on enhancing the stability period and network lifetime. This method suffered because CH selection and routing criteria was inefficient. Kuila et al. [44] proposed a GA protocol to perform load balancing in WSN, but it didn't discuss much on CH selection. Gupta et al. [45] recommended genetic algorithm-based clustering and routing (GACR) to achieve an improved number of rounds until the first gateway gets dead.

Elhoseny et al. [15] proposed dynamic cluster head selection using genetic algorithm (DCH-GA). It uses six factors for CH selection and designed for the multi-hop communication. DCH-GA suffered from the hot spot and extreme overhead problem. Genetic algorithm-based distance aware-leach (GADA-LEACH) [16], genetic algorithm for heterogeneous network (GAHN) [46], and application specific low power routing (ASLPR) [47] protocols are discussed in Table 1. These protocols were largely influenced by scalability and by node density as a factor of CH selection was avoidable in these methods. Yuan et al. [48] suggested genetic algorithm-based, self-organizing network clustering (GASONeC) for CH selection on the premise of predicted energy usage, distance, and node density parameter, respectively.

But it suffered because: (a) computation cost was high and (b) ineffective routing (occurs due to long-haul communication next to the time of sink placement at the boundary). Moreover, there is no approach is characterized instead of multi-hop communication appearing in huge network. In fact, there is no obvious manifestation of network remains rendered. Hamidouche et al. [49] reported low energy-efficient hierarchical clustering and routing protocol based on genetic algorithm (LECR-GA) that not only focused on the network lifetime but also on the Quality of

Table 1. Comparison of genetic algorithm-based optimization developed in literature.

Reference	GA-based technique	Targeted key factors	Integration of parameters in proposed method	Research gap
Kuila et al. [44] (2013)	GA	Consumption of energy, time of execution, active member nodes and CHs, and convergence rate number	Cluster load distribution happens while the standard variance of the gateway load is retained	An algorithm has been proposed that utilizes a sink beyond the network and uses multi-hop contact with hot-spot concern is raised.
Gupta et al. [45] (2015)	Genetic Algorithm based Clustering and Routing (GACR)	Number of hops, total distance covered in a round, number of node dead, energy consumption, and number of gateways dead.	Ratio of the gateway's residual energy included standard deviation, average distance from cluster head and sensor node and number of member sensor nodes of a CH	Gateway node collection should not take into account node density and a multi-hop connectivity issue exists.
Elhoseny et al. [46] (2015)	Genetic algorithm for heterogeneous network (GAHN)	Remaining energy of the node and network lifetime.	Sensor node density, spatial distances through energy factor.	There is a serious problem of scalability, but routing is not addressed.
Shokouhifar et al. [47] (2015)	GA-SA with ASLPR	Efficiency of the network with various sensor nodes.	Distance from the base station, distance between the CHs and the residual energy	A long-haul transmission facility drains a lot of energy since sink is situated at the middle of the network.
Bhatia et al. [16] (2016)	GADA LEACH	Throughput and network lifetime	Number of CHs, distance between CH and BS, distance between member nodes of cluster and CH, and node energy.	Inter-cluster coordination is ineffective, the process of choosing CH is also unpromising. Inefficient solution of implementing the relay node.
Elhoseny et al. [15] (2017)	DCH-GA	Efficiency of lifetime of the network with three various values of energy	Density of the node, residual energy, energy consumption of node, distance of energy consciousness, and degree of mobility	The main challenge is scalability, overheads are too high, algorithm complexity is really high
Yuan et al. (2017) [49]	GASONeC	Network lifetime with varying number of nodes	Overall distance of CH from BS and the density of the local node, the remaining energy, and initial energy	Sink positioning is carried out beyond the network, so long-distance transport contributes to heavy energy drainage
Hamidouche [49] et al. (2018)	LER-GA	Data packets received (QoS) at the base station and network lifetime	Distance of node from neighbor node, residual energy of the cluster and for routing: distance, hop and energy.	Due to the multi-hop routing hot spot dilemma, the difficulty of the algorithm significantly outweighs the benefits.
Mittal et al. [52] (2018)	Genetic algorithm-based threshold-sensitive energy-efficient routing protocol (GATERP)	Efficiency of the network with various sensor nodes, number of active CHs and active sensor nodes	Residual energy, schedule time, expected energy consumption	Scalability is the major concern and hot-spot problem arises.
Bhola et al. [57] (2020)	O-LEACH with GA	Efficiency of the network which optimum probability for selection of CH	Total energy consumption of the node and distance between member nodes of cluster and CH	Only energy usage factors are regarded when selecting CH.
Osamy et al. [32] (2020)	ETDMA-GA	Minimizes the total network latency	TDMA schedule, slots and routes of the broadcasts	Long haul transmission drains a lot of energy
Shahzad et al. [33] (2021)	GAFOR	Network lifetime with energy efficiency	Fuzzy optimized re-clustering and en-route filtering	The main problem is scalability; overheads would be excessive, and the algorithm complexity is very high.

Service (QoS) parameter to achieve an improved performance. This algorithm suffered due to its complex nature of implementation and, was not suitable to deal with the hot-spot problem.

Guo et al. [50] proposed scheduling for linked coverage enhancement in WSNs. Some literature [51] [53-54] discussed the role of other the metaheuristic algorithms for obtaining the network's optimum efficiency. Different streamlined methods [55-56] were proposed to discover the most energy-efficient variation of CH to reduce energy consumption of the network.

It can be contemplated from the study reported in the state-of-art techniques that for multi-hop communication, and the large scale of the networks suffers from the hot-spot problem [52]. The perspicacious aspect that can be obtained from Table 1 is reported as follows:

a) While various techniques have been suggested to function against CH selection by using GA, none of the techniques have optimally conserved energy. CH must be chosen such that multi-hop connectivity along with the other variables is not properly considered. Consequently, the current work incorporates a GA, which employs CMD in its fitness function.
b) The work is focused on the topology of the data scenario is considered in the SEECP protocol which includes the sink within the network.

This research uses a GA among all the metaheuristic algorithms to improve network performance. GA is a search and optimization algorithm. It remains one of the most popular optimization processes based on searching the best optimum solution. It enhances the process of optimization exploration effectively, which improve during mimicking the flexible evolution method which is accessible from nature [51]. GA have been successfully utilized proactively to devise the resolution for problems optimization concerning several disciplines viz. wireless networking, artificial intelligence, bio-medical engineering etc., [59]. GA is those metaheuristic optimization algorithms that imitate the genetic selection and advancement process, which is done naturally and biological, respectively. In this process at very first individuals are selected randomly produced as of the candidate solution and then they are included within population. GA is one of the most popular optimization procedures that supports to adapt the natural evolution process. Through GA, individuals are randomly chosen and then included in the community. Subsequently, the individuals have developed through numerous phases, including selection, crossover, and mutation. The new population of individuals is created after choosing the most appropriate or qualified individuals. The robust actions behind preferring GAs over other algorithms in metaheuristics and it has another characteristic that not only with the single point of operations, but it can also perform with parallel computations as well. Additionally, it helps to avoid the local optimal point and helps in acquiring the global optimal solution. GAs may be used to decrease the difficulty of the applications that are extracted from the derivative details by turning over the capability to the fitness function [58][60].

3. The system structure of NCOGA

This section sheds light on the proposed NCOGA algorithm. Here, we present the operational steps involved in the working of the NCOGA.

3.1 Heterogeneous model of NCOGA

For operation the proposed NCOGA employs energy heterogeneous nodes as well as 3-level of energy heterogeneity remains engaged during this network. In heterogeneous wireless sensor network with additional energy supply in some nodes. The main functions of the high-energy nodes is to act as repositories for the other nodes on the network. These nodes ensures that this role is done effectively using the GA algorithm. The quantity of normal node as N_{NORM}, advanced node as N_{ADVN} and super nodes as N_{SUP} used in the network respectively as given equation (1-9). In this model, advance node and super node are signified as high energy nodes as fraction correspond by θ and $\mho$ with total number of nodes i.e., represented by n, separately.

$$\mathrm{N_{SUP}} = \mathrm{n} * \mho \tag{1}$$

$$\mathrm{N_{ADVN}} = \mathrm{n} * \theta \tag{2}$$

$$\mathrm{N_{NORM}} = \mathrm{n} * (1 - \mho - \theta) \tag{3}$$

As compared to normal nodes, the advanced and super nodes remain ϕ and ϖ times higher in the sphere of energy. Eq. (4-9) compute the network total energy is denoted by (E_T). E_{NORM} denote the energy of normal node as well as E_{ADVN} and E_{SUP} denoted by energy of advanced and super nodes, respectively.

$$\mathrm{E_{SUP}} = \mathrm{E_O} * (1 + \varpi) * \mathrm{n} * \mho \tag{4}$$

$$\mathrm{E_{ADVN}} = \mathrm{E_O} * (1 + \phi) * \mathrm{n} * \theta \tag{5}$$

$$\mathrm{E_{NORM}} = \mathrm{E_O} * (1 - \mho - \theta) * \mathrm{n} \tag{6}$$

$$E_T = E_{ADVN} + E_{SUP} + E_{NORM} \quad (7)$$

$$E_T = E_O * (1 + \phi) * n * \mho + E_O * (1 + \varpi) * n * \theta + E_O * (1 - \mho - \theta) * n \quad (8)$$

From the above equation 8, the final expression of E_T is as follows

$$E_T = n * E_O * (1 + \varpi * \theta + \mho * \phi) \quad (9)$$

While integration parameters of fitness function in the following Subsection this above process is used in CH selection

3.2 Operational working of NCOGA

The bit frame structure will be recalculated using the validation procedure for NCOGA protocol. If the bit is '1,' then the node is a CH; otherwise, it is a '0.' To begin, the validation procedure helps select the nodes that are appropriate for further consideration.

3.2.1 Initialization

There are certain chromosomes which remain initialized based upon their performance of the validation process. The possibilities of CH nodes are considered through GA in this initialization stage. It can manage a population of various chromosome solutions where the length of one complete solution of a chromosome is equals to the number of sensor nodes in the network. It also helps to identifying the location of CH nodes and their member nodes in WSN.

Let $V_i = (V_{i1}, V_{i2}, \dots V_{iM})$ represents the i^{th} population vector of a sensor network with M sensors and their binary assignment to sensor nodes is given in equation (10).

$$V_i(j) = \begin{cases} 1, & if\ rand \le CH_P \\ 0, & if\ rand > CH_P \end{cases} \quad (10)$$

Where, $V_i(j) \in \{1, 0\}$, $j \in \{1, 2, 3, \dots, M\}$. CH_P is the suggested probability of cluster heads. Member nodes and cluster head nodes represent the values as 0 and 1 respectively and uniform random number is denoted by rand.

3.2.2 Derivation of fitness function

Fitness function stands for combination of various types of performing parameters towards set up for develop the minimization or maximization of an expression. Fitness function discusses different fitness criteria that define the individual's fitness. Fitness function incorporated in GA eliminates the weak chromosomes so as to exclude them from their further evaluation. However, the chromosomes selected via computing the fitness function are included in the competition of selecting the fittest chromosome. The optimal point is observed and the chromosome converging to the point is declared as fittest chromosome. Fitness parameter is evaluated due to its present value dependent based on several significant aspects. It should be observed that the higher the significance of the parameter, the higher the optimization would be obtained. The fitness parameters (FP) here focus on minimizing energy usage and allowing the network durability through the network. When designing fitness function, the following criteria are taken into consideration, that is to be used for selection of CH as discussed as follows.

- ***Fitness parameter 1 (Residual energy):*** A node's residual energy is one of the most important elements in determining the CH after each cycle. The CH's rotation is dependent on the node residual energy, which is why this factor was chosen. For a long time, the network's energy must be balanced by rotating the CH. Due to the heterogeneous network, the nodes with maximum energy are favoured for selection as the CH. The equation (11) is denoted as below.

$$FP_{1st} = 1 \Big/ \sum_{i=1}^{M} \left(\frac{E_{R(i)}}{E_T}\right) \tag{11}$$

In above equation (11), $E_{R(i)}$ represents residual energy and E_T represents the total energy considered to calculate the FP_{1st}. The symbol of M represents the number of nodes.

- ***Fitness parameter 2 (Initial energy):*** In this aspect, the initial energy is assigning to a node for consideration to select CH of the network. Moreover, deployment of nodes in the network as per initial energy consideration together with heterogeneity nodes. According to heterogeneity of nodes in the comparison of super nodes maintain lengthier time instead of advanced nodes as well as advances nodes remain ideal than normal nodes. FP_{2nd} is the second parameter for fitness function as shown in eq. (12), a node of initial energy is normalized to have value between 0 and 1.

$$FP_{2nd} = 1/E(i) \tag{12}$$

- ***Fitness parameter 3 (Distance between node and sink):*** The energy consumption of a node decides the communication among the nodes or with the sink. The sum of energy that is consumed by the sink remains strictly proportional to the distance from the sink to the node. Therefore, the networking technique for CH selection takes into account of the parameter under which the median gap between the nodes and the sink can be optimized accordingly. The target purpose rather than CH selection decisions should be framed, FP_{3rd} is presented in conjunction with the equation (13) with the distance variable.

$$FP_{3rd} = \sum_{i=1}^{M} \left(\frac{D_{(N(i)-S)}}{D_{AVG(N(i)-S)}} \right) \tag{13}$$

where,

$$D_{AVG(N(i)-S)} = \left(\frac{\sum_{i=1,}^{N} D_{(N(i)-S)}}{M} \right) \tag{14}$$

The above equation (13) evaluates the sum of the distance costs obtained for any i^{th} node by third fitness parameter (FP_{3rd}), where i ranges from 1 to M (Number of nodes) and in the equation (14), $D_{N(i)-S}$ denotes the Euclidean distance from the sink to the i^{th} node, $D_{AVG(N(i)-S)}$ signifies an average distance at the middle of the i^{th} node and sink.

- ***Fitness parameter 4 (Neighbors of a node):*** In this respect, the intra-cluster connectivity improves a controlling entity until the network has developed throughout the larger area. If the selection of CH is independent of the number of adjacent nodes, it contributes to the selection of a node as a CH that is far from other nodes. The CH node would also use more resources when gathering data from the other nodes in a cluster. To avoid this kind of selection, we must consider the number of neighboring nodes. FP_{4th} is defined by the following eq. (15).

$$FP_{4th} = \frac{N_{dist}}{N_{CL}} \tag{15}$$

where,

$$N_{dist} = \sum_{i=1,j=1,i\neq j}^{N_{CL}} D_{(N(i)-N(j))} \tag{16}$$

In the above equation (15), the ratio of N_{dist} and N_{CL} is evaluated by $\mathrm{FP}_{4\mathrm{th}}$. Whereas N_{dist} denotes the distance intervals of neighbor nodes and N_{CL} denotes the total number of nodes in the cluster. In equation (16), the distance interval between i^{th} and j^{th} nodes is denoted by $\mathrm{D}_{(\mathrm{N(i)-N(j)})}$.

- ***Fitness parameter 5 (Communicating mode decider):*** One of the essential contributions in this work is the involvement of CMD factor in the CH selection. CMD decides a node's ability to be relay CH for performing dual hop communication. When the average distance among the node and sink remains lesser than the threshold distance, the node is preferred to be CH and it will be forwarding data to the sink directly. The aforementioned threshold distance is the standard distance of the whole nodes from the sink. It is to be noted that the CH forwarding data from the farther located clusters is termed as Relay CH (RCH). CMD factor is computed by calculating the distance between the nearby CHs or between the sink whichever is nearest. Minimum CMD value to be enhance a node as a CH in the network.

Therefore, the fifth fitness parameter ($\mathrm{FP_{5th}}$) computes CMD such that (CMD=1/ $\mathrm{FP_{5th}}$) for a node as follows. $\mathrm{N_C}$ is the number of clusters or the number of CHs in the network.

$$FP_{5th} = \left(\frac{\sum_{i=1,j=1}^{N_C} D_{(CH(i)-CH(j))}}{N_C} + D_{(N(i)-S)} \right) \tag{17}$$

Through the equation (17), $D_{(CH(i)-CH(j))}$ signifies the distance between the CHs and $D_{(N(i)-S)}$ denotes the distance between the i^{th} node and sink. Lesser the value of FP_{5th}, more will be value of CMD. Hence, the node will be selected with least distance such that it follows single hop communication. Otherwise, it will act as RCH node.

It is noted that the distance and CMD factors inclusion hold their individual significance. The CH selection of a node is governed through the distance factor, which supports the selection of nearest node towards the sink. However, CMD factor helps to select appropriate node as relay CH.

- ***Fitness parameter 6 (Load balancing):*** It improves during network durability as it improves in preventing the imprisonment of excessive energy nodes have been carefully chosen as CH repeatedly no matter that they have energy. This one, like a CH, may pick a node at any time throughout its ten-round period of time. The node that hasn't yet become CH is still increasing its chances of becoming CH as the term concludes. Initially, each node in the competition for improving CH is given a value of 0.1. With each the increase in intensity during the round, the value will be improved through 0.1 till it reaches 1 as shown in Fig. 1. As quickly as possible as it reaches 1 that one becomes CH. Therefore, sixth fitness parameter includes load balancing given as follows in equation (18).

$$FP_{6th} = \frac{1}{\sum_{i=0.1}^{1} L.B.(i)} \tag{18}$$

It is to be noted that higher the value of L.B. for a node, more chances to enhance a node as a CH. The reason for employing these six fitness parameters is the fact that the inclusion of these parameters altogether brings the energy balancing through the network. Every parameter is essentially explained individually, and the fitness function thus framed leaves a great impact on the CH selection by selecting the best profile node as CH. The existing techniques used different parameters for CH selection, but the above-discussed parameters are not included altogether. It is believed that this process of inclusion will help in acquiring the energy efficient CH selection.

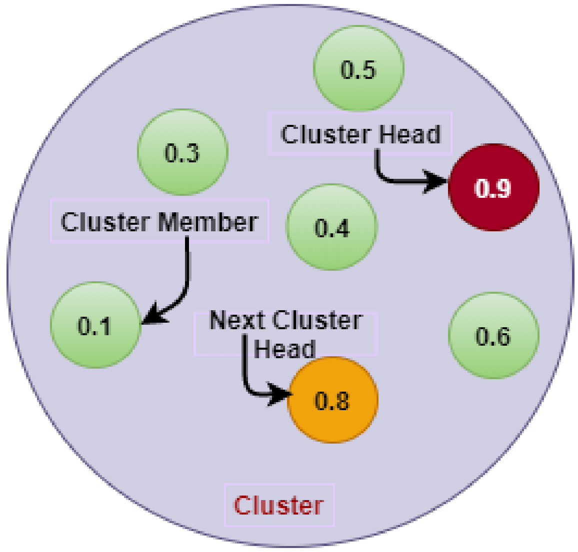

Fig. 1. Load balancing scenario.

- Fitness function: We considered above the integration of various fitness parameters of the network is completely express in given fitness function as follows in equation (19).

$$F = \frac{1}{\varphi \times FP_{1st} + \delta \times FP_{2nd} + \gamma \times FP_{3rd} + \alpha \times FP_{4th} + \sigma \times FP_{5th} + \beta \times FP_{6th}} \quad (19)$$

To achieve maximum network efficiency at the lowest feasible cost, it is necessary to reduce the fitness function F in equation (16). This basically includes various fitness parameters which are computed in eq. (11-18). In this eq. (19), the weight coefficients are used which are used to give different weightage to the different parameters used in the integration of fitness function. It depends upon the user to tune these parameters according to the application for which the sensor network is employed.

The reason for keeping the sum of all weighted fitness parameters in denominator is the fact that the proposed algorithm is to minimize the objective function and must maximize the weighted sum. In eq. (19), consequent fitness parameters are increased with weight coefficients i.e., φ,δ,γ,α,σ, and β. Such elements are uniformly weighted as given as follows in equation (20). To be precise, select two fittest indices by using fitness function and the weight coefficients are all set to 1/6. However, we obtained the best outcome by setting the weight coefficients to 1/6.

$$\varphi + \delta + \gamma + \alpha + \sigma + \beta = 1 \tag{20}$$

Hence, the main objective of GA is to minimize this function over metaheuristic processes to perform for optimum performance of the network.

3.2.3. Binary tournament selection

The selection process is adopted for selecting most suitable individuals to be used as parents for next generation through crossing and mutation. In this work, the process of binary tournament selection is used to determine the fittest chromosome. Such approach is implemented in a way that any fittest chromosome is assigned a rank and distance metric.

Binary tournament method starts with two players chosen randomly with a comparison operator, $(\leq_n)$ where one winner is selected with the help of equation (21). Two attributes are taken into consideration for comparison operator and are as follows: (a) non-domination rank (i_{rank}); and (b) Crowding distance ($i_{cr_{dist}}$). Let i and j are two fittest chromosomes chosen for tournament, then the crowded operator is described using equation (21).

$$i \leq_n j \ if \ (i_{rank} < j_{rank}) \ or \ ((i_{rank} = j_{rank}) and \left(i_{cr_{dist}} > j_{cr_{dist}}\right)) \tag{21}$$

3.2.4 Adaptive crossover

After performing mating operation, the recombination of component material is done. When two parents are crossed over, a new set of offspring is created. Adaptive crossover technique is adopted here to enhance the GAs performance that selects two better individuals as parents for generating new individuals [60]. The crossover probability of each individual chromosome is obtained from their genome. In this process a chromosome C_k is chosen randomly as a parent and the second partner for this crossover is obtained from the matrix variance U_{kl} for each chromosome C_l associated with C_k having highest U_{kl}, as given in equation (22).

$$U_{kl} = \sum_{n=1}^{M} \sum_{i=1}^{\max(|G_{kn}|,|G_{ln}|)} (g_{kni} - g_{lni})^2 \tag{22}$$

Where G_{kn} is the gene ϵ C_k and $|G_{kn}|$ is the number of sets that belongs to s_k sensor and g_{kni} is the set number. The Chromosome C_l having superior U_{kl} is consider as other parent. If the genetic coding of all chromosomes is same in the population, then the other parent will be chosen with maximum crossover rate P_c between 0.5 to 1. The Crossover takes place between C_k and C_l chromosomes for production of new offspring $\bar{C}_k$ and $\bar{C}_l$.

3.2.5 Mutation

To explore a better chromosome, we have implemented move mutation, which ensures that a new gene sequence is inserted into a chromosome. Move mutator approach, transfers one data point from one offspring chromosome to another offspring of the cluster to avoid the stuck of evolutionary optimization among the local optimum. The mutation rate determines how often mutation is introduced for better chromosomes. It is observed that the novel mutation operation contributes to improved results. Mutation () presents steps used for performing mutation operation.

Mutation ()

1. Randomly generate $n \in \{1, M\}, l \in \{1, J_i\}$ and let $x_n \in \bar{C}_k$.
2. **If** $\bar{C}_l \neq \bar{C}_k$ **then** move x_n from $\bar{C}_k$ into $\bar{C}_l$, **else** go to the step 1.
3. Again, evaluate the offspring of $\bar{C}_l$ and $\bar{C}_k$ and change the corresponding values of the individual.

3.2.6 Termination

These stages are repeated as maximum times as necessary until the prerequisites are met. A new version of the chromosome is added until the best possible value is found. Fitness values are saved and compared to prior values after each iteration, and only the fittest are saved. After iteration, the node with the highest fitness value is associated with the fittest chromosome. Following that, the better chromosomes or nodes are selected as CHs accordingly.

3.2.7 Pseudocode of the NCOGA

Algorithm 1 presents the pseudocode of the proposed NCOGA algorithm. We highlighted terms used, input and output for the proposed algorithm.

Algorithm - 1. NCOGA Algorithm for CH selection.

Term used: *N*: number of chromosomes; *M*: number of sensor nodes; *Pc*: crossover rate; *Pm*: mutation rate; CH_P: The probability of cluster head selection; *gen*: generation number; *itermax*: number of iterations; *genmax*: maximum number of generations; FP_{1st}, FP_{2nd}, FP_{3rd}, FP_{4th}, FP_{5th}, FP_{6th}: fitness parameters.

Input: Chromosomes *Vi, 1≤ i ≤ N*

Output: *Best_fit_chrome*

1. **Begin**
2. Initialization of *Pc, Pm*, and *gen=0*.

While *round* <= *itermax* **Do** /*Loop to check termination */

For *i =1: N* ***Do*** */*Initialization of nodes for CHs with probability CHp */*

  For *j = 1: M* **Do**

$$V_i(j) = \begin{cases} 1, & if\ rand \leq CH_P \\ 0, & if\ rand > CH_P \end{cases}$$

  End For

End For

No_of_CHs = 0 /*Obtained the number of eligible CHs */

For *j = 1: M* **Do**

  If (*Vi(j) = = 1*) **then**

    No_of_CHs = No_of_CHs + 1

  End If

End For

For *i = 1: No_of_CHs* **Do**

  For *j = 1: M* **Do**

    Obtained i_{rank} and $i_{cr_{dist}}$ to add into cluster

  End For

End For

Determine FP_{1st},FP_{2nd}, FP_{3rd}, FP_{4th}, FP_{5th}, FP_{6th} through *Equations (11) – (18).*

Fitness_value(i) = F(*Vi*) /*using *Equation (19)* */

For *gen = 1: genmax* **Do**

  For *i= 1: N* **Do**

    Select two fittest indices (by using fitness function) using Equation (21) /*Selection*/

  End For

  For *j = 1: M* ***Do***

    New offspring $\bar{C}_k$ and $\bar{C}_l$ from parent C_k and C_l using Eq (22) /*Crossover*/

    Perform mutation operation using **Mutation ()** /*Mutation*/

  End For

    For *i= 1: N* **Do**

      If *Best_fit_chrome* < *Fitness_value(i)* **then**

        Best_fit_chrome = *Fitness_value(i)*

33. **Else**

34. *Best_fit_chrome = V(i)*

35. **End For**

36. **End For**

37. **End While**

38. End

**Terms used are discussed in abbreviations table.*

We assume the set of chromosomes *Vi = {V1, V2, . . ., VN}* of the population where *N* is the number chromosomes. Each chromosome is represented by a bit frame structure or vector. The size of the bit frame or vector is *M*, where *M* is the number of nodes. The first location in the frame corresponds to a node with ID one. If the node selected as CH, its corresponding location in the vector updates with '1' otherwise '0'. The objective of the GA-based algorithm is to the selection of CHs. After initialization, each chromosome is evaluated using the fitness function defined in Eq. (19), The CH selection method used in the proposed GA is presented as Algorithm 1. The fitness value for each chromosome in *Vi* is evaluated (line 21). In the initial iteration each chromosome *Vi* itself is the best fittest value. The output *Best_fit_chromo* represents the final selection of CHs. Next, the process continues until the stopping criteria are acquired (line 3-37). In line 22, the generation count is increased by one, i.e., (gen=gen+1) and. we used tournament selection to choose two fittest indices (by using fitness function) using Eq. (21). For offspring generation, crossover operation has been implemented as indicated in Eq. 22. To get the new offspring $\bar{C}_k$ and $\bar{C}_l$ from parent C_k and C_l the perform crossover operation with crossover rate *Pc*. The new offspring *mutation ()* is further exploited by the mutation operation with pre-defined mutation rate *Pm.* After mutation operation, again calculated the fitness value of new generation and update the best fitness value, i.e., *Best_fit_chromo*. The stopping criteria is the maximum number of iterations is set by the user. Algorithm 1 presents the overall process of the GA implemented for CH selection.

As discussed above, the operation of GA follows the steps defined in Algorithm 1 and also demonstrated in the Fig. 2. Once the heterogeneous nodes are deployed in the network of defined dimensions and sink being placed at the middle of the network, the role of GA comes into play. The process of clustering and CH selection is comprehended by operation of GA involving various stages as explained above. Once the CH is selected the network operates in steady state phase. Henceforth, the operation is terminated once the energy of all nodes is exhausted.

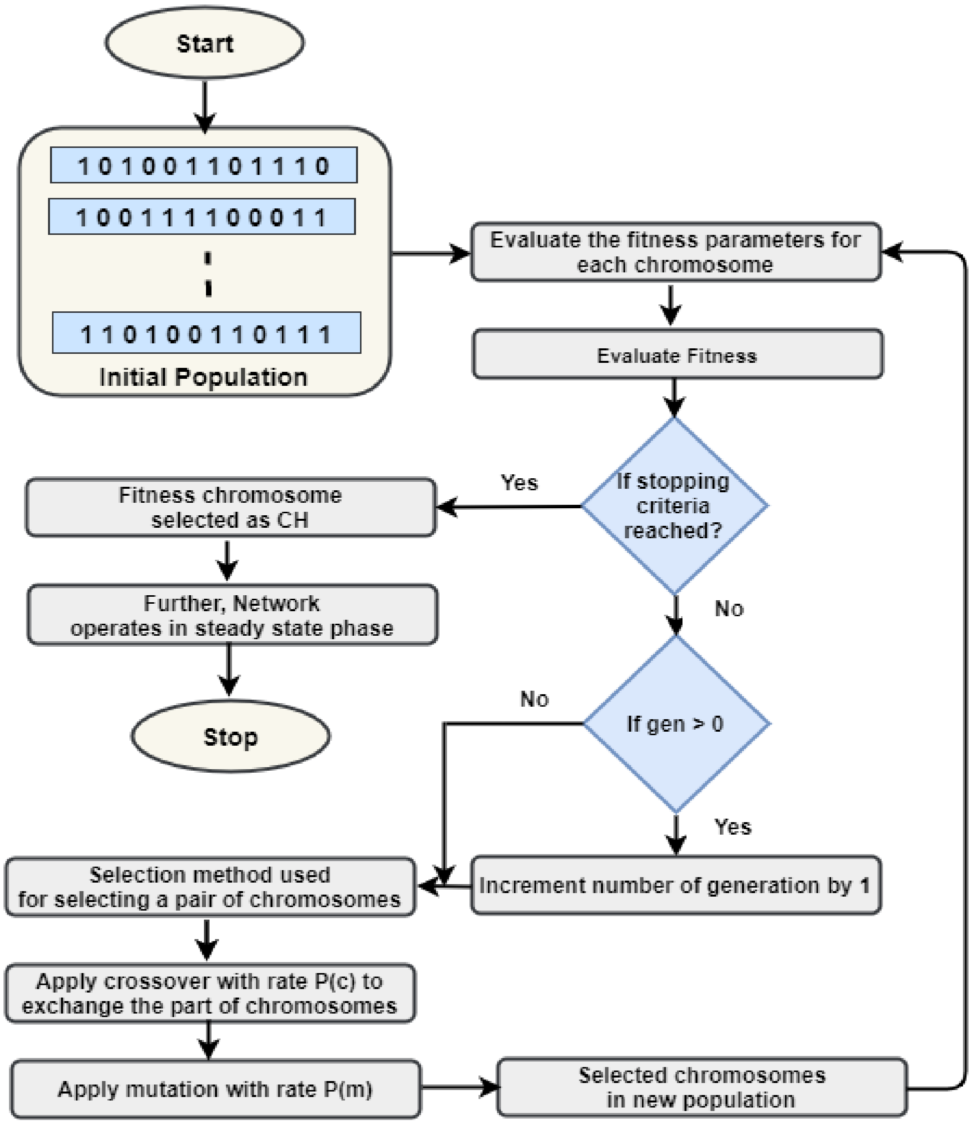

Fig. 2. GA outfit flowchart in NCOGA.

3.2.8 Complexity analysis of NCOGA

In terms of the complexity of the viability of proposed algorithms, it is imperative to conduct the real-time analysis of the proposed algorithm. The complexity of the algorithm is found to be O ($r_{max} \times N$) which can be seen in Algorithm 1, where N is represented as the size of population and r_{max} is represented the maximum number of rounds for which the network is performed.

3.3 The operational process of NCOGA

In Fig. 3, the operational process of NCOGA is elaborate systematically. The protocol is operated in one of the phases called set-up-phase and another called steady-state-phase that are described given as get along.

Set-up-phase: A network planned in compliance with the following phases. The node with energy heterogeneity such as normal, intermediate, and advance node levels is used arbitrarily within the intended network region. The node is then transferred to somewhere in the center of the communication network for transport from the cluster to the consumer over the Network. Subsequently completing the deployment of the heterogeneous node, the selection of CH by nodes is performed for each cluster using GA. By this set up phase it seems this clustering technique is conventional, but there is a unique promising approach at the selection of CH. CH selection from each cluster is performed by exploiting six parameters becoming it to enhance the durability of network.

Steady-state-phase: In this phase NCOGA executes between Cluster Head and sink communication after completing set up phase. As the NCOGA is a responsive protocol in such a way that it does about idea of hard threshold as well as soft threshold like the implementation in the TSEP protocol. The outcomes of the transition determine whether or not to use a hard or soft threshold. If the actual sensed value indicated by (C(V)) is larger than a hard threshold (H(T)), intra-cluster data will be transported from the node into the CH. Furthermore, only the following round of contact is established if the difference between the current value and the previously detected value surpasses the predetermined soft threshold. If this doesn't happen, the following phase of data transmission will be delayed. Moreover, if the node is gone, and the energy of the node is depleted, the dead nodes would be connected to the dead nodes by an increase of 1. However, since the data was transmitted in accordance with CH and afterwards an aggregation of valuable data was transmitted to the sink from CH, aggregation of data for CH persists. The same procedure is replicated until all nodes are dead, and it is said that the network is stop working when all nodes are dead.

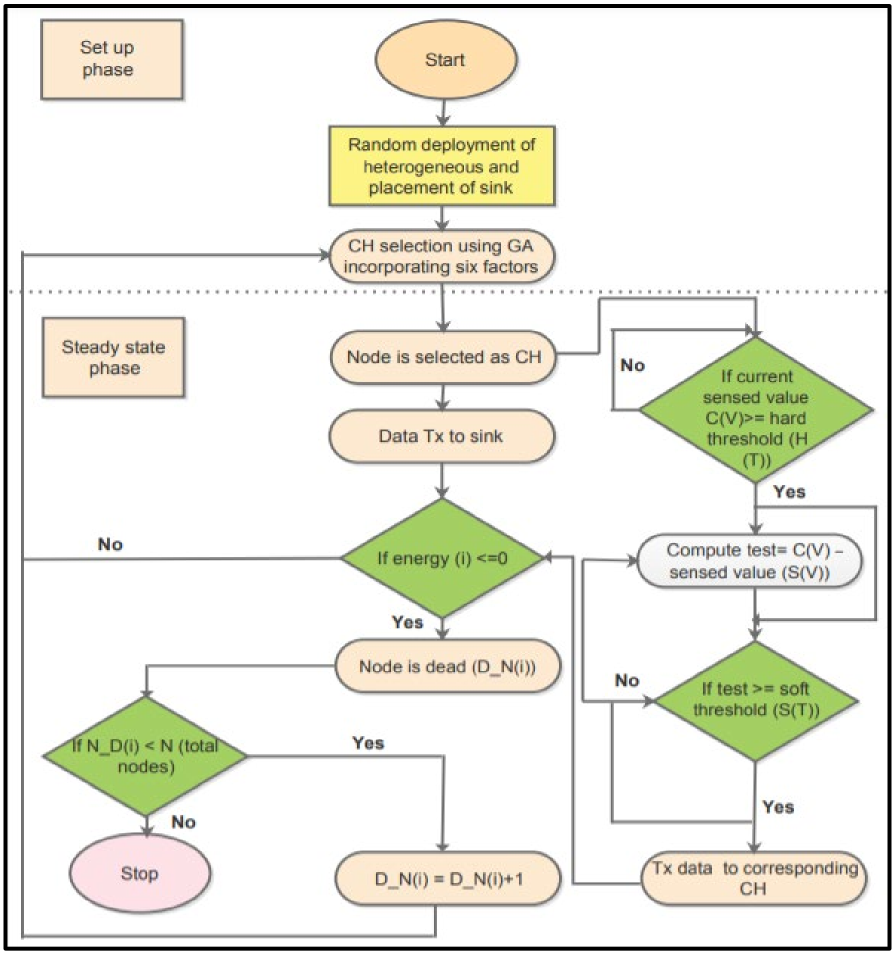

Fig. 3. Overall Flowchart for working method of NCOGA Protocol.

Lemma 1. NCOGA ceases throughout preset iterations $It_r = O(1)$ and CH selection is enhanced in each and every important aspect.

***Proof*:** After completion of the setup process, NCOGA performs the steady-state process, which gathers data and combines it to the CH. Moreover, the unavoidable data is transferred over to the sink and this phase can proceed until the network is dead, i.e., all the network nodes are drained. As nodes begin to transfer data on to CH, they concurrently deplete their energies, and it depends on the difficult and soft threshold. Until the node loses its resources or becomes a dead node the next round of data transfer cannot be done. This method would stay unchanged until all nodes have been depleted. Thus, as the number of iterations grows, the number of nodes logically reduces. Which ensures that the numbers of death nodes and iterations are inversely proportional. In addition, hence, the number of dead nodes stays the same and the NCOGA procedure finishes with a fixed iteration.

Residual energy and original energy are more desirable for CH selection when more energy has been implanted. Consequently, NCOGA's main purpose involves many primary variables that consider beneficial energy that also takes into account the energy source for that operation. Furthermore, it's another significant element in power usage that separates the nodes from the drain. This distance component and energy consumption are mutually proportional. This interval between the nodes and the sink node often enables energy usage to be decreased to be reduced. One of the major significant facts is that optimizes the CH selection to the significant level is the involvement of CMD and load balancing factor that makes it possible for nodes to avoid hot spot problem in the network as well.

3.4 Network structure of NCOGA

Here, we explore a heterogeneous network with three levels of heterogeneity energy. These nodes consume enough energy when processing data. To restrict this energy consumption here we discussed about an energy model i.e., sensor radio energy paradigm. This model adopts the quantity of energy consumed that make up the nodes for data transmission for NCOGA until it became completely exhausted.

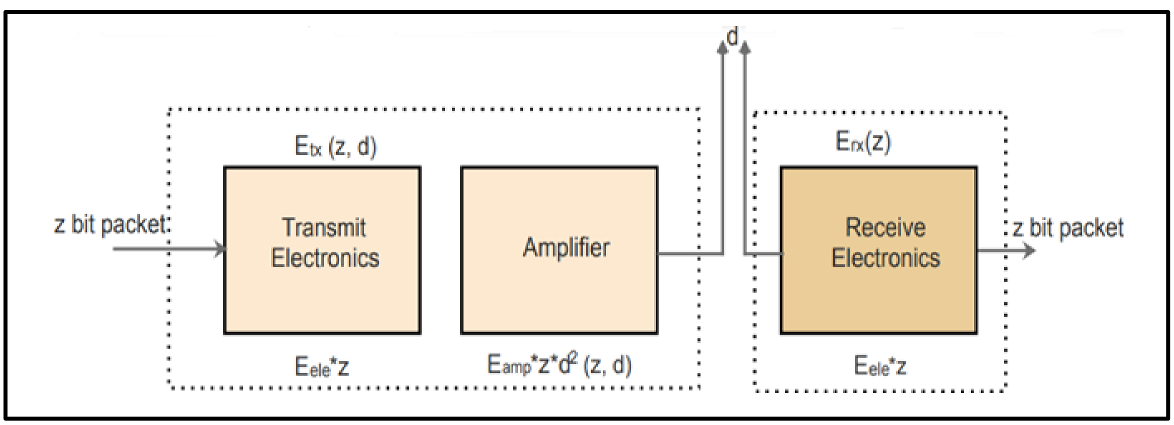

Fig. 4. Debauchery Paradigm of Radio Energy.

3.4.1 Sensor radio energy paradigm

Fig. 4 depicts the suggested system structure for NCOGA's sensor radio energy paradigm, which is responsible for consuming the nodes' energy consumption. The energy consumption of the data transceiver is elaborated in equations (23-27). The consumption of total energy during data transmission is inversely conditional on the distance between nodes. Explanation for the following equation let assume, transferred data by nodes is z-bit and the distance between the nodes is 'd 'and the consumed energy is indicated by $E_{tx}(z, d)$ and provided equation looks like the following.

$$\mathrm{E_{tx}}(\mathrm{z},\mathrm{d}) = z * \mathrm{E_{ele}} + z * \mathrm{E_{efs}} * \mathrm{d}^2 \quad \text{for } \mathrm{d} \leq \mathrm{d_o} \tag{23}$$

$$\mathrm{E_{tx}}(\mathrm{z},\mathrm{d}) = z * \mathrm{E_{ele}} + z * \mathrm{E_{amp}} * \mathrm{d}^4 \quad \text{for } \mathrm{d} > \mathrm{d_o} \tag{24}$$

The distance between nodes and sink in eq. (23-24) is denoted by 'd'. E_{ele} denotes the energy consumed instead of initiating a transmitter and a receiver circuitry. *'d_o'* indicates a threshold distance and shall be expressed as in eq. (25).

$$\mathrm{d_o} = \sqrt{\frac{\mathrm{E_{efs}}}{\mathrm{E_{amp}}}} \tag{25}$$

Transmission amplifier characteristics are specified by E_{efs} and E_{amp} where E_{efs} is defined as "free space energy model (power loss d^2)" and E_{amp} is defined as the "energy consumption for multi path energy model (power loss d^4)".

Energy consumption for the data per bit is provided by the equation (26).

$$\mathrm{E_{rx}}(\mathrm{z}) = z * \mathrm{E_{ele}} \tag{26}$$

The eq. (27) specifies the amount of energy required by the CH to aggregate the data.

$$\mathrm{E_{dx}}(z) = \mathrm{x} * z * \mathrm{E_{da}} \tag{27}$$

where $\mathrm{E_{rx}}(\mathrm{z})$ is defined as the energy consumed while receiving z-bit of data. E_{da} is defined as the energy consumption of the data aggregation of 1-bit data. Moreover, $E_{dx}(z)$ is defined as the "energy expenditure during data aggregation of received z-bit data of x number of data packets".

3.4.2 Assumption of network model for NCOGA

The sensor node properties had some impacts on the framing of the network. Therefore, we have to look after the characteristics of the frame for NCOGA with sensor nodes and the attributes are the following:

i. The network remains utilize in order which shows that the nodes stay stagnant and sinks in the network.
ii. Particularly super nodes, advanced nodes, and regular nodes are three-level energy heterogeneity nodes inside a heterogeneous framework. Super nodes are made up with maximal energy between the nodes and also the smallest of regular nodes.
iii. For the nodes, the energy use is limited although there is no constraint for the drain. Therefore,

the nodes are drained after a certain round of data transfer to CH. Therefore, no such constraint on power usage for sinks is different for nodes.

iv. Location is not mentioned anywhere in the circuit of the nodes.

v. There are some other considerations regarding signal dilution which are not considered here.

vi. The shape of the network area is supposed to be square.

vii. In Euclidean node, distance is considered the measured signal intensity of the receiver Signal strength Signal Indicator (RSSI).

Real time application for GA based CH selection:

Monitoring the temperature inside of a building using WSNs in real world environment allows for an examination of GA-based CH selection. The sensor node incorporates a microcontroller, a low power transceiver, and a thermistor into its construction to measure the temperature of its surrounding environment. The temperature of the surrounding environment is garnered by the sensor node, which then relays this information to the sink through the CHs. In the experiment that we are carrying out, the placement of each sensor node has already been decided upon by computing the Euclidean distance between the sensor nodes. The WSN node that is employed in the experimental set up is equipped with both a transmitter and a receiver, and the communication distance between two sensor nodes is calculated based on the value of the received signal strength. Table 2 represent the $RSS(i)$ value of node *i* of a cluster and CH_P value of sensor nodes.

Table 2 Initial CH information of CH_P and $RSS(i)$

Number of Sensor Node	**Received Signal Strength ($RSS(i)$)**	CH_P
M1	0.7	0.11
M2	**0.3**	**0.05**
M3	0.8	0.13
M4	**0.2**	**0.03**
M5	0.4	0.06
M6	0.5	0.08

$$CH_P = \frac{RSS(i)}{\sum_{i=1}^{M} M_i} \tag{28}$$

where, CH_P is the suggested probability of cluster heads at initial selection as per eq. (10). Sensor nodes and cluster head nodes represent the values as 0 and 1 respectively. $RSS(i)$ is denotes the received signal strength of node i, and M is the total number of sensor nodes represent in eq. (28).

The initial CH is chosen by calculating the RSS values of all of the sensor nodes in the network. The RSSI values of the sensor nodes that are located within 10 meters of the sink are measured as follows: [0.7, 0.3, 0.8, 0.2, 0.4, 0.5] and the eq. (28) used to determine the probability of a sensor node becoming the initial CH as follows: [0.11, 0.05, 0.13, 0.03, 0.06, 0.08]. As per eq. (10), the CH_P value of **[0.05, 0.03]** is designated as CH and others are the member nodes in the first round. For each round, the fitness value is calculated using eq. (19) and the node with a best fitness is chosen as the CH. The outcome is recorded in a database that is referred to as the node information data table. This information is sent to the CH and other cluster members during a first-round network operation. In order to establish a transmission and receiving schedule for each of its member nodes, the CH must first assemble its members. After obtaining the transmission schedule from the cluster's CH, the members of the cluster transmit messages to that CH. The CH gathers data from its member nodes, then sending the aggregated data to sink. The microcontroller's channel is utilized to monitor the real temperature.

3.5 Time Complexity Analysis

The computational complexity of CH selection and relay CH selection is the feasibility for its real time implementation. In NCOGA method, time complexity reduces and convergence speed increases. The time complexity of CH selection and relay CH selection is represented as O ($r_{max} \times M \times k^2 \times T_{fit}$), and O ($r_{max} \times M \times k \times T_{fit}$), where M is represented as the total nodes and r_{max} is represented the maximum number of rounds, k represent the total number of CH and T_{fit} is the time complexity of fitness function.

4. Simulation, results and analysis

This section elaborates on simulation setting, performance metrices, state-of-the-art methods used for performance comparison and statistical analysis.

4.1 Simulation settings

All the simulations have been conducted using MATLAB R2016a with system configuration as 4 GB RAM, 1 TB HD, Intel i5 processor with 2.60 GHz and Window 7. Network parameters and sensor radio energy paradigm used in simulations are summarized in Table 3.

Table 3. Simulation parameters of NCOGA

Parameters	***Values***
Size of Area	*100 m ×100 m*
Nodes (N)	*100-200*
Sink for NCOGA	*1*
Initial energy (Eo) of nodes	*0.5 Joule*
Heterogeneous nodes	*Advanced, Super, and Normal node*
Super nodes (Ʊ), advanced nodes (θ) as per energy fraction	*Ʊ=1 Joule, θ=2 Joule*
Proportions value of Super node (ϖ), and Advance node (ϕ)	*ϖ=0.1, ϕ=0.2*
Essential transceiver energy (E_{elc})	*50nJ/bit*
Threshold-distance (d_o)	*87m*
Size of the data packets	*2000bits*
Small distance energy $d \le d_o$ (E_{efs})	*10pJ/bit/m*^{2}
Large distance energy $d > d_o$ (E_{mp})	*0.0013pJ/bit/m*^{4}
Utilization of energy in data aggregation (E_{da})	*5nJ/bit/signal*
CH_P	*0.05*
Size of the population (P)	*30*
P_c	*0.5-1.0*
P_m	*0.031*
Selection method	*Binary Tournament*
Total chromosomes	*30*
No. of Generations	*30*
Simulation run	*20*

In our simulation, 100 nodes were deployed randomly over (100 m x 100 m) with the network of different energy nodes. The energy heterogeneous nodes which include 50% of normal nodes (low energy nodes), 30% advanced nodes (Medium energy nodes), and 20% super nodes (High energy nodes). In this aspect, the initial energy is assigning to a node for consideration to select CH of the network. Moreover, deployment of nodes in the network as per initial energy consideration together with heterogeneity nodes. According to heterogeneity of nodes in the comparison of super nodes maintain lengthier time instead of advanced nodes as well as advances nodes remain ideal than normal nodes. A node of initial energy is normalized to have value between 0 and 1. Three level of sensor nodes and energy fraction nodes, and GA parameters remain summarized in the Table 3, so it precisely provides the standard values for population size, 'mutation and crossover rate', number of generation, and other parameters are employed into deliberation for performance GA processes for CH selection. The binary tournament selection method was utilized before termination, elitism process was followed.

It is observed that its nodes are separated into various genres based on their initial energy resources and established specific identifiers. However, the energy of the node reduces due to the radical data transfer. The designed algorithm works in a way that the algorithm designed for different categories of sensor nodes works according to the energy profile of these nodes.

There would be a period when the nodes in one group will move between the genres. The truth is, though, that the guidelines for all node genres remain the same for the network's entire working. For starters, if a node called a super node goes beyond normal node energy, then it also has the same CH selection criterion as the super nodes.

4.2 Performance metrics

Standard performance metrics are used to validate the performance of the proposed NOCGA. Five performance metrics such as (a) stability period, (b) network longevity, (c) number of dead nodes against rounds, (d) throughput, and (e) network's remaining energy are considered. The rationale behind selection of these metrics is discussed below:

4.2.1 Stability period

The number of rounds completed may be calculated until each node has completely depleted its energy supply, at which point it becomes a dead node. This parameter may have a substantial impact on the performance of tasks in which a loss of information simply cannot be accepted. The higher stability benefit assures better durability.

4.2.2 Network longevity

To assess the efficiency of covered rounds while there is no node left alive in the network to communicate. For certain roles where the management of the network is a continuous operation, it is important. Any of the scenarios where the durability of the network is extremely relevant are applications such as farming, flood prediction, etc.

4.2.3 Number of dead nodes against rounds

The efficiency appraisal observes the protocol's responses as energy is expended through data transfer, the rate of energy exhaustion of the nodes before they are completely dead.

4.2.4 Throughput

The number of data packets following the efficient transmission can be described as an output. The QoS status for each node is explained. If QoS is inadequate for any routing protocol, the stability period and network long life have no importance.

4.2.5 Network's remaining energy

At various number of rounds, the network remaining energy is monitored. It displays the network load balance. Network energy is simply the amount of energy used within the network by all nodes.

4.3 Performance comparison with state-of-the-art algorithm

The performance of the proposed NCOGA is tested against the most recently developed state-of-the-art algorithms such as: (a) BMHGA [60]; (b) GAOC [18]; (c) GASONeC [48]; and GATERP [52]. We selected these algorithms because of the following reasons: (a) these algorithms have exploited GA technique, (b) having compared with these techniques it renders a fair comparison to reflect the fact that the results obtained are not just due to the characteristics of GA, rather it is due to the proposed approach of CH selection.

4.4 Results and analysis

This section presents the results and analysis of the NCOGA based on the five-performance metrics are discussed in Section 4.2

4.4.1 Stability Period

It is seen that in NCOGA, after 6631 rounds first node is dead but in the situation of GATERP [52], BMHGA [60], GAOC [18], and GASONeC [48], it remains just 5853, 5446, 5136, and 4180 rounds, respectively as presented in Fig. 5. The important thing is the understanding that NCOGA enhances stability period in accordance with the 21.75%, 13.29%, 29.1%, and 58.63% as compare to the protocols BMHGA, GATERP, GAOC, and GASONeC, respectively. An improvement such as stability period as well as HND is the unification of six fitness factors that make sure the energy conservation even as in the process of data transmission. Therefore, distance between the nodes and nodes as well as sink to nodes is efficiently decreased.

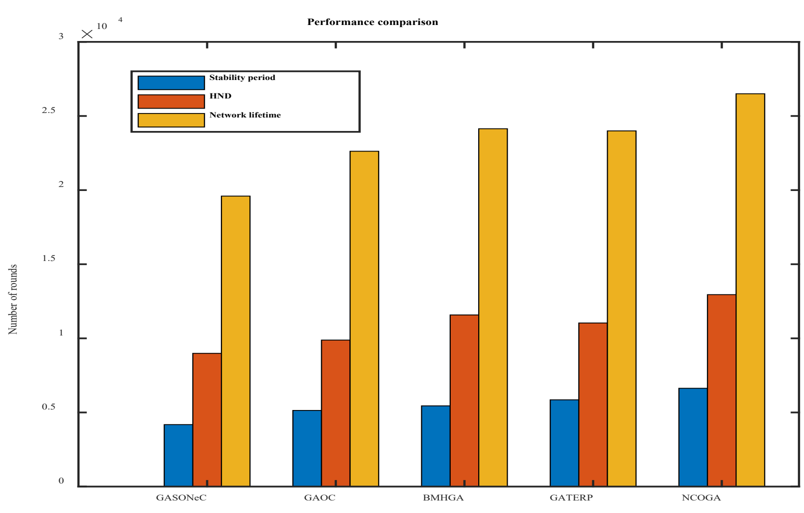

Fig. 5 Stability Period, HND and Network lifetime comparative analysis of NCOGA with other protocols.

Stability period can be grasped that NCOGA outperform the state-of-art protocols from the perspective of the networks. The incorporation of energy efficient fitness parameters in fitness function is a major factor for this improvement in stability period. The integration of load balancing and energy factors in CH selection decreases unbalanced and unexpected energy usage and improves preserve energy.

4.4.2 Network Lifetime

It is seen that in NCOGA is completed at 26500 rounds while the network lifetime just for the BMHGA [60], GATERP [52], GAOC [18], and GASONeC [48] has been seen on 24140, 23996, 22621, and 19595 rounds, respectively. From the analysis, it can be seen that 2360, 2504, 3879, and 6905 more rounds are covered by NCOGA in comparison with the BMHGA, GATERP, GAOC, and GASONeC protocols, correspondingly outlined in the fig. 6. Enhancement of network lifetime is monitored because of load-balancing and CMD factors combined in objective function in NCOGA. The communication mode decider (CMD) factor encourages the CH selection of a node that has more surrounding nodes. Thus, total network energy is conserved, extending network lifespan. Therefore, with large number of surrounded nearby nodes, a standard distance amongst node and CH is reduced broadly.

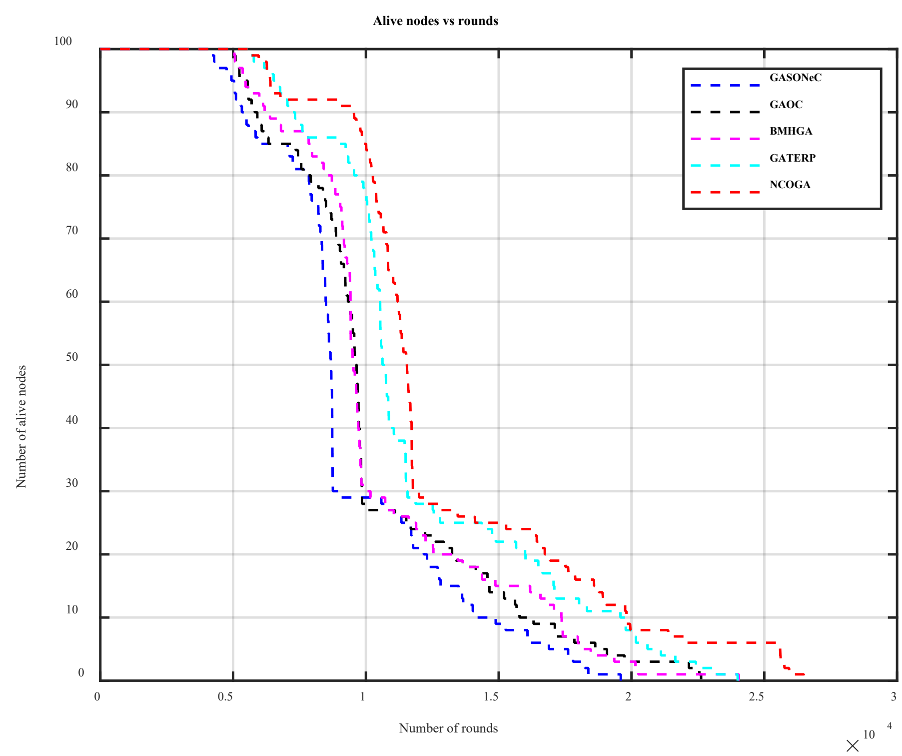

Fig. 6. Comparative analysis of alive nodes vs rounds of NCOGA with other protocols.

4.4.3 Number of dead nodes against rounds

NCOGA is seen to have a lower number of rounds compared to other protocols with regard to the number of dead nodes. According to the Fig. 7, after 6631 rounds the First Node Dead (FND) in NCOGA whereas 5853, 5446, 5163, and 4180 rounds FND in GATERP, BMHGA, GAOC, and GASONeC, respectively and after 12948 rounds Half Nodes Dead (HND) in NCOGA but in just 11579, 11035, 9882, and 8988 rounds as per the BMHGA [60], GATERP [52], GAOC [18], and GASONeC [48] protocols separately. Moreover, in enhancement of last node dead (LND) that is, also referred to as the network lifetime, likewise stated in NCOGA covering 26500 rounds whereas BMHGA, GATERP, GAOC, and GASONeC covers 24140, 23996, 22621, and 19595 rounds, respectively. In the above analysis, NCOGA is able to complete more rounds at various phases of dead nodes related to reduced energy consumption during intra-cluster communication and CH selection. Enhancement has been described while after the CH selection has been optimized in accordance with several factors, higher energy conservation is obtained in NCOGA in comparison with other protocols, separately.

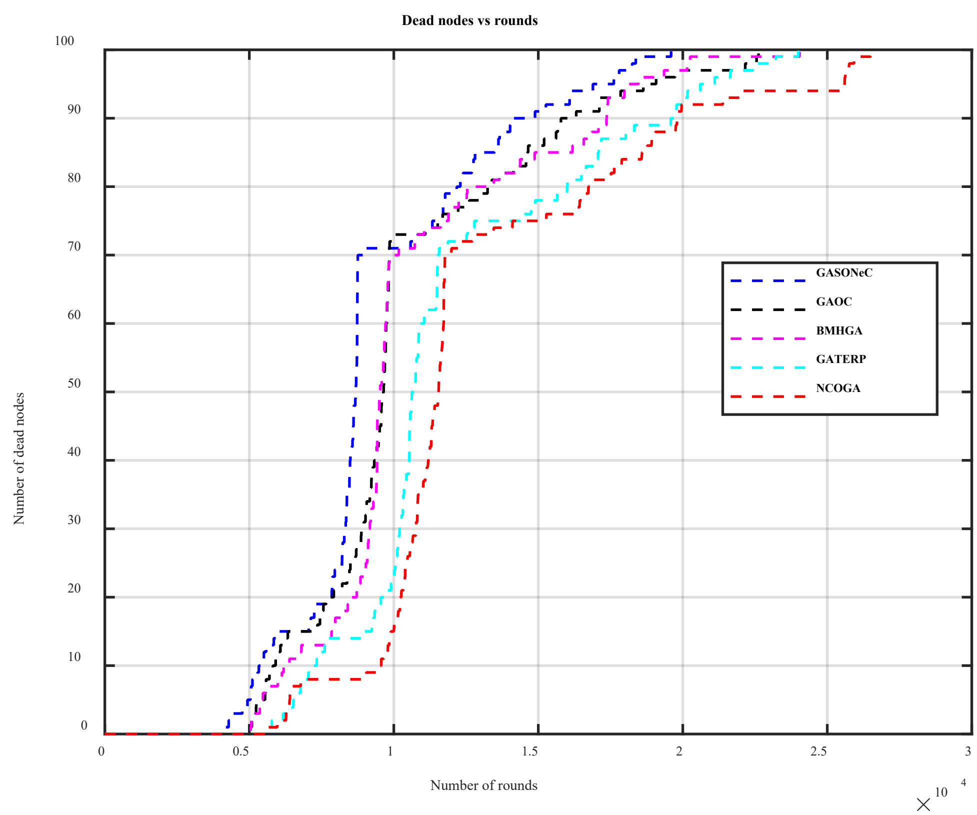

Fig. 7. Comparative analysis of dead nodes vs rounds of NCOGA with other protocols.

4.4.4 Network's remaining energy

It can be shown that the NCOGA protocol assumes a reduction of network energy based on data transfer. The performance of networks residual energy through the upsurge into the number of rounds as per the observation. NCOGA achieves better as compared GATERP [53], BMHGA [60], GAOC [18], and GASONeC [49] protocols, correspondingly in a manner that it comprises a larger number of rounds whereas the data transmission is improvement such as presented in Fig. 8. Moreover, in dual hop communication the energy consumption of a node in each round for NCOGA is less over in order to compare protocols. NCOGA performs better than GATERP and GAOC because the optimum selection of CH helps maintain the energy of nodes. Due to the distance and CMD factors, inclusion holds their significance. The CH selection of a node is governed through the distance factor, which supports the selection of the nearest node towards the sink.

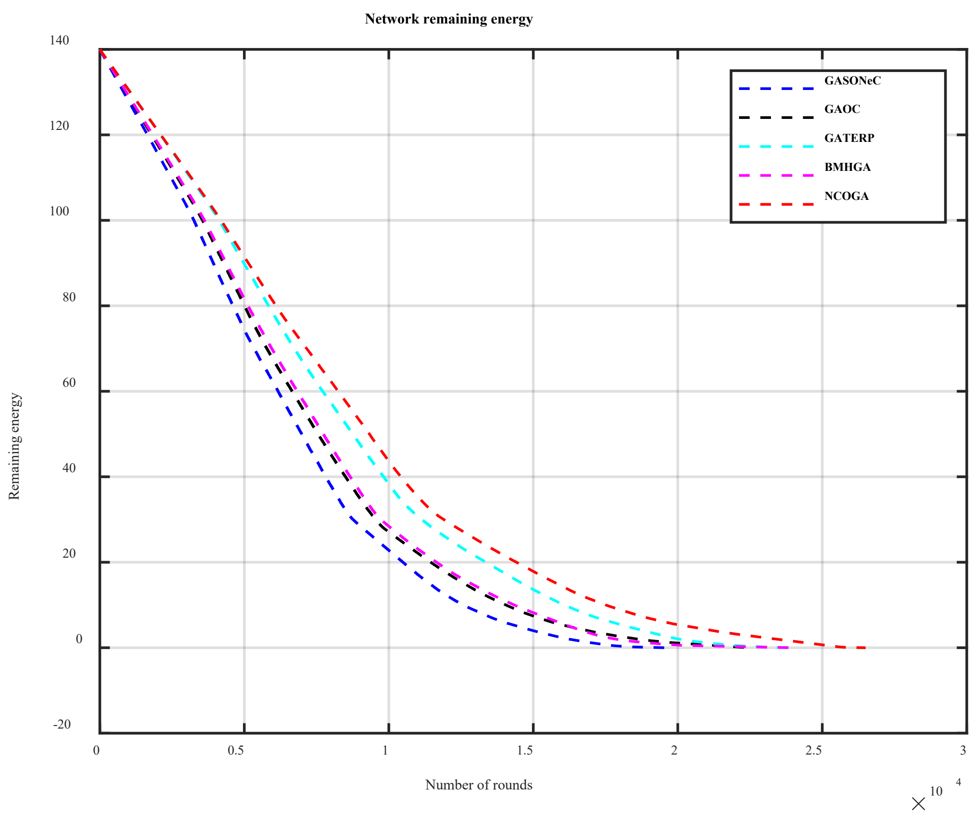

Fig. 8. Comparative analysis of Network's remaining energy of NCOGA with other protocols.

4.4.5 Throughput

As shown in Fig. 9, throughput is improved systematically due to the fact that it effectively transmits 746390 data packets for NCOGA while GATERP [52], BMHGA [60], GAOC [18], and GASONeC [48] transmit 630032, 568458, 532837, and 486432 data packets, correspondingly. It has been observed that, comparison of throughput, NCOGA enhances throughput by 18.46%, 31.3%, 40.07%, and 53.44% as compared to GATERP, BMHGA, GAOC, and GASONeC protocols, separately. Throughput has been improved gigantically since during the transmission of data packet successfully forwarded due to reduction of loss reported and with a choice of enhanced CH in proposed protocol.

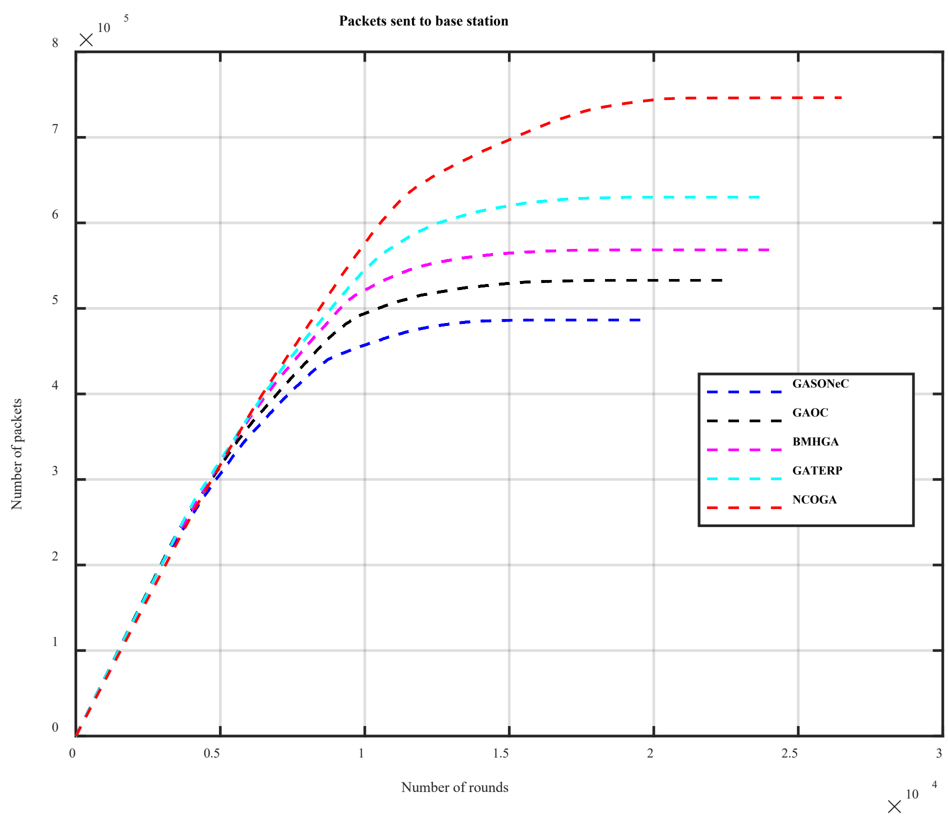

Fig. 9. Comparative analysis of throughput of NCOGA with other protocols.

4.4.6 Convergence Analysis

Fig. 10 shows the comparison of the convergence rate of NCOGA with BMHGA, GAOC, GATERP, and GASONeC in terms of the fitness function. To perform this analysis, the network size was kept being 100×100 m^2, the number of nodes was 100, and the number of rounds was 800. It can be observed that the convergence of the plot is gradual which is attributed to the fact that the GA variant used in our approach performs better search as compared to the other existing GA variants. It is noteworthy from the plot here that NCOGA converges faster as compared to BMHGA, GAOC, GATERP, and GASONeC.

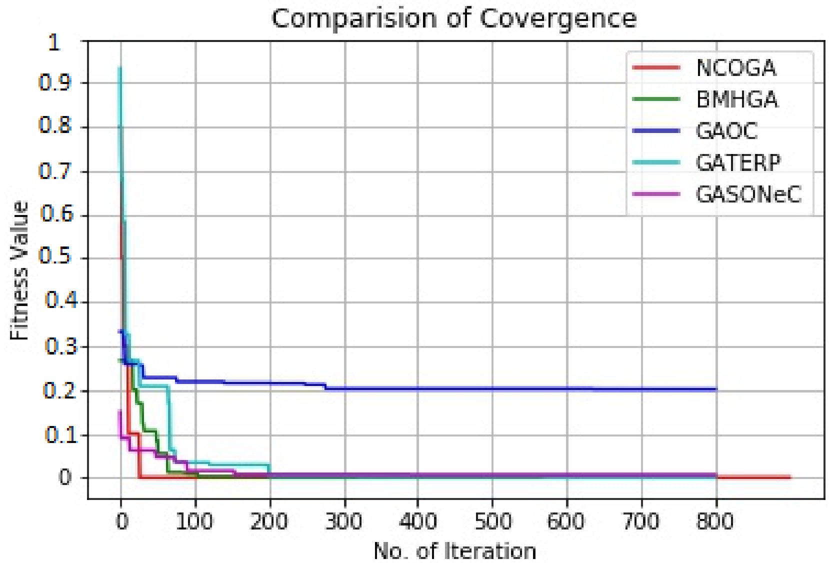

Fig. 10. Convergence analysis of NCOGA with other protocols

In a nutshell, Evaluation performance summary is improvement stated by NCOGA is summarized in Fig. 11. The performance of NCOGA indicates that, it much better than other protocols as per the comparative analysis in terms of different performance metrics.

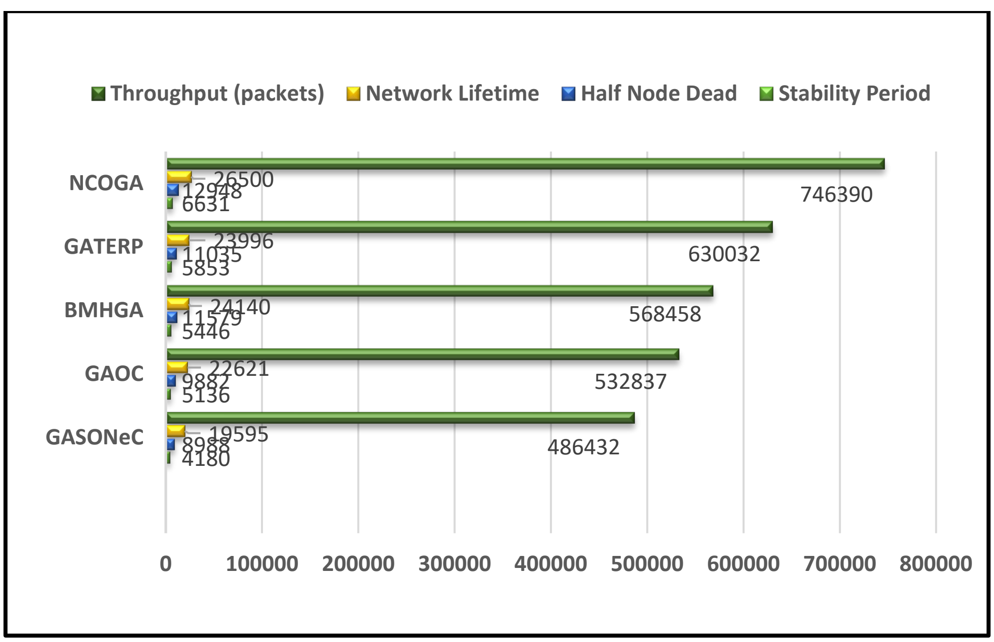

Fig. 11. Comparative analysis of NCOGA with other protocols in different metrics

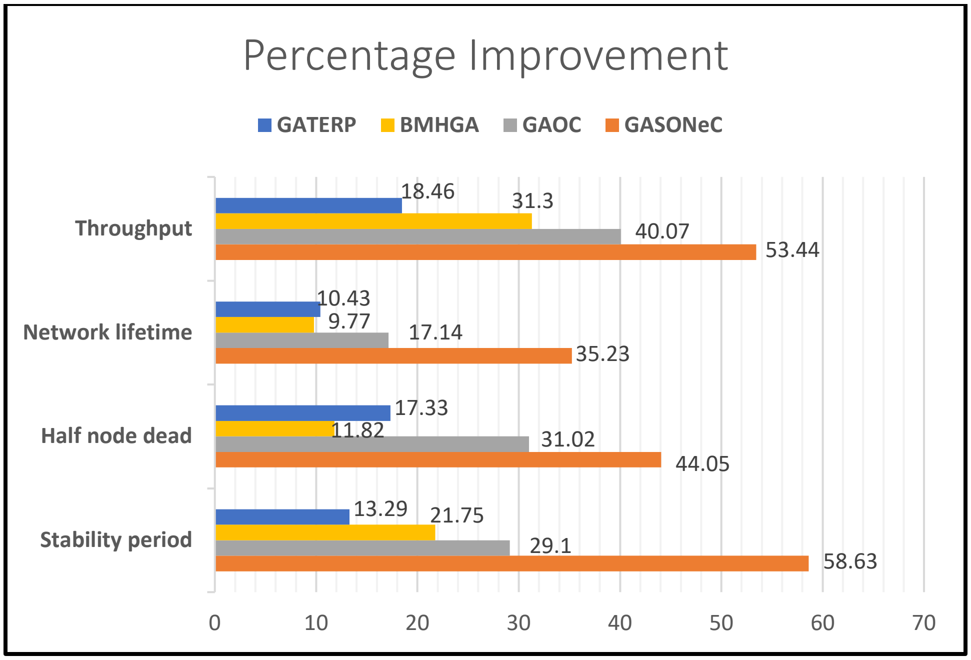

Fig. 12. Percentage improvement by the NCOGA against the other protocols.

The percent improvement in accordance with the NCOGA in the sense of stability period, HND, Network Lifetime and Throughput are provided in the Fig 12.

4.5 Statistical analysis

Statistical tests have been performed to determine performance significance of the NCOGA. F-test is conducted on the collected sample at 5% level of confidence. Throughput and remaining energy have been considered to perform statistical analysis. Equation (29) presents the null (H_0) and assumed (H_A) hypothesis.

$$\begin{aligned} H_0 &: \mu_{NCOGA} = \mu_{BMHGA} = \mu_{GAOC} = \mu_{GATERP} = \mu_{GASONeC} \\ H_A &: \mu_{NCOGA} \neq \mu_{BMHGA} \neq \mu_{GAOC} \neq \mu_{GATERP} \neq \mu_{GASONeC} \end{aligned} \tag{29}$$

Table 4 and 5 present descriptive statistics respectively for remaining energy and throughput with respect to the algorithms. This result statistics were achieved on 30 samples were collected from each of the algorithm. Result reveals that NCOGA has outperformed all the other algorithms such as GATERP [52], BMHGA [60], GAOC [18] and GASONeC [48]. ANOVA test was conducted for both remaining energy and throughput. Result of ANOVA test showed that the p-value is less than 0.05 (($p = 0.000 < 0.05$) respectively for remaining energy and throughput. Hence, from equation (29), H_0 is rejected. Further, this result concludes that one of sample is better than the other one. But looking the results it is hard to say which algorithms are responsible for the difference in the sample. Therefore, multiple comparisons or posthoc test has been conducted to compare the individual group or pair results.

Table 4. Descriptive analysis of remaining energy with respect to algorithms.

Algorithm	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
GATERP	30	-0.00694	0.00122	0.00022	-0.00739	-0.00648	-0.00896	-0.00493
BMHGA	30	-0.00728	0.00124	0.00022	-0.00774	-0.00681	-0.00933	-0.00524
NCOGA	**30**	**-0.00647**	**0.00110**	**0.00020**	**-0.00689**	**-0.00606**	**-0.00831**	**-0.00465**
GAOC	30	-0.00730	0.00134	0.00024	-0.00780	-0.00680	-0.00952	-0.00509
GASONeC	30	-0.00953	0.00150	0.00027	-0.01009	-0.00897	-0.01202	-0.00705
Total	150	-0.00750	0.00165	0.00013	-0.00777	-0.00724	-0.01202	-0.00465

Table 5. Descriptive analysis of throughput with respect to algorithms.

Algorithm	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
GATERP	30	630032.00	0.00	0.00	630032.00	630032.00	630032.00	630032.00
BMHGA	30	568457.56	0.50	0.09	568457.37	568457.75	568457.00	568458.00
NCOGA	**30**	**746390.00**	**0.00**	**0.00**	**746390.00**	**746390.00**	**746390.00**	**746390.00**
GAOC	30	532837.00	0.00	0.00	532837.00	532837.00	532837.00	532837.00
GASONeC	30	486432.00	0.00	0.00	486432.00	486432.00	486432.00	486432.00
Total	150	592829.71	90277.39	7371.11	578264.28	607395.14	486432.00	746390.00

TukeyHSD (Tukey–Kramer honestly significant difference) test has been conducted. It is less liable to type-1 error and well suited for equal samples (in our case N = 30). Table 6 and 7 show the result of TukeyHSD respectively variable remaining energy and throughput. It is noted from Table 6 that NCOGA showed significantly better performance as compared to GASONeC for remaining energy. On the other hand, NCOGA's performance is significantly better than GATERP, BMHGA, GAOC and GASONeC for throughput as shown in Table 7.

Table 6. Multiple comparisons test (Posthoc test) with respect to algorithms (dependent variable: remaining energy).

(I) Algorithms	(J) Algorithms	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
GATERP	BMHGA	0.00034	.00033	0.845	-0.00058	0.00126
	NCOGA	-0.00046	.00033	0.633	-0.00138	0.00045
	GAOC	0.00036	.00033	0.810	-0.00055	0.00128
	GASONeC	0.00259*	.00033	0.000	0.00166	0.00351
BMHGA	GATERP	-0.00034	.00033	0.845	-0.00126	0.00058
	NCOGA	-0.00080	.00033	0.117	-0.00172	0.00011

	GAOC	0.00002	.00033	1.000	-0.00089	0.00094
	GASONeC	0.00225*	.00033	0.000	0.00132	0.00317
NCOGA	**GATERP**	**0.00046**	**.00033**	**0.633**	**-0.00045**	**0.00138**
	BMHGA	**0.00080**	**.00033**	**0.117**	**-0.00011**	**0.00172**
	GAOC	**0.00082**	**.00033**	**0.100**	**-0.00009**	**0.00175**
	GASONeC	**0.00305***	**.00033**	**0.000**	**0.00213**	**0.00397**
GAOC	GATERP	-0.00036	.00033	0.810	-0.00128	0.00055
	BMHGA	-0.00002	.00033	1.000	-0.00094	0.00089
	NCOGA	-0.00082	.00033	0.100	-0.00175	0.00009
	GASONeC	0.00222*	.00033	0.000	0.00130	0.00314
GASONeC	GATERP	-0.00259*	.00033	0.000	-0.00351	-0.00166
	BMHGA	-0.00225*	.00033	0.000	-0.00317	-0.00132
	NCOGA	-0.00305*	.00033	0.000	-0.00397	-0.00213
	GAOC	-0.00222*	.00033	0.000	-0.00314	-0.00130

*. The mean difference is significant at the 0.05 level.

Table 7. Multiple comparisons test (Posthoc test) with respect to algorithms (dependent variable: throughput).

(I) Algorithms	(J) Algorithms	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
GATERP	BMHGA	61574.43*	0.058	0.00	61574.27	61574.59
	NCOGA	-116358.00*	0.058	0.00	-116358.16	-116357.83
	GAOC	97195.00*	0.058	0.00	97194.83	97195.16
	GASONeC	143600.00*	0.058	0.00	143599.83	143600.16
BMHGA	GATERP	-61574.43*	0.058	0.00	-61574.59	-61574.27
	NCOGA	-177932.43*	0.058	0.00	-177932.59	-177932.27
	GAOC	35620.56*	0.058	0.00	35620.40	35620.72
	GASONeC	82025.56*	0.058	0.00	82025.40	82025.72
NCOGA	**GATERP**	**116358.00***	**0.058**	**0.00**	**116357.83**	**116358.16**
	BMHGA	**177932.43***	**0.058**	**0.00**	**177932.27**	**177932.59**
	GAOC	**213553.00***	**0.058**	**0.00**	**213552.83**	**213553.16**
	GASONeC	**259958.00***	**0.058**	**0.00**	**259957.83**	**259958.16**
GAOC	GATERP	-97195.00*	0.058	0.00	-97195.16	-97194.83
	BMHGA	-35620.56*	0.058	0.00	-35620.72	-35620.40
	NCOGA	-213553.00*	0.058	0.00	-213553.16	-213552.83
	GASONeC	46405.00*	0.058	0.00	46404.83	46405.16
GASONeC	GATERP	-143600.00*	0.058	0.00	-143600.16	-143599.83
	BMHGA	-82025.56*	0.058	0.00	-82025.72	-82025.40
	NCOGA	-259958.00*	0.058	0.00	-259958.16	-259957.83
	GAOC	-46405.00*	0.058	0.00	-46405.1608	-46404.83

*. The mean difference is significant at the 0.05 level.

Table 8. TukeyHSD homogeneous subsets test with respect to algorithms (dependent variable: remaining energy).

Algorithms	N	Subset for alpha = 0.05	
		1	2
GASONeC	30	-0.00953	
GAOC	30		-0.0073
BMHGA	30		-0.0072
GATERP	30		-0.0069
NCOGA	**30**		**-0.0064**
Sig.		1.000	.100

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 30.000.

Table 9. TukeyHSD homogeneous subsets test with respect to algorithms (dependent variable: throughput).

Algorithms	N	Subset for alpha = 0.05				
		1	2	3	4	5
GASONeC	30	486432.00				
GAOC	30		532837.00			
BMHGA	30			568457.56		
GATERP	30				630032.00	
NCOGA	**30**					**746390.00**
Sig.		1.00	1.00	1.00	1.00	1.00

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 30.000.

TukeyHSD homogeneity test is conducted for remaining energy and throughput to verify the performance similarity of the algorithms. The means of algorithms in homogenous subsets are presented in Table 8 and 9 respectively for remaining energy and throughput. Table 8 shows that performance of NCOGA – GASONeC is significantly different while other pair of algorithms (NCOGA – GAOC, NCOGA - BMHGA and NCOGA – GATERP) showed almost similar results where NCOGA's performance is better. Table 9 reveals the homogeneity test results for throughput. We can see that the performance of the proposed NCOGA is significantly better than the other algorithms. In other words, we can say that NCOGA has outperformed the other algorithms.

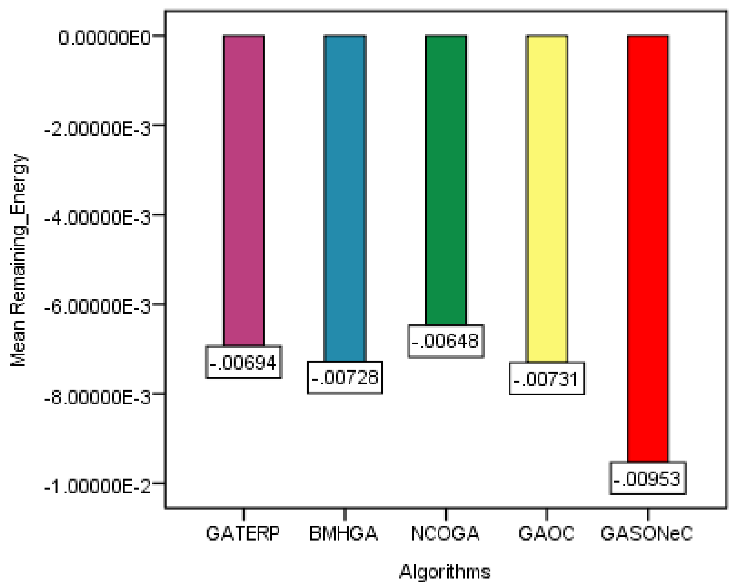

Fig. 13. Estimated marginal mean plot for remaining energy with respect to each algorithm.

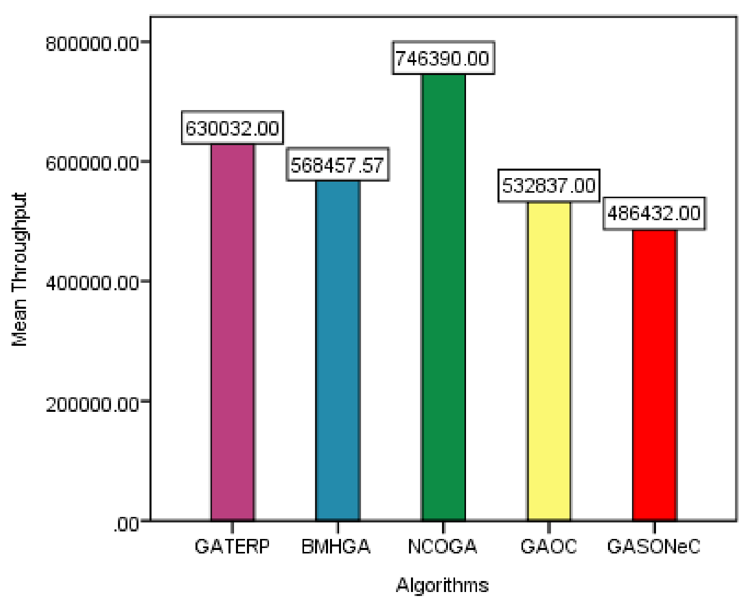

Fig. 14. Estimated marginal mean plot for throughput with respect to each algorithm.

The estimated marginal mean plots are presented in Figure 13 and 14, respectively, for remaining energy and throughput. The graphical result reveals that the NOCGA has demonstrated significantly better performance over the other algorithms. This result also indicates that the GATERP [52] showed fairly good results, GAOC [18] and BMHGA [60] average performance, while GASONeC's [48] reported worst results.

5 Conclusions

In this paper, NCOGA has been proposed that integrates objective function by considering six parameters viz. residual energy, initial energy, distance to the sink, number of neighbors surrounded by a node, load balancing factor and communicating mode decider (CMD). The fitness parameters have been mathematically modelled and integrated to generate a fitness function. The performance of NCOGA has been evaluated on the basis of different performance metrics. It has been found that NCOGA elongates stability period by 21.75%, 13.29%, 29.1%, and 58.63% and network lifetime by 9.77%, 10.43%, 17.14%, and 35.23% as compared to BMHGA, GATERP, GAOC, and GASONeC, respectively. The underlying reason of the enhancement in the stability period is due to the energy preservation attributed to the CMD and load balancing factors. These factors not only helped in making dual hop communication energy efficient but also balances the load distribution over the clusters. Statistical tests have also been conducted which reveals the superiority of the proposed NCOGA over the other state-of-the-art methods. We observed that NCOGA converges faster as compared to existing methods. Furthermore, the multiple sinks which should be used, is another crucial concern which seeks optimization. In the future, this work can be extended to introduce the mobility of sink in the network that will improve its Quality of Service (QoS) parameters.

References

1. Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., & Cayirci, E. (2002). Wireless sensor networks: a survey. *Computer networks*, *38*(4), 393–422.
2. 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* (pp. 719–724). IEEE. Retrieved from http://ieeexplore.ieee.org/abstract/document/1467103/
3. Verma, S., Sood, N., & Sharma, A. K. (2018). Design of a novel routing architecture for harsh environment monitoring in heterogeneous WSN. *IET Wireless Sensor Systems*.
4. Pant, D., Verma, S., & Dhuliya, P. (2017). A study on disaster detection and management using WSN in Himalayan region of Uttarakhand. In *2017 3rd International conference on advances in computing, communication & automation (ICACCA)(Fall)* (pp. 1–6). IEEE.

5. Pottie, G. J. (1998). Wireless sensor networks. In *Information Theory Workshop, 1998* (pp. 139–140). IEEE.
6. Akkaya, K., & Younis, M. (2005). A survey on routing protocols for wireless sensor networks. *Ad hoc networks*, *3*(3), 325–349.
7. Abbasi, A. A., & Younis, M. (2007). A survey on clustering algorithms for wireless sensor networks. *Computer communications*, *30*(14), 2826–2841.
8. Nguyen, T.-T., Pan, J.-S., Dao, T.-K., & Chu, S.-C. (2018). Load balancing for mitigating hotspot problem in wireless sensor network based on enhanced diversity pollen. *Journal of Information and Telecommunication*, *2*(1), 91–106.
9. Singh, S. K., Kumar, P., & Singh, J. P. (2018). An Energy Efficient Protocol to Mitigate Hot Spot Problem Using Unequal Clustering in WSN. *Wireless Personal Communications*, *101*(2), 799–827.
10. Bara'a, A. A., & Khalil, E. A. (2012). A new evolutionary based routing protocol for clustered heterogeneous wireless sensor networks. *Applied Soft Computing*, *12*(7), 1950–1957.
11. Saleem, M., Di Caro, G. A., & Farooq, M. (2011). Swarm intelligence based routing protocol for wireless sensor networks: Survey and future directions. *Information Sciences*, *181*(20), 4597–4624.
12. Kulkarni, R. V., Forster, A., & Venayagamoorthy, G. K. (2011). Computational intelligence in wireless sensor networks: A survey. *IEEE communications surveys & tutorials*, *13*(1), 68–96.
13. Nanda, S. J., & Panda, G. (2014). A survey on nature inspired metaheuristic algorithms for partitional clustering. *Swarm and Evolutionary computation*, *16*, 1–18.
14. Bhushan, S., Pal, R., & Antoshchuk, S. G. (2018). Energy Efficient Clustering Protocol for Heterogeneous Wireless Sensor Network: A Hybrid Approach Using GA and K-means. In *2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP)* (pp. 381–385). IEEE.
15. Elhoseny, M., Farouk, A., Zhou, N., Wang, M.-M., Abdalla, S., & Batle, J. (2017). Dynamic multi-hop clustering in a wireless sensor network: Performance improvement. *Wireless Personal Communications*, *95*(4), 3733–3753.

16. Bhatia, T., Kansal, S., Goel, S., & Verma, A. K. (2016). A genetic algorithm based distance-aware routing protocol for wireless sensor networks. *Computers & Electrical Engineering*, *56*, 441–455.
17. 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.
18. Verma S, Sood N, Sharma AK (2019) Genetic Algorithm-based Optimized Cluster Head selection for single and multiple data sinks in Heterogeneous Wireless Sensor Network. Applied Soft Computing 105788.
19. Tyagi, S., & Kumar, N. (2013). A systematic review on clustering and routing techniques based upon LEACH protocol for wireless sensor networks. *Journal of Network and Computer Applications*, *36*(2), 623–645.
20. Tanwar, S., Kumar, N., & Rodrigues, J. J. (2015). A systematic review on heterogeneous routing protocols for wireless sensor network. *Journal of network and computer applications*, *53*, 39–56.
21. Smaragdakis, G., Matta, I., & Bestavros, A. (2004). *SEP: A stable election protocol for clustered heterogeneous wireless sensor networks*. Boston University Computer Science Department. Retrieved from http://open.bu.edu/handle/2144/1548
22. Qing, L., Zhu, Q., & Wang, M. (2006). Design of a distributed energy-efficient clustering algorithm for heterogeneous wireless sensor networks. *Computer communications*, *29*(12), 2230–2237.
23. Elbhiri, B., Saadane, R., Aboutajdine, D., & others. (2010). Developed Distributed Energy-Efficient Clustering (DDEEC) for heterogeneous wireless sensor networks. In *I/V Communications and Mobile Network (ISVC), 2010 5th International Symposium on* (pp. 1–4). IEEE. Retrieved from http://ieeexplore.ieee.org/abstract/document/5656252/
24. Kumar, D., Aseri, T. C., & Patel, R. B. (2009). EEHC: Energy efficient heterogeneous clustered scheme for wireless sensor networks. *Computer Communications*, *32*(4), 662–667.
25. Javaid, N., Qureshi, T. N., Khan, A. H., Iqbal, A., Akhtar, E., & Ishfaq, M. (2013). EDDEEC: Enhanced developed distributed energy-efficient clustering for heterogeneous wireless sensor networks. *Procedia Computer Science*, *19*, 914–919.

26. Qureshi, T. N., Javaid, N., Khan, A. H., Iqbal, A., Akhtar, E., & Ishfaq, M. (2013). BEENISH: Balanced energy efficient network integrated super heterogeneous protocol for wireless sensor networks. *Procedia Computer Science*, *19*, 920–925.
27. Akbar, M., Javaid, N., Imran, M., Amjad, N., Khan, M. I., & Guizani, M. (2016). Sink mobility aware energy-efficient network integrated super heterogeneous protocol for WSNs. *EURASIP Journal on Wireless Communications and Networking*, *2016*(1), 66.
28. Naranjo, P. G. V., Shojafar, M., Mostafaei, H., Pooranian, Z., & Baccarelli, E. (2017). P-SEP: A prolong stable election routing algorithm for energy-limited heterogeneous fog-supported wireless sensor networks. *The Journal of Supercomputing*, *73*(2), 733–755.
29. Mittal, N., & Singh, U. (2015). Distance-based residual energy-efficient stable election protocol for WSNs. *Arabian Journal for Science and Engineering*, *40*(6), 1637–1646.
30. Mittal, N., Singh, U., & Sohi, B. S. (2017). A stable energy efficient clustering protocol for wireless sensor networks. *Wireless Networks*, *23*(6), 1809–1821.
31. Hussain, S., Matin, A. W., & Islam, O. (2007). Genetic algorithm for hierarchical wireless sensor networks. *JNW*, *2*(5), 87–97.
32. Osamy, W., El-Sawy, A. A., & Khedr, A. M. (2020). Effective TDMA scheduling for tree-based data collection using genetic algorithm in wireless sensor networks. Peer-to-Peer Networking and Applications, 13(3), 796-815.
33. Shahzad, M. K., Islam, S. M., Hossain, M., Abdullah-Al-Wadud, M., Alamri, A., & Hussain, M. (2021). GAFOR: Genetic Algorithm Based Fuzzy Optimized Re-Clustering in Wireless Sensor Networks. Mathematics, 9(1), 43.
34. Chandirasekaran, D., & Jayabarathi, T. (2019). Cat swarm algorithm in wireless sensor networks for optimized cluster head selection: a real time approach. *Cluster Computing*, *22*(5), 11351–11361.
35. Chawra, V. K., & Gupta, G. P. (2020). Salp: Metaheuristic-Based Clustering for Wireless Sensor Networks. In *Nature-Inspired Computing Applications in Advanced Communication Networks* (pp. 41–56). IGI Global.
36. John, J., & Rodrigues, P. (2019). MOTCO: Multi-objective Taylor Crow Optimization Algorithm for Cluster Head Selection in Energy Aware Wireless Sensor Network. *Mobile Networks and Applications*, 1–17.

37. Lee, J.-G., Chim, S., & Park, H.-H. (2019). Energy-Efficient Cluster-Head Selection for Wireless Sensor Networks Using Sampling-Based Spider Monkey Optimization. *Sensors*, *19*(23), 5281.
38. Poluru, R. K., & Kumar R, L. (2019). An Improved Fruit Fly Optimization (IFFOA) based Cluster Head Selection Algorithm for Internet of Things. *International Journal of Computers and Applications*, 1–9.
39. Rambabu, B., Reddy, A. V., & Janakiraman, S. (2019). Hybrid Artificial Bee Colony and Monarchy Butterfly Optimization Algorithm (HABC-MBOA)-based Cluster Head Selection for WSNs. *Journal of King Saud University-Computer and Information Sciences*.
40. Vijayalakshmi, K., & Anandan, P. (2019). A multi objective Tabu particle swarm optimization for effective cluster head selection in WSN. *Cluster computing*, *22*(5), 12275–12282.
41. Sahoo, B. M., Amgoth, T., & Pandey, H. M. (2020). Particle Swarm Optimization Based Energy Efficient Clustering and Sink Mobility in Heterogeneous Wireless Sensor Network. *Ad Hoc Networks*, 102237.
42. 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.
43. Singh, V. K., & Sharma, V. (2012). Elitist genetic algorithm based energy efficient routing scheme for wireless sensor networks. *International Journal Of Advanced Smart Sensor Network Systems (IJASSN)*, *2*(2).
44. Kuila, P., Gupta, S. K., & Jana, P. K. (2013). A novel evolutionary approach for load balanced clustering problem for wireless sensor networks. *Swarm and Evolutionary Computation*, *12*, 48–56.
45. 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.
46. Elhoseny, M., Yuan, X., Yu, Z., Mao, C., El-Minir, H. K., & Riad, A. M. (2015). Balancing energy consumption in heterogeneous wireless sensor networks using genetic algorithm. *IEEE Communications Letters*, *19*(12), 2194–2197.

47. Shokouhifar, M., & Jalali, A. (2015). A new evolutionary based application specific routing protocol for clustered wireless sensor networks. *AEU-International Journal of Electronics and Communications*, *69*(1), 432–441.
48. Yuan, X., Elhoseny, M., El-Minir, H. K., & Riad, A. M. (2017). A genetic algorithm-based, dynamic clustering method towards improved WSN longevity. *Journal of Network and Systems Management*, *25*(1), 21–46.
49. Hamidouche, R., Aliouat, Z., & Gueroui, A. M. (2018). Genetic Algorithm for Improving the Lifetime and QoS of Wireless Sensor Networks. *Wireless Personal Communications*, *101*(4), 2313–2348.
50. Guo, Y., Cheng, J., Liu, H., Gong, D., & Xue, Y. (2017). A novel knowledge-guided evolutionary scheduling strategy for energy-efficient connected coverage optimization in WSNs. *Peer-to-Peer Networking and Applications*, *10*(3), 547–558.
51. Zhang, Y., Gong, D., Gao, X., Tian, T., & Sun, X. (2020). Binary differential evolution with self-learning for multi-objective feature selection. *Information Sciences*, *507*, 67–85.
52. Mittal, N., Singh, U., & Sohi, B. S. (2019). An energy-aware cluster-based stable protocol for wireless sensor networks. Neural Computing and Applications, 31(11), 7269-7286.
53. Sahoo, B. M., Pandey, H. M., & Amgoth, T. (2021). "A Whale Optimization (WOA): Meta-Heuristic based energy improvement Clustering in Wireless Sensor Networks," 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2021, pp. 649-654, doi: 10.1109/Confluence51648.2021.9377181.
54. Karthick, P. T., & Palanisamy, C. (2019). Optimized cluster head selection using krill herd algorithm for wireless sensor network. *Automatika*, *60*(3), 340–348.
55. Du, M., Ding, S., & Jia, H. (2016). Study on density peaks clustering based on k-nearest neighbors and principal component analysis. *Knowledge-Based Systems*, *99*, 135–145.
56. Sahoo, B. M., Pandey, H. M., & Amgoth, T. (2021). GAPSO-H: A hybrid approach towards optimizing the cluster based routing in wireless sensor network. Swarm and Evolutionary Computation, 60, 100772.
57. 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.

58. Wikaisuksakul, S. (2014). A multi-objective genetic algorithm with fuzzy c-means for automatic data clustering. *Applied Soft Computing*, *24*, 679-691.
59. Goldberg, D. E., & Holland, J. H. (1988). Genetic algorithms and machine learning. *Machine learning*, *3*(2), 95–99.
60. Li, J., Luo, Z., & Xiao, J. (2020). A Hybrid Genetic Algorithm with Bidirectional Mutation for Maximizing Lifetime of Heterogeneous Wireless Sensor Networks. IEEE Access, 8, 72261-72274.