

Flower Pollination Algorithm–Based Cluster Head Optimization And Compression-Aware Data Transmission For Iot-Driven Wireless Sensor Networks

Kavita Agrawal¹, Dr. Shish Ahmad²

¹Department of Computer Science and Engineering Integral University, Lucknow, India
kavita@iul.ac.in.

²Department of Computer Science and Engineering Integral University, Lucknow,
Indiashish@iul.ac.in.

Manuscript Communication Number (MCN): IU/R&D/2026-MCN0004285

Abstract: Wireless Sensor Networks (WSNs) form a fundamental component of Internet of Things (IoT)–based environmental monitoring systems; however, their practical deployment is severely constrained by limited energy resources and excessive communication overhead. Although clustering-based protocols are widely used to enhance network lifetime, most existing solutions emphasize only optimal cluster head (CH) selection and rely on direct or inefficient data transmission strategies, leading to rapid energy depletion. Moreover, the large volume of time-series weather data generated by sensor nodes further intensifies energy consumption during transmission. To address these challenges, this paper proposes an integrated optimization framework that jointly considers energy-aware CH selection and compression-based data transmission. The Flower Pollination Algorithm (FPA), inspired by natural pollination behavior, is employed to select an optimal set of CHs based on multiple network parameters, including residual energy, node distribution, convergence, divergence, centrality, and distance to the Base Station (BS). Under the stage of data aggregation at the CHs, simple, fast but efficient compression methods—known as ASWDR and Zstandard (ZSTD)—are applied. This data compression stage is used for reducing the size of data before the transmission from CHs to the BS. The results generated under simulation model are used to demonstrate the proposed FPA-based CHs selection associated with data compression reduce energy consumption significantly. Hence it introduces the performance improvement in terms of load balancing, network stability, and prolonging the lifetime of network compared to conventional and other optimization-based protocols for sustainable and energy-efficient WSN-based IoT applications.

Keywords: Wireless sensor network, IoT, Data compression, Optimization, Clustering

I. INTRODUCTION

Modern application based on IoT-enabled platforms significantly transforms the information technology in various domains by data extraction and monitoring from sensor data. These IoT based systems are supporting a wide range of applications, like agriculture, data analytics, weather forecasting, crop yield. They have reduced operational costs by reduction of unnecessary use of fertilizers and herbicides [1, 2]. The IoT has emerged from the interconnection of lot of devices that communicate in real time and exchanging the information on Internet. These communicating devices are electronic chips, sensors, mechanical components, and radio frequency identification (RFID) tags [3, 4].

They are integrated with sensors for providing an application-based data. Interaction of these devices creates a Wireless Sensor Network (WSN) inherently, which works as the backbone for serving IoT applications [3]. WSN is consisting of a large number of sensor nodes deployed in a region of interest randomly or in planned manner. The primary focus behind the use of these sensor nodes is environmental monitor by transmitting the recorded data to a Base Station (BS) [4]. But the WSNs face critical challenges due the limited operational lifetime of nodes due to limited capacity battery [5]. Clustering-based protocols have widely explored to address such issues. Clustering protocols for data

transmission is effective in prolonging the lifetime of the network [6–9]. In clustering based data propagation architecture, nodes are divided into groups. Each group is associated with a cluster head (CH), and other nodes are called as cluster members [10–12]. The CHs are performing data aggregation and forwarding to the BS/ Due to this extra task they consume higher energy than the cluster members [13–15]. As a result, the selection of CH requires critical consideration of multiple nodes with optimal characteristics. These characteristics reflect the network like residual energy, spatial location, node density, and distance to the BS. Decision using collection using these parameters helps to achieve efficient and sustainable performance of the network [9]. Several research methods are proposed for the optimal selection of CHs in WSN. Diverse methodologies are employed that vary from simple fitness-value, weighted approach, and rule-based scheme to highly advanced optimization mechanism [16, 17]. However, in WSNs achievement of high efficiency is not limited only with selection of CHs. It also depends critically on the technical advancement in data aggregation and transmission stage. In the clustering protocols the CHs follows single-hop or direct communication to BS. It causes substantial large energy consumption compared to multi-hop communication schemes [18–20]. In this way reduction of energy usage under the phase of data transmission exhibits a significant concern for modern research.

At the phase after data aggregation and before data forwarding to BS, efficient management of large size data collected from sensor at CH introduces a significant challenge [5]. The large volumes sensed data transmission is leading towards outcome of increasing energy consumption and high communication bandwidth utilization. This can be mitigated by applying data compression methods that helps to reduction of storage requirements and overhead on communication phase [6]. Lot of compression algorithms are developed for addressing application-specific requirements, with many schemes exhibit the lossy compression mechanism. These lossy compression schemes significantly reducing data size, but introduce errors that are often introduces difficulties [7]. To overcome such difficulties, several recent proposed methods incorporate isolation and quantification of the errors associated with lossy compression schemes [8–10].

In the field of big data analytics, time-series weather data has high significance. It is very useful for capturing the temporal changes in environmental conditions depending on temperature, pressure, humidity, rainfall, solar radiation, and wind speed. It is essential for climate analysis and long-term forecasting. Despite the importance of such applications, there is a considerable lack of research that works for integrating optimization algorithm-based CH selection with data compression of data aggregated before the transmission to the BS. Although these lossy compression methods are crucial in efficient storing of weather data but existing research works are limited in relevance. For example, the work implementation in [9] incorporated compression using only temperature and humidity data, leaving the weather parameters unexplored (Figure 1). In this paper, an optimization scheme proposed for selection of best set of the cluster heads followed with data compression after aggregation stage for efficient transmission of weather data in WSN platform based on two different compression algorithms. The evaluation of lifetime, energy consumption and amount of data transmitted during the complete network lifetime is performed to evaluate enhancement. The objective is to utilize the featured information of network as parameters to define fitness of CH and compression schemes that ensure minimum energy consumption during sensor node to CH (S2CH) the data transmission and CH to BS (CH2BS).

A. Motivation

Clustering-based routing protocols in WSN are widely used for improving the energy efficiency. While some of the existing clustering schemes are focusing on the optimal

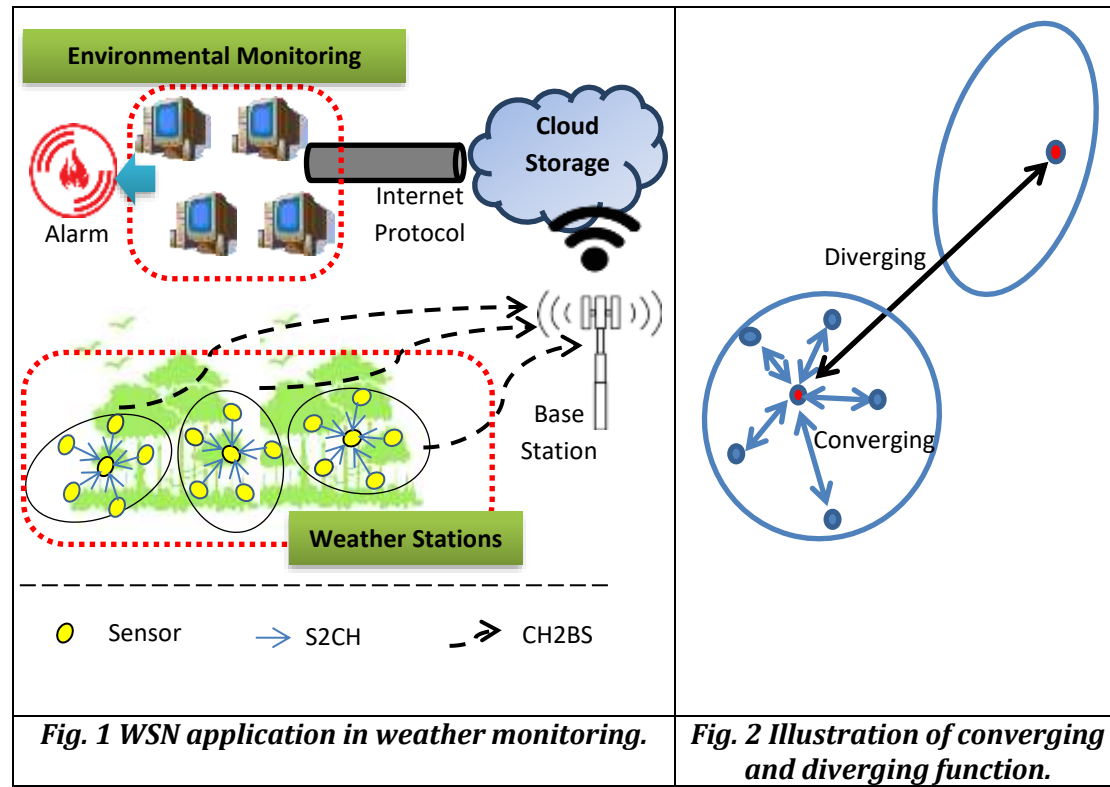
selection of CHs, some of them are emphasizing on efficient data transmission process management between CHs and the server. In this regard, replacement of direct communication under single hop transmission between nodes to the BS to multi hop scheme found to be very effective strategy for saving the energy. The research works that is going on to achieve an efficient routing have associated smart mechanism like machine learning (ML) and swarm intelligence algorithms. However, sensor nodes are inherently resource-constrained, that makes the application of machine learning-based schemes unsuitable for applying in real-time conditions. ML has basic requirement of extensive training process that incorporates substantial computational task, and, sometimes associates specialized hardware like GPUs. Consequently, their use of ML becomes impractical for sensor nodes devices with low power resources and simple hardware. Similarly, the optimization-based clustering for data routing is computationally simpler, but most of the existing algorithms depends only on CHs for planning the data forwarding route. The CHs are already consuming large energy considerable during the collection and aggregation of discrete data packets from the nodes inside respective cluster. Hence additional assignment of the responsibility to CH as an intermediate relaying node increases the extra energy consumption. In such cases, CHs is receiving large aggregated data packets from member nodes and forward them to the BS, it leads to instant energy depletion in high amount at the CHs. Therefore, it is very significant to enhance the energy efficiency of network by optimization of the data transmission in a planned way that is avoiding the excessive utilization of resource and preventing overburden on CHs. The work proposed in this article is motivated by this observation. It formulates the process of selection of CHs. The routing path paths are updated to best combination of group of CHs using optimization algorithms. It is offering a resource-efficient alternative for cluttering protocols that maintains an effective routing performance.

B. Contribution

The paper presented a research work that is addressing the issue of excessive consumption of energy during transmission of data from CHs to the BS by introduction of a multi-hop mechanism routing. This strategy of routing scheme is involving the selection of optimal nodes that forward data packets toward the BS with least consumption of energy. This objective is achieved by the utilization of an optimization framework integrated with data compression schemes. The Flower pollination algorithm (FPA), based by the pollination behavior of seeds, is employed for selection of CHs to get efficient solution of data forwarding while gathering data prior to forwarding towards BS. This metaheuristic approach is providing a resilient, robust and adaptive solution to suppress the impact of challenges observed in WSN, like energy-efficient routing, data aggregation, and overall optimization of the system. The FPA algorithm is incorporating various network parameters for finding the fitness of set of CHs. Strong capability of FPA is avoiding local optima and adapting to the dynamic selection of solution in WSNs. The algorithm is systematically exploring neighboring solutions while maintains a history of previous solutions for preventing the revisiting of previously explored possibilities. Results obtained from simulation model are indicating that the residual energy of the network has decreased marginally in comparison to existing approaches. It is reflecting load balancing and uniform energy consumption in sensor nodes. Finally, the proposed protocol helped in improving network stability and gives a long lifetime than other protocols.

The paper is organized as follows, in the section 2 the literature review of recent works related to optimization theory based WSN schemes are covered. It is providing a systematic comparative analysis of approaches proposed in recent time. Section 3 is presenting the simulation model in detail along with the description of the proposed

protocol. Section 4 has discussed the results and respective performance evaluation, and Section 5 covers the conclusion drawn from the paper with and future directions.



II. RELATED WORK

The rapid growth in the application of IoT and WSNs is intensifying on the utilization of energy-efficient communication and management of data due to the limited power, processing capability, and bandwidth associated with sensor devices. For addressing these constraints, a lot of research work is focusing on intelligent routing mechanism, optimization methods, and data compression algorithm developed for prolonging lifetime of network, improvement of QoS, and enhancement of the scalability.

A. Routing-Based Approaches

Modern routing methods basically employing the biologically or nature inspired algorithms for optimizing the method of cluster head selection and planning the routing paths. Cai et al. [1] and Cui et al. [2] worked for enhancing the conventional method LEACH protocol for improvement and modification in Bat Algorithms, to achieve better balance of energy and lifetime of network, but it increases the computational overhead. PSO-based routing proposed by Ghawry et al. [7] showing a fast converging and enhanced delivery of data packet, but suffering from premature convergence in dynamically changing network conditions. Sparrow Search Algorithm associated routing mechanism proposed by Kathirolu and Selvadurai [8] shows improvement in the energy efficiency through optimum searching, but sensitivity towards parameter affecting the robustness. Hybrid routing related approach further improved the throughput and utilization of energy but introducing control overhead [9]. Khan et al. [10] addressed mobility-aware routing in FANETs that used the nature-inspired techniques. Ma et al. [13] worked to optimize hierarchical APTEEN routing by applying Ant Colony Optimization for reducing redundant transmissions. Rani and Sharma [17] applied the combination of nature-

inspired routing with optical communication theory to achieve better data rates at the cost of deployment complexity.

B. Routing-Based Approaches

Optimization-based research extends beyond routing to include node deployment, green communication, and resource management. Dai et al. [3] optimized node placement in industrial IoT networks to improve coverage efficiency, though long-term energy effects were not examined. Harris Hawks Optimization was employed by Dev et al. [4] to reduce communication energy, yielding significant savings but relying on centralized computation. Mnasri and Alrashidi [14] introduced a lightweight optimizer for energy-efficient routing with faster convergence, while Bat Algorithm variants were used to improve cluster head selection and routing stability in IoT and NoC environments [15], [16]. Hybrid optimization frameworks such as GREPHRO and WAOA demonstrated improved convergence and robustness but increased computational complexity [11], [12]. Improved Ant Colony Optimization further enhanced load balancing and energy efficiency at the expense of higher control traffic [20].

C. Compression-Based Approaches

Data compression techniques focus on reducing communication energy by minimizing transmitted data volume. Dhou and Cruzen [5] proposed biologically inspired chain coding for IoT data compression, achieving energy savings with added processing complexity. Piątkowski et al. [21] experimentally showed that moderate compression significantly reduces transmission energy in microcontroller-based systems. Wavelet with optimization algorithm assisted compression methods applied for IoMT and IoT environments for balancing the compression efficiency and fidelity of data [22]. Learning mechanism-based compression and sparse reconstruction methods helped in reducing data size while keeping the accuracy on reconstruction, though the issues of increased computation and memory usage are observed [23], [25]. TinyML associated adaptive compression introduced adaptability for dynamic data patterns but incorporated the training overhead [26]. Simple and light data reduction techniques are giving efficient transmission with low amount of complexity, but they face the issue of adaptability on heterogeneous type data [27], [28]. Overall, the use of routing, optimization, and compression techniques is independently demonstrating improvements in efficiency and performance in IoT applications. However, the increasing computation overhead, centralized processing, and absence of integrated mechanism is the key limitations. The need of unified, lightweight, and scalable frameworks is still desired that can collectively address routing, optimization, and data compression.

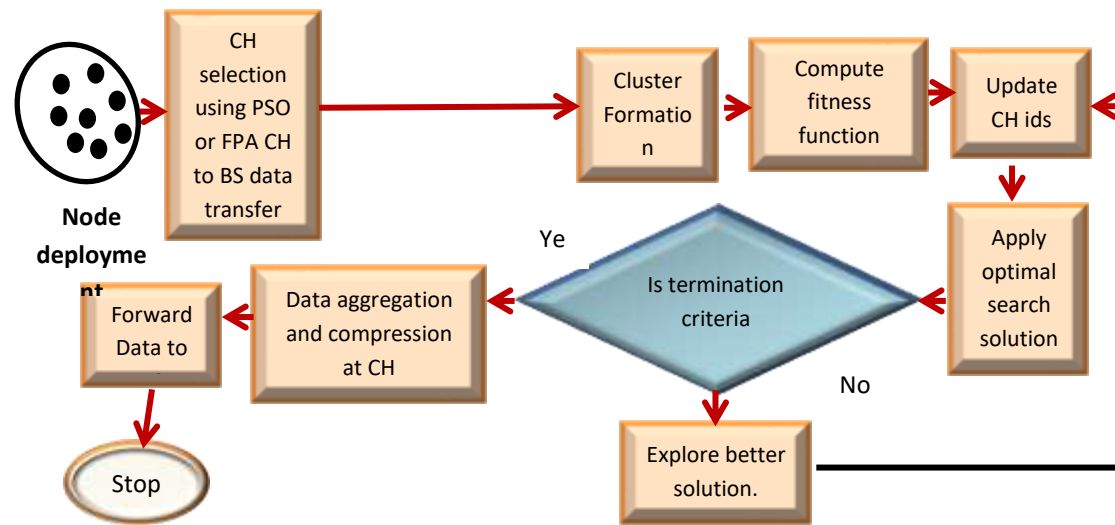
III. METHODOLOGY

In this work randomly deployed sensor nodes are considered to be scattered in a square field. The base station is assumed to locate at center. The data in packets are sent from sensor node to cluster head (CH) and from CH to the base station. The selection of set of CH is to be executed by finding the best set of sensor nodes by applying two different optimization algorithms. The selection criterion for set of CHs calculates the fitness of the group of randomly chosen nodes for becoming as the set of CHs. The node position and energy status are responsible for various factors like distance from neighboring nodes, distance from BS, eligibility to participate as CH etc. Hence these basic parameters are used to define higher function that more deeply represents the network and node parameters and reflects collective possibility of suitability of set of nodes as participant for CHs. In this work the optimization algorithm initially generates multiple sets of nodes as possible solution of CHs set.

A. Network parameters

The network parameters that are combined as per the proportionality to form the fitness function. These parameters are named as Converging function (F_{conv}), Diverging function (F_{Diverg}), Centrality, Node degree (N_{degree}), Distance to base station (D_{S2BS}) and Residual energy (E_{Res}). F_{conv} is defined as the sum of distance of nodes to the CH within the cluster (Figure 2). The best cluster is that which have most of the nodes nearest to the associated CH. Lower the F_{conv} better is the cluster geometry to consume lower amount of energy. F_{Diverg} is defined as the sum of the Distances between two cluster centers. It reflects the inter-cluster separation. If the clusters are far apart means less number of CHs covered highest number of nodes. Centrality describes the closeness of CH to the central position of cluster. Nearer the CH to center is better for all the nodes to transmit with lower energy expenses hence balance of energy consumption throughout the nodes in a cluster. Node degree is defined as the number of nodes in the cluster divided by total nodes.

Fig. 3 Block diagram illustrating the steps involved in execution of proposed protocol.



If count of CH in a solution set = C then count of cluster is also C, the count of nodes in a cluster = N_c , $D_{cluster}$ = sum of distance between nodes and CH.

$$D_{cluster} = \sum_{i=1}^{N_c} D_{Si-CH} \quad (1a)$$

$$F_{converging} = \sum_{j=1}^C D_{cluster}^j \quad (1b)$$

Distance between CHs = $D_{CHi \text{ to } CHj}$

$$F_{diverging} = \sum_{i=1}^C \sum_{j=1}^C D_{CHi \text{ to } CHj} \quad (2)$$

$$Centrality_{CHi} = \sqrt{\frac{\sum dis^2}{n}} \quad (3)$$

$\sum dis^2$: sum of squared distances of nodes to cluster heads i.

$$N_{degree} = \text{var} \left\{ \frac{(\text{number of nodes in Ckuster i})}{(\text{total number od nodes in network})} \right\} \quad (4)$$

B. Fitness Function

There is total six parameters but all do not have supportive impact on fitness measurement criteria for the set of nodes to be best suitable as CHs. Some are desired to be high and some are desired to be low. Fitness of set of nodes is high if F_{diverg} and E_{res}^{ch}

is high. It means the selected nodes as a CH are far apart and possess high residual energy. In this way fitness function is directly proportional to F_{diverg} and $E_{\text{res}}^{\text{ch}}$. Similarly, the fitness is inversely proportional to F_{conv} , centrality, D_{ch2bs} and N_{degree} . The total fitness function F_{total} is the sum of the six fitness factor of network parameters in direct or inverse proportionality as given below in equation 5.

$$F_{\text{total}} = (a_1 * F_{\text{Diverg}} + a_2 * E_{\text{res}}^{\text{ch}}) + (a_3 / F_{\text{conv}} + a_4 / \text{centrality} + a_5 / D_{\text{ch2bs}} + a_6 / N_{\text{degree}}) \quad (5)$$

$$F_{\text{total}} = f(f_1, f_2, 1/f_3, 1/f_4, 1/f_5, 1/f_6) \quad (6)$$

{a1 to a6} is the set of scaling factor that are individually multiplied to each factor f_1 to f_6 to bring all the factors in similar range to maintain uniform weightage of each parameter in F_{total} fitness function value. Hence the final value of fitness function is given in equation 7.

$$\text{fitness value} = \sum_{i=1}^6 F_i * a_i = f_1 * a_1 + f_2 * a_2 + 1/f_3 * a_3 + 1/f_4 * a_4 + 1/f_5 * a_5 + 1/f_6 * a_6 \quad (7)$$

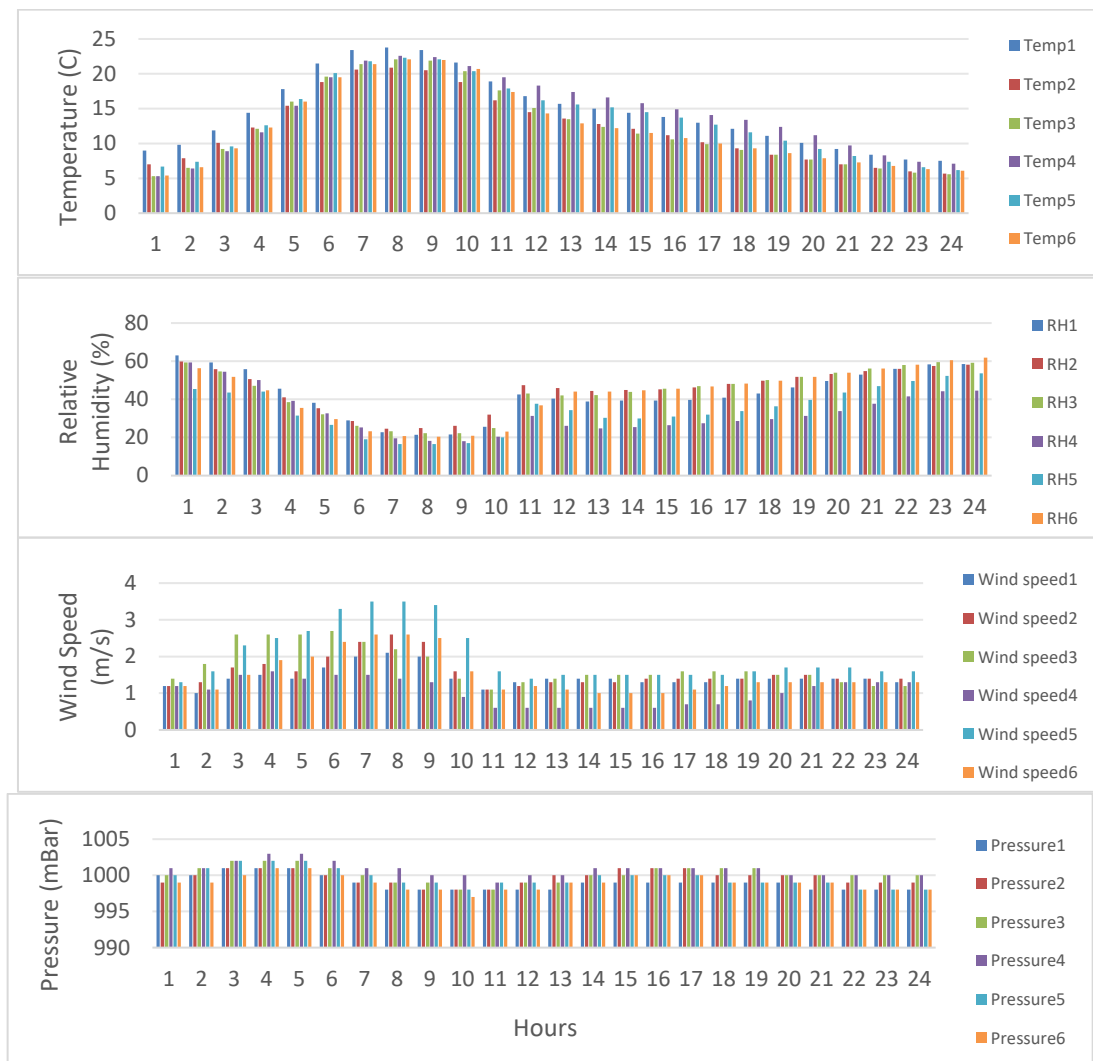
C. Data transmission and compression

The optimization algorithms select the different node ids and make multiple sets of proposed possible solution as the CHs. The fitness value of each proposed solution from the population of sets of expected CHs is calculated and the best set is traced out and further I next iteration (generation) the node ids is updated from neighboring nodes as per the displacement (movement) rules of PSO or FPA to find a better solution. The optimization process runs for several iterations until the no further updates in proposed solution is achieved. After finding the best cluster head set the cluster is formed and energy consumed with respect to distance between nodes to CH is calculated for transmission of k bits of data. After the phase of first hop transmission S to CH the transferred data is aggregated at cluster head and two different data compression algorithms (ASWDR and ZSTD) are run to reduce the data size prior to transmission phase of CH to BS. After data compression phase the energy consumed in sending packets from CH to BS is calculated. The complete mechanism is demonstrated in the flowchart given in figure 4.

C. Data set

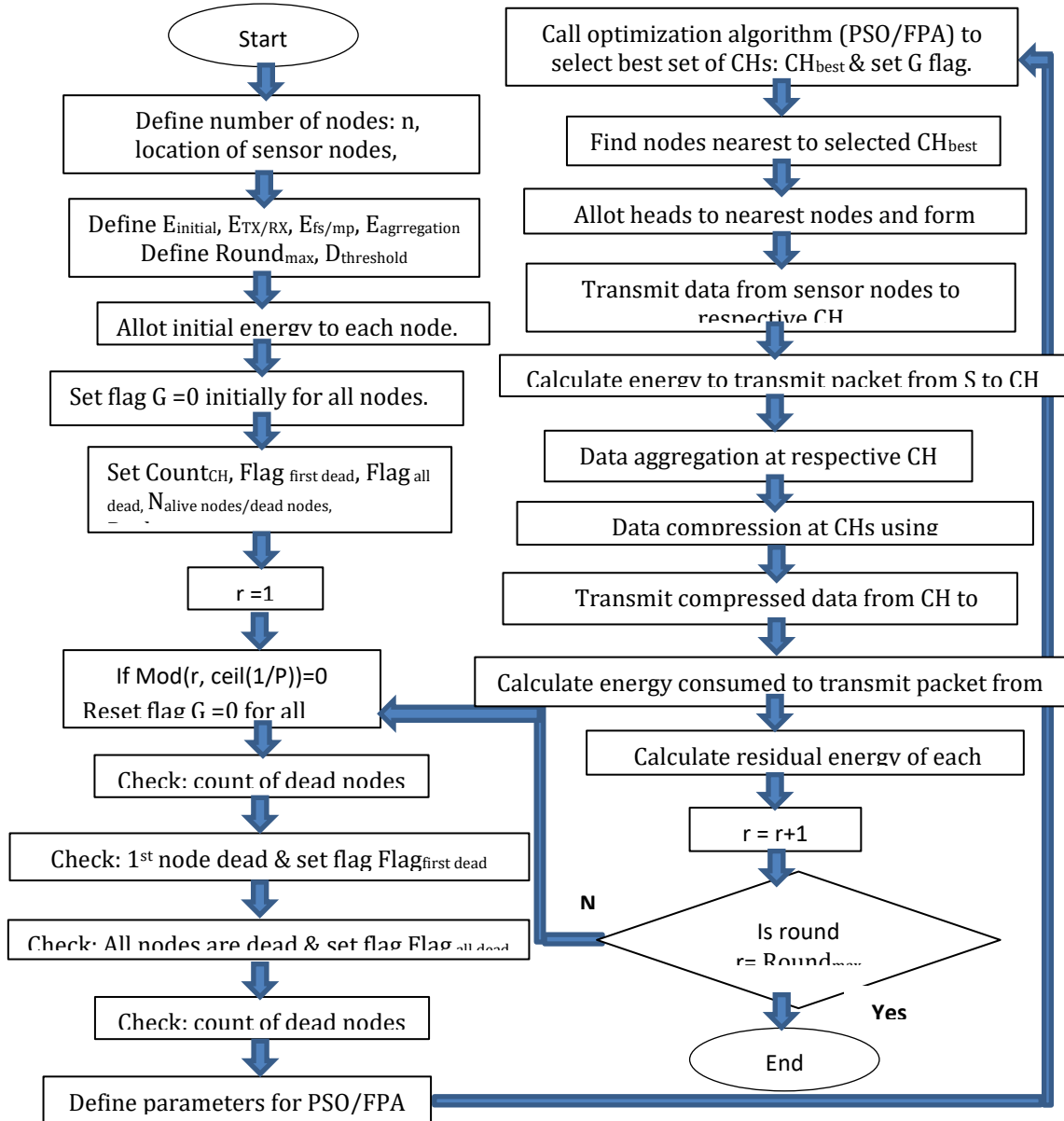
The dataset is collected from National Solar Radiation Database (NSRDB) website through the link <https://mapfiles.nrel.gov/data/solar>. The excel sheet consist of Temperature (°C), Relative Humidity (%), Pressure (mBar), Wind Speed (m/sec) data records recorded for the location at Latitude, Longitude of [26.93, 80.98]. This location belongs to sensor records of metrological data placed at Kukrail forest area, Lucknow, U. P., India. The NSRDB offers high quality temporal and spatial resolution dataset that consist of measurements of solar radiation with other meteorological data. NSRDB is established for data-based model using cloud and weather information primarily useful for airports. These data are free available on user registration to NSRDB website. Figure 4 is showing the hour wise plot of data samples by six locations of sensor for four different types sensor records of temperature, relative humidity, pressure and wind speed. This figure is only demonstrating data samples of 24 sensors (4 sensor types at six locations) for 24 hours only for illustration to understanding. However, the data base is large and consists of data records of throughout the year.

Fig. 4 Representation of data records used to simulate the WSN based transmission of weather data.



To mechanism for moving current best solution to next best solution is shown in figure 6a. Similarly, the selection process of CHs set is also executed using FPA algorithm but in this algorithm each sensor node is considered as flower and the search for new cluster head is carried out for nodes situated in area inside the cluster named as self-pollination and the nodes present in neighboring clusters named as cross pollination as shown in figure 6b.

Fig. 5 Flowchart of proposed algorithm for selection of set of cluster heads with optimum fitness value.



PSO Algorithm

Input: Lower bound, Upper Bound, Parameters

Parameters: [pop, iter, α]

1. Call Sub Program to Initialize Particle position to get best solution: PSO_{initialize} %% Best Solutions: best_{NCH x Pop}
 2. Fbest=0 %consider small value of fitness
- Loop i= 1 to iter
Loop j= 1 to pop

```

3. Xa=Best{i} % Input best particle one by one
4. Find fitness value (fval) for ith best particle
5. If Fval of ith best particle > Fval of Gbest
6. Update the best particle
end
end
7. Call Sub Program to update the particle position: PSOmove
8. Xa new position = Xa old position + velocity
9. Best{i} = Xa new position
End

```

Sub Program: PSO_{initialize} (Initialize Particle position to get best solution)

Description: Set random cluster head ids set as an initial solution(population)

Input: Pop, lower bound, upper bound, n

Generate random value for each node: temp_{rand} | pop x n

```

1. Calculate probability for selection of node as a cluster head: p
2. p = P * n * (1 + a) * E. / (n + A) * (Ea)
Loop k=1 to pop %select CH for each particle
3. if temprand < (p / (1 - p * mod(r, round(1/p))))
4. select the sensor node as CH
end
5. save sensor node it to CHid
6. Return : CHid as Best solution for first iteration to PSO algorithm

```

Sub Program: PSO_{move} (Update Particle position to get best solution)

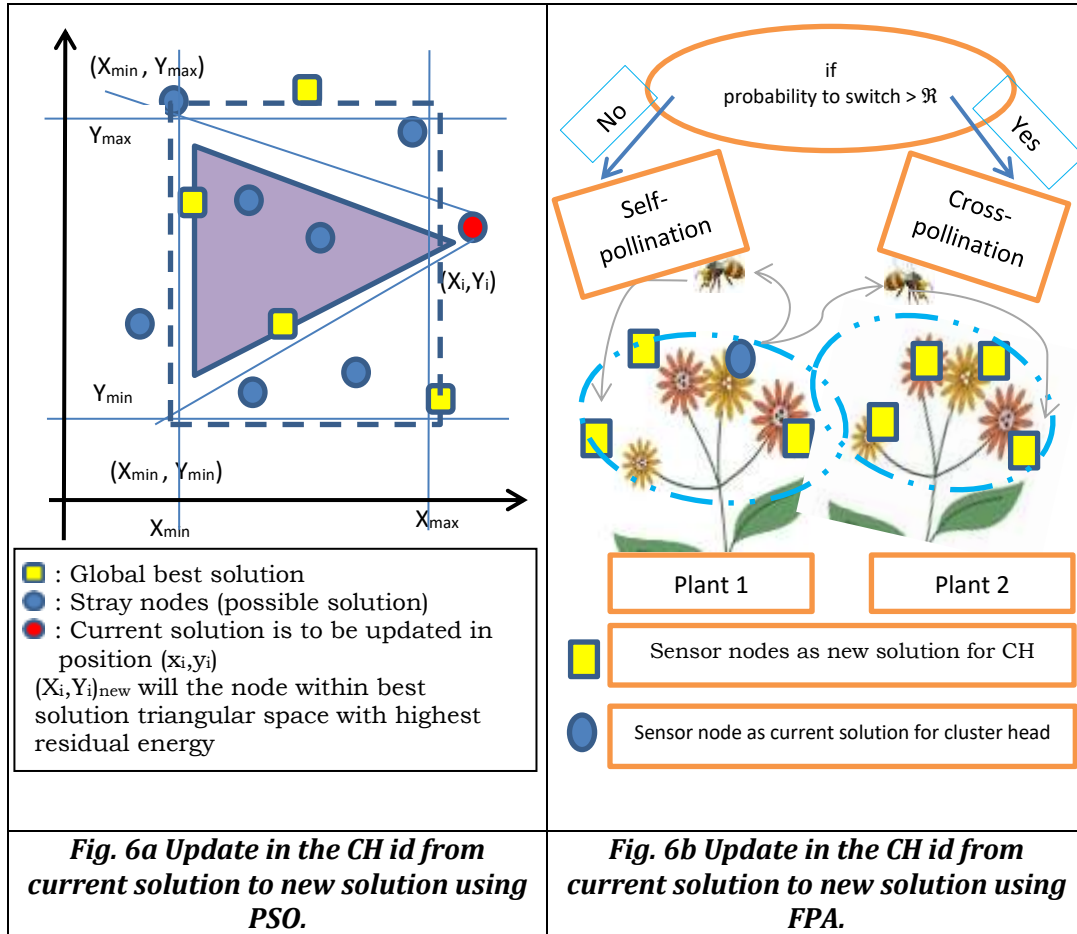
Description: Update the cluster head ids set with respect to best set of CH ids (best solution)

Input: best_{current}, best_{global}, pop, α , Lower bound, upper bound

```

1. Xr, Yr: Get Location of all the node
2. Xr gbest, Yr gbest: Get Location of all the CH ids of global best solution
3. Find minimum / maximum value of x and y co-ordinates of CHid gbest {Xmin, Xmax, Ymin, Ymax}
4. Set a square region using {Xmin, Xmax, Ymin, Ymax}
   Loop j= 1 to number of nodes in bestcurrent
5. Local best: Lbest(j1): get CHids of j1th local best solution
6. Generate triangle with vertices as : [ Lbest(j1)x,y , Farthest points of square from
   Lbest(j1)x,y ]
7. Set triangle vertices as : [ XΔ1, YΔ1, XΔ2, YΔ2, XΔ3, YΔ3 ]
8. x/yinside : Find all the sensor nodes inside triangle
9. Select nearest node in set x/yinside to Lbest(j1)x,y as the next best solution
10. Update position Lbest(j1)x,y by nearest x/yinside : Lbest(j1)x,ynew = nearest x/yinside to
   Lbest(j1)x,yold
11. X/Ynew solution = coordinates of updated Lbest(j1)x,ynew % New position of particle
12. Add new solution with P * n / 2 sensor nodes with highest energy
   end
13. Return updates solution of CHid ids to PSO algorithm

```



FPA Algorithm

Input : Lower bound, Upper Bound, Parameters. iter

Parameters: [pop, probability of switching (p_{sw})]

%Initialize population of flower to get best solution: FPA_{initialize}

1. Allot random value to each i^{th} node in each population: Rand1_i^{Pop}
2. Find probability of selection: $p = P * n * (2 + a) * E. / (n + A) * (Ea)$;
3. If Rand1_i^{Pop} < p set the i^{th} node as CH as initial solution as CH_{id}^{temp}
4. %% Best Solutions: best_{NCH x Pop}
5. Find fitness value (f_{val}) for all the CH_{id}^{temp} in respective Pop
6. Set best solution as CH id with maximum fitness: Bset=Max Fitness (CH_{id}^{temp})
7. Loop: t = 1 to iter %run loop for maximum number of iterations
8. Loop: i1 = 1 to pop_size %run loop for population size i.e. number of flowers
9. generate random value: R₂
10. if R₂ < p_{sw} % Apply main search mechanism via global flower pollination
11. Find x,y coordinates of best solutions (Set of CH IDs)_{x,y}^{pop} as L_{best}^{x,y}
12. Find nearest node from best solution to all the nodes of current solution, g_{xy}^{near}
13. Calculation displacement in current solution node id:
14. Generate random numbers from a Levy distribution, L_x and L_y
15. Calculate the step increments, dS_x and dS_y
16. dS_x or y = L_x or y, (X or Y coordinates of node from current solution – X or Y coordinates of nearest node from best solution)

17. Update new solution: X_{new} or $Y_{new} = X$ or Y coordinates of nodes of current solution dS_x
or y
18. Find node Id nearest to position X_{new}/Y_{new}
19. else if $R_2 > p_{sw}$ % Find random pollen/flowers in the neighborhood
20. Randomly select two solutions from whole population as Sol_j and Sol_k
21. Find nearest CH_{id} of Sol_j to Sol_k
22. $(X_{new}/Y_{new})^{local} = X/Y$ coordinate of $sol_j + \text{epsilon}$. (X/Y coordinates of nearest CH_{id} of Sol_k)
23. Find nodes nearest to $(X_{new}/Y_{new})^{local}$ and set them as new solution for CH_{id_set}
24. end of if condition for switching probability
25. $F_{new} = \text{fitnessfunctionmetrics}(\text{New Solutions}_{i1})$
26. if ($F_{new} \geq \text{Fitness}(i1)$)
27. $\text{Fitness}(i1) = F_{new}$;
28. end
29. if $F_{new} \geq f_{max}$
30. $best = S(i1)$;
31. $f_{max} = F_{new}$;
32. end
33. end loop $i1$ % For-loop if all pop size covered
34. end loop t % For-loop if maximum iteration achieved

ASWDR Algorithm

1. Load the data signal
2. $x = \text{Data}$
3. $N = \text{Data length}$
4. $t = \text{time records}$
5. Define Wavelet Decomposition parameters as:
6. wavelet type: 'db4', level of decomposition: 2
7. Perform wavelet decomposition and save wavelet coefficients as C
- %Apply adaptive scanning based on magnitude
9. Sort the magnitude values of C in descending order
10. Define the Threshold for significance as $T = 0.5 * \max(\text{abs}(C))$
11. Save ids of sorted value as significant_{idx} for sorted values $> T$
12. Save ids of sorted value as $\text{insignificant}_{idx}$ for sorted values $< T$
- % Quantize significant coefficients (Difference Reduction)
13. Define quantization step size $Q_{step} = 0.1$
14. Calculate quantized value of C as $C_{quantized} = \text{round}(C / Q_{step})$
15. Only keep quantized value of significant coefficients
16. Perform Huffman encoding of significant $C_{quantized}$

ZSTD Algorithm

Step 1: Input Framing: The frame contains:

- Magic number (for format identification)
- Frame header (flags, content size, checksum)
- One or more compressed blocks

Step 2: Block Segmentation: Input data is divided into blocks (up to 128 KB each).

Step 3: LZ77 Match Finding: Apply variant of the LZ77 sliding window algorithm.

- Maintains a search window of previously seen data.
- Uses hash tables and binary trees for fast matching.
- Searches for repeated byte sequences and Long-distance matches

Output: Matches represented as (offset, length) pairs.

Step 4: Literal Extraction: Bytes not covered by matches are stored as literals and grouped into literal blocks.

Step 5: Sequence Generation: Compressed data represented as sequences, each consisting of: (literal_length, match_length, match_offset)

- literal_length: number of raw bytes before a match
- match_length: length of the matched sequence
- match_offset: distance to the match source

Step 6: Entropy Coding Preparation: Encodes sequence components separately as:

Component	Encoding Method
Literal lengths, Match lengths, Match offsets	FSE
Literal bytes	Huffman

Step 7: Finite State Entropy (FSE) Encoding: The encoding steps are as follow:

1. Symbol frequency analysis
2. Normalization of frequencies
3. Construction of FSE tables
4. Bitstream encoding using state transitions

Step 8: Huffman Encoding for Literals: Literal bytes are compressed using Huffman coding.

Step 9: Block Assembly: Each compressed block contains:

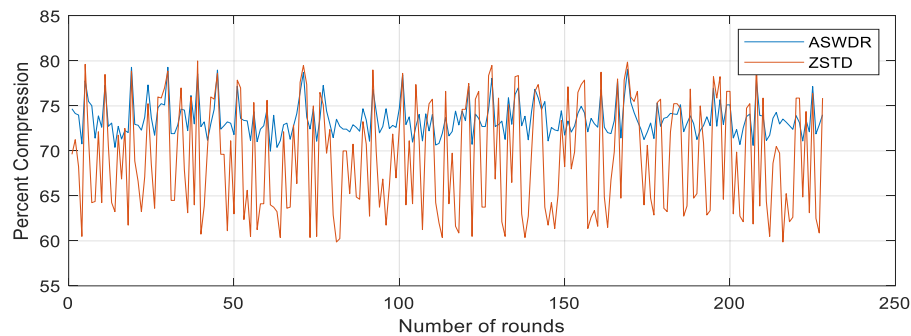
- Literal section (Huffman encoded)
- Sequence section (FSE encoded)
- Block header

Blocks are marked as Last block or non-last block and compressed or raw block

Step 10: Frame Construction: All blocks are concatenated.

IV. RESULTS and DISCUSSIONS

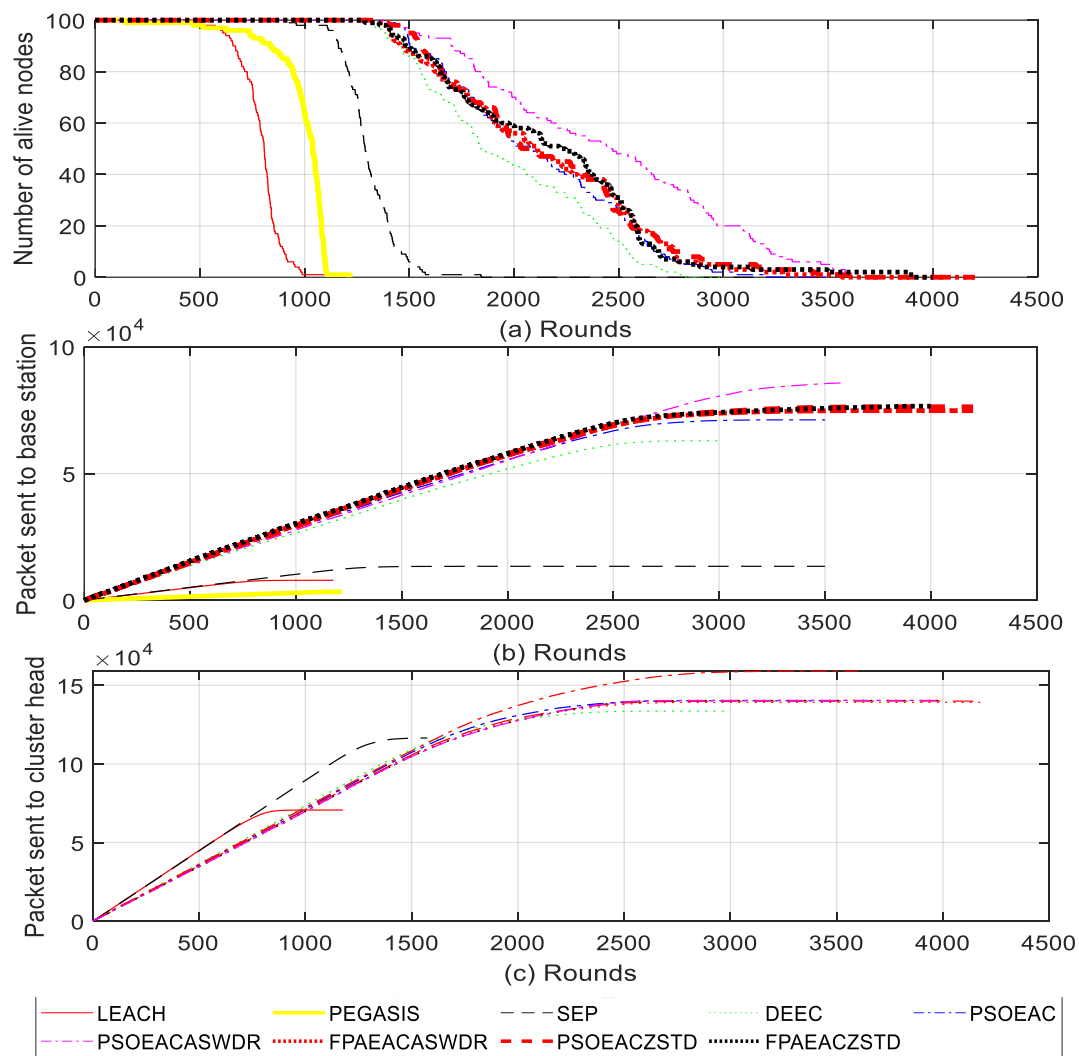
Fig. 7 Percent compression ratio at different rounds for ASWDR and ZSTD data compression



The code is developed in Matlab programming language for standard network parameters of initial energy, free space/ multipath losses, data aggregation energy and transmitter/receiver electronics energy. The maximum number of rounds are taken to be 4000 rounds for transmission of data packets. For the validity of results conventional

methods are also run for LEACH, EGASIS, SEP and DEEC. The code for without using any compression is also executed and named as PSO energy aware routing (PSO EAC) and FPAEAC. The CH set optimum selection with data compression algorithms are finally run for combination of ASWDR and ZSTD compression and named as PSOEACASWDR and FPAEACASWDR, PSOEACZSTD and FPAEACZSTD. Figure 7 is showing the result for variation in compression ratio at different rounds for the packet sent from CHs to BS. For the ASWDR the compression ratio is varying in between 70 to 80% but in the case of ZSTD the variation is large from 60 to 80%. On average the compression ratio for ASWDR is 0.73 observed to be and ZSTD has 0.69. In this way the ASWDR is found to be giving higher compression on average.

Figure8. Variation of number of alive nodes, Packet sent from CH to BS and packet sent from node to CH with respect to number of rounds



Along with the results further calculated in terms of network performance in terms of network lifetime as number of alive nodes at each round (Figure 8a), number of packets sent from CH to base station (Figure 8b) and number of packets sent from node to CHs. The figure 8a is used for figuring out the network lifetime in terms of round at which first node is dead, half of the nodes are dead and all the nodes are dead and named as First

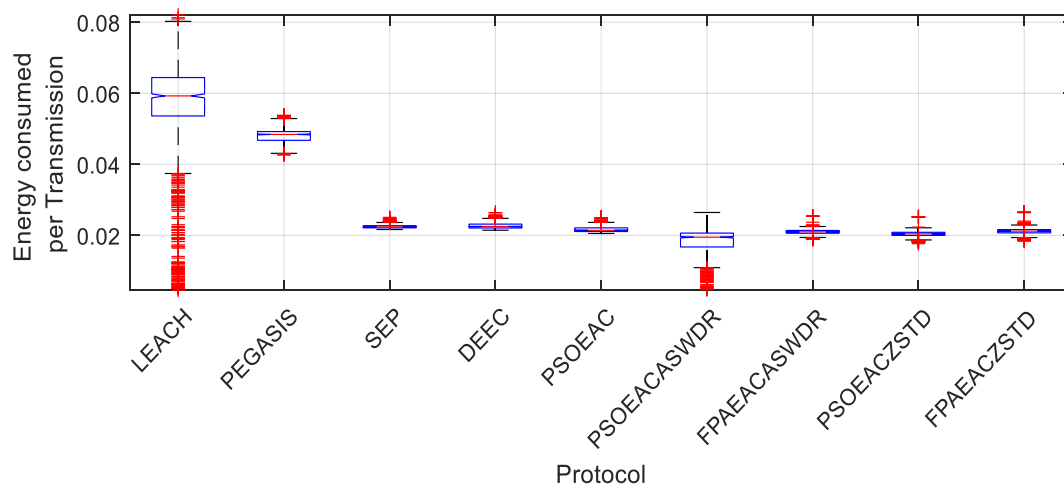
dead, Half dead and all dead. The total packets sent by all algorithms is finally summed up and tabulated as table 1. It shows that the highest number of packets is sent by FPAEACASWDR for sensor node to CH as 4.40×10^8 and CHs to BS as 2.11×10^8 . In this way the FPAEACASWDR is found to be best one to select best combination of CHs using FPA and higher compression ratio using ASWDR to gives finest performance.

Table1. Packet Sent

Packet Sent	PSOEAC ASWDR	FPAEAC ASWDR	PSOEAC ZSTD	FPAEAC ZSTD
Sensor node to CH	3.83×10^8	4.40×10^8	3.82×10^8	4.15×10^8
CH to Base Station	1.74×10^8	2.11×10^8	1.70×10^8	2.03×10^8

Since the energy consumed depends upon two factors distance of transmission and number of packets sent. Due to compression the packet size is reduced at CHs about 60 to 70% hence lower energy is used at the phase of CH to BS data transmission while the selection of CHs at minimum convergence, higher centrality, node degree and divergence distance supports the minimum distance of transmission. The boxplot of energy consumed by the all algorithm is shown in figure 9.

Fig. 9 Boxplot of energy consumed by different conventional and proposed algorithms



The values of energy consumed are shown in table 2 for all the algorithms and the minimum, maximum, average and standard deviation value are shown. It is observed that FPAEACASWDR consumes on average 0.0178 J energy per transmission that is lowest one.

Table2. Energy Consumed

Energy Consumed (Joules)	LEACH	PEGASIS	SEP	DEEC	PSOEAC ASWDR	FPAEAC ASWDR	PSOEAC ZSTD	FPAEAC ZSTD
Minimum	0.0033	0.0426	0.0217	0.0215	0.0035	0.002	0.0033	0.0184

Maximum	0.1208	0.0538	0.0251	0.0264	0.0265	0.0255	0.0255	0.0265
Average	0.0555	0.0484	0.0225	0.0228	0.0182	0.0178	0.0179	0.0212
Standard Deviation	0.0178	0.0025	0.0005	0.0009	0.0037	0.0039	0.0036	0.0011

For further validating the results the simulation scenario is further varied by changing the number of nodes and the network area. The number of nodes is taken to be 100 and 200 while the network area is taken as 100x100m² and 200x200m². In this way total 4cases are created as [nodes, area]: [100, 100x100m²], [100, 200x200m²], [200, 100x100m²], [200, 200x200m²]. All the simulation results are shown in terms of rounds at which number of nodes observed to dead as first dead, half dead and all dead as shown in table 3.

Table 3 Network lifetime in terms of respective rounds at given recorded number of dead nodes for different simulation setup in terms of total number of nodes and node deployment area

Nodes	Field Area length/width		First Dead	Half Dead	All Dead
100	100	PSO ASWDR	1455	2146	3803
200	100		1277	1964	3787
100	200		840	1747	3386
200	200		1055	1693	3561
100	100	PSO ZSTD	1361	2078	3687
200	100		1056	1977	3650
100	200		796	1700	3433
200	200		1030	1676	3868
100	100	FPA ASWDR	1299	2112	3565
200	100		1053	1994	3536
100	200		681	1591	3324
200	200		968	1632	3921
100	100	FPA ZSTD	1288	2247	3909
200	100		1148	1944	3872
100	200		921	1674	3242
200	200		1009	1641	3230

TABLE 1 Table caption

Table Head	Table Column Head		
	Table column subhead	Subhead	Subhead
copy	More table copy ^a		

IV. CONCLUSION

This paper presented an energy-efficient clustering and data transmission framework for wireless sensor networks aimed at addressing the challenges posed by limited node

energy and high communication overhead in IoT-based weather monitoring applications. Unlike traditional clustering approaches that focus solely on cluster head selection or rely on direct CH-to-BS communication, the proposed method integrates optimization-based CH selection with compression-aware data transmission to achieve improved energy efficiency across the network.

The Flower Pollination Algorithm was employed to identify an optimal set of cluster heads by considering multiple network and node-level parameters, thereby ensuring balanced cluster formation and uniform energy utilization. By applying lightweight data compression techniques at the aggregation stage, the volume of transmitted data from CHs to the Base Station was significantly reduced, leading to lower transmission energy consumption. Comparative performance evaluation against classical protocols such as LEACH, PEGASIS, SEP, and DEEC demonstrated that the proposed FPA-based approach achieves superior network stability, extended lifetime, improved packet delivery, and reduced average energy consumption under different network sizes and deployment scenarios.

Overall, the proposed framework effectively combines metaheuristic optimization and data compression to enhance the sustainability and scalability of wireless sensor networks. Future work may explore adaptive compression thresholds, heterogeneous sensor environments, and real-time deployment scenarios to further improve robustness and applicability in large-scale IoT systems.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

REFERENCES

1. X. Cai, Y. Sun, Z. Cui, W. Zhang, and J. Chen, "Optimal LEACH protocol with improved bat algorithm in wireless sensor networks", *KSII Transactions on Internet and Information Systems*, vol. 13, no. 5, 2019.
2. Z. Cui, Y. Cao, X. Cai, J. Cai, and J. Chen, "Optimal LEACH protocol with modified bat algorithm for big data sensing systems in internet of things", *Journal of Parallel and Distributed Computing*, vol. 132, 2019, pp. 217–229.
3. L. Dai, B. Wang, L. T. Yang, X. Deng, and L. Yi, "A nature-inspired node deployment strategy for connected confident information coverage in industrial internet of things", *IEEE Internet of Things Journal*, vol. 6, no. 6, 2019, pp. 9217–9225.
4. K. Dev, P. K. R. Maddikunta, T. R. Gadekallu, S. Bhattacharya, P. Hegde, and S. Singh, "Energy optimization for green communication in IoT using Harris hawks optimization", *IEEE Transactions on Green Communications and Networking*, vol. 6, no. 2, 2022, pp. 685–694.
5. K. Dhou and C. Cruzen, "An innovative chain coding technique for compression based on the concept of biological reproduction: An agent-based modeling approach", *IEEE Internet of Things Journal*, vol. 6, no. 6, 2019, pp. 9308–9315.
6. U. Draz, T. Ali, S. Yasin, et al., "Hybridization and optimization of bio and nature-inspired metaheuristic techniques of beacon nodes scheduling for localization in underwater IoT networks", *Mathematics*, vol. 12, no. 22, 2024, p. 3447.
7. M. Z. Ghawy, G. A. Amran, H. AlSalman, et al., "An effective wireless sensor network routing protocol based on particle swarm optimization algorithm", *Wireless Communications and Mobile Computing*, vol. 2022, 2022, pp. 1–13.
8. P. Kathirolu and K. Selvadurai, "Energy efficient cluster head selection using improved sparrow search algorithm in wireless sensor networks", *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 10, 2022, pp. 8564–8575.

9. K. Kaur and S. Kaur, "Hybrid bio-inspired optimization-based routing protocol for enhancing data transmission in clustered network", *Array*, vol. 27, 2025, p. 100481.
10. U. Khan, I. M. Qureshi, M. A. Aziz, T. A. Cheema, and S. B. H. Shah, "Smart IoT control-based nature inspired energy efficient routing protocol for flying ad hoc network (FANET)", *IEEE Access*, vol. 8, 2020, pp. 56371–56378.
11. G. Kumar, R. Saha, M. Conti, T. Devgun, and R. Thomas, "GREPHRO: Nature-inspired optimization duo for internet-of-things", *Internet of Things*, vol. 25, 2024, p. 101067.
12. N. Kumar, K. Singh, and J. Lloret, "WAOA: A hybrid whale-ant optimization algorithm for energy-efficient routing in wireless sensor networks", *Computer Networks*, vol. 254, 2024, p. 110845.
13. J. Ma, S. Wang, C. Meng, Y. Ge, and J. Du, "Hybrid energy-efficient APTEEN protocol based on ant colony algorithm in wireless sensor network", *EURASIP Journal on Wireless Communications and Networking*, no. 1, 2018, p. 102.
14. S. Mnasri and M. Alrashidi, "Energy-efficient IoT routing based on a new optimizer", *Simulation Modelling Practice and Theory*, vol. 119, 2022, p. 102591.
15. B. N. K. Reddy and A. S. Kumar, "Evaluating the effectiveness of bat optimization in an adaptive and energy-efficient network-on-chip routing framework", *Journal of Parallel and Distributed Computing*, vol. 188, 2024, p. 104853.
16. M. P. K. Reddy, "Cluster head selection in IoT using enhanced self adaptive bat algorithm", *Journal of Networking and Communication Systems*, vol. 2, no. 4, 2019.
17. R. Rani and S. Sharma, "Nature-inspired routing and optical communication for next-generation IoT systems", *Journal of Optical Communications*, 2025.
18. M. Raval, S. Bhardwaj, A. Aravelli, J. Dofe, and H. Gohel, "Smart energy optimization for massive IoT using artificial intelligence", *Internet of Things*, vol. 13, 2021, p. 100354.
19. G. Santhosh and K. V. Prasad, "Energy optimization routing for hierarchical cluster based WSN using artificial bee colony", *Measurement: Sensors*, vol. 29, 2023, p. 100848.
20. L. Wang, Y. Luo, and H. Yan, "Optimization analysis of node energy consumption in wireless sensor networks based on improved ant colony algorithm", *Sustainable Energy Technologies and Assessments*, vol. 64, 2024, p. 103680.
21. D. Piątkowski, T. Puślecki, and K. Walkowiak, "Study of the impact of data compression on the energy consumption required for data transmission in a microcontroller-based system", *Sensors*, vol. 24, no. 1, 2023, p. 224.
22. D. S. Khafaga, E. A. Aldakheel, A. M. Khalid, H. M. Hamza, and K. M. Hosny, "Compression of bio-signals using block-based Haar wavelet transform and COVIDOA for IoMT systems", *Bioengineering*, vol. 10, no. 4, 2023, p. 406.
23. J. Azar, G. B. Tayeh, A. Makhoul, and R. Couturier, "Efficient lossy compression for IoT using SZ and reconstruction with 1D U-Net", *Mobile Networks and Applications*, vol. 27, no. 3, 2022, pp. 984–996.
24. Y. Wang, K. Yang, W. Wan, Y. Zhang, and Q. Liu, "Energy-efficient data and energy integrated management strategy for IoT devices based on RF energy harvesting", *IEEE Internet of Things Journal*, vol. 8, no. 17, 2021, pp. 13640–13651.
25. M. Zhang, H. Zhang, D. Yuan, and M. Zhang, "Learning-based sparse data reconstruction for compressed data aggregation in IoT networks", *IEEE Internet of Things Journal*, vol. 8, no. 14, 2021, pp. 11732–11742.
26. G. Signoretti, M. Silva, P. Andrade, I. Silva, E. Sisinni, and P. Ferrari, "An evolving TinyML compression algorithm for IoT environments based on data eccentricity", *Sensors*, vol. 21, no. 12, 2021, p. 4153.
27. S. A. Abdulzahra, A. K. M. Al-Qurabat, and A. K. Idrees, "Compression-based data reduction technique for IoT sensor networks", *Baghdad Science Journal*, vol. 18, no. 1, 2021, p. 0184.

28. S. Sawalha and G. Al-Naymat, "Internet of things data compression based on successive data grouping", Turkish Journal of Electrical Engineering and Computer Sciences, vol. 29, no. 1, 2021, pp. 32–45.