Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

An Augmented Fuzzy C-Means Algorithm for Efficient Data Clustering in IoT-Enabled Wireless Sensor Networks

Sonia Singh[1], Dr. Neha Gupta[2]

[1]Research Scholar, School of Computer Applications, Manav Rachna International Institute of Research and Studies, Faridabad, India
[2]Professor, School of Computer Applications, Manav Rachna International Institute of Research and Studies, Faridabad, India
Email: 14sonia.singh@gmail.com

In the world of wireless sensor networks (WSN), the constraint of limited energy resources owing to battery dependency needs effective energy management solutions to extend network longevity. Among the various options, clustering has emerged as a pivotal technique for increasing energy efficiency. The paper delves into the critical challenge of optimally clustering sensor nodes (CN) to extend the operational longevity of WSNs amidst the burgeoning data demands. The paper presents an analytical methodology to generate the optimal number of clusters, which is supplemented by the development of a centralized clustering algorithm. Furthermore, the present research outlines the implementation of a decentralized clustering framework that employs the fuzzy C-means algorithm. Comprehensive simulations are used to compare the proposed algorithm with the established Hybrid Energy-Efficient Distributed (HEED) clustering technique. The findings clearly demonstrate the increased efficiency of our proposed algorithm in optimizing energy consumption and extending network lifespan, indicating a significant step forward in optimizing Wireless Sensor Networks (WSNs) for future high-density data processing requirements.

Keywords: Big data wireless sensor networks, Clustering, Cooperative nodes, Fuzzy C-means, Spectral partitioning.

1. Introduction

WSNs have become a vital data-gathering platform in a broad spectrum of applications, including home security [1]. A WSN comprises many sensor nodes communicating wirelessly to carry out distributed sensing tasks. The volume of sensor data is growing explosively in many applications due to the advancement of IoT equipped WSNs, that can be categorized into heterogeneous and homogeneous types. In homogeneous WSNs, sensor nodes are uniformly distributed, ensuring equal spacing throughout. On the other hand, heterogeneous WSNs distinguish themselves by selecting, crucial nodes as cluster heads, while the rest serve as standard nodes within the cluster. Even though the data produced by a single sensor might not seem important, WSNs with many sensors can generate a sizeable amount of data in various civilian and military applications [2]. In multi-hop wireless sensor networks, the data from each and every sensor node must be sent over multiple hops in order to reach the cluster heads. At its core, any clustering technique involves management processes, which include determining the method for selecting cluster heads, optimizing the number of data transmissions within and across clusters, and facilitating communication from the clusters to their respective heads [8]. One significant limitation of current clustering methodologies discussed in various studies is their failure to address energy consumption patterns and the lack of research on determining the optimal number of clusters for a network. Few researchers have adopted a two-layer approach where hybrid networks slightly increase the initial cost but significantly enhance the network's lifespan and stability post-deployment [6]. Reducing energy consumption in sensors, computation, and transmission is crucial for Wireless Sensor Networks (WSNs), as replacing batteries in harsh environments is often impractical once they are deployed. Extensive data analysis makes data collection and analysis more convenient, yet sensor node counts and volume of data are expanding quickly [3]. Furthermore, many existing algorithms fall short in efficiently handling large data environments.

The present research introduces an optimal cluster count based on a power utilization model where each group comprises many cluster members that transfer the collected data to their corresponding collecting head referred as Cluster Head (CH). The CH then relays this information directly to the sink node [4]. This method is designed to conserve energy within the sensor network by streamlining the data transmission process. The task of cluster formation is determined by the sensor nodes' capability to transfer the data and communicate directly with their Cluster Heads (CHs). Given the constraints of limited battery life and the need for low cost, sensor nodes are typically embedded with basic computing components and radio transmitters [5]. However, unlike previous works, the present study aims to achieve efficient data clustering amidst the challenges of handling extensive datasets. Clustering has been demonstrated as a successful strategy for increasing the scalability of WSN.

The research contributions are compared with existing research, and have been illustrated as follows:

- For the broadly distributed wireless sensor networks, the present paper proposes a framework to analyze energy consumption and determine the optimal number of clusters required. This framework employs collaborative transmission to facilitate communication between clusters, thereby significantly reducing the energy consumption and extending the lifespan of Cluster Heads (CHs) through the deployment of cooperative nodes.

- The paper recommends the adoption of decentralized clustering algorithms, employing fuzzy C-means to select the cluster heads, alongside a centralized clustering strategy that leverages spectral partitioning.

● The effectiveness of the proposed algorithm here is determined by the network's lifetime and the residual energy of nodes. This algorithm has the potential to decrease the time to the first and last sensor network failures, highlighting its capability to analyze and optimize the energy exhaustion of sensor nodes.

2. Related work:

The development of efficient data clustering algorithms in IoT-enabled Wireless Sensor Networks (WSNs) is crucial for optimizing resource utilization and improving network performance. These algorithms are designed to group sensor nodes to form the clusters. Usually, each cluster has a leader, called a cluster head, that handles the collection of data and communicate with the main sink node. Various approaches have been explored to achieve efficient data clustering in IoT-enabled WSNs. The notable contributions of the prominent authors are outlined below.

Hussein, A., & Abdulrazzak, H. (2024) [7] identified Cluster-based routing or hierarchical routing as an approach with benefits of scalability and communication efficiency. In IoT equipped wireless sensor networks, the utilization of clustered routing was a cost-effective method of data collection and transmission, where nodes with the most remaining energy could be utilized to collect and transmit data. The authors proposed a soft Fuzzy C-Means Dynamic Clustering (FCM-DC) approach with a Cluster Head (CH) selection technique. Their model was compared with an existing routing protocol, LEACH and LEACH-C protocol. The proposed FCM-DC model reduced the packet delay by about 16% in a 50-node scenario and 86% in a 100-node scenario as compared with LEACH.

Lahmar, A. et.al, (2024) [8] discussed the progression of wireless sensor network-based systems (WSN) with decentralized Internet of Things (DIoT). Various obstacles encountered the DIoT and WSN networks, with the key one being decreased energy consumption to extend network lifetime. They introduced the Type-2 Fuzzy Harris Hawks Optimization method to greatly improve energy-efficient routing for DIoT and WSN networks. This improvement has important benefits for different sectors and applications, leading to longer network lifetimes and more sustainable IoT setups.

Kumar, P. S., & Barkathulla, A. (2023) [13] introduced a new method called Fuzzy Clustering and Optimal Routing (FCOR) to reduce energy use and delay, while improving network lifespan and node density. It was tested FCOR's performance using measures like energy efficiency, packet delivery, and network lifespan. The results showed that FCOR used much less energy compared to other methods like FRNSEER, E-ALWO, ACI-GSO, and CRSH.

Naderloo, S. et.al (2023) [23] introduced a fuzzy-based clustering protocol for routing. This protocol utilized various performance parameters such as network zoning, remaining energy of node, distance from the center and node-to-base station for the selection of cluster head. The efforts were made to reduce the energy consumption of the network by assigning the responsibility of selecting Cluster Heads in each area and calculating distances to the center, and computing node angles to the Base Station. Their comparative results demonstrated that the proposed method outperformed existing protocols across various criteria, notably enhancing network lifetime.

Singh, S., and Anand, V., (2023) [28] proposed a hybrid optimization approach involving the fuzzy c-means clustering algorithm and Grey Wolf optimization (GWO) for clustering algorithms. This technique is applied to Internet of Things (IoT) equipped wireless sensor networks (WSNs) devices that are characterized by extensive deployments in challenging areas and require energy-efficient operations due to their autonomous nature. The proposed scheme's performance is assessed against various metrics, including energy requirement and network lifespan. It is concluded that the developed protocol effectively extends the network lifetime, making it well-suited for IoT monitoring systems.

Cherappa, V., Thangarajan, S. e.t al (2023) [31] worked to extend the lifespan of IoT quipped wireless sensor networks (WSNs) and to reduce power usage. The selection of cluster heads is noted as a persistent challenge in minimizing WSN energy consumption. In their study, sensor nodes (SNs) have been clustered by an Adaptive Sailfish Optimization (ASFO) algorithm along with K-medoids clustering algorithm. The research primarily focused on optimizing the Cluster head selection to stabilize the consumption of energy, reduce distance, and minimize the time latency between nodes.

Tumula, S., Ramadevi, Y., e.t al (2023) [29] discussed a notable rise in demand for IoT across sectors like industrial automation, healthcare, and surveillance, driving the integration of optimal WSNs into transmission methods. WSNs, comprising distributed sensor nodes, were pivotal in capturing and relaying environmental data. With energy consumption becoming a key concern, research focused on tailored design techniques and protocols for IoT applications. The study proposed the opportunistic energy-efficient dynamic self-configuration routing (OEDSR) algorithm. This algorithm computed optimal routes to the base station by leveraging sensor node residual energy and mobility factors, employing a hierarchical tree architecture for dynamic cluster generation to streamline connections.

Sharmin, S., et.al (2023) [27] conducted research for the enhancement of efficiency in terms of energy and network longevity in wireless sensor networks (WSNs). The study acknowledged that clustering conserves power, but the lack of an effective cluster head (CH) selection process could complicate data collection and increase power usage, potentially leading to early node failure and shortened network life. In order to resolve the issue, the authors introduced a hybrid particle swarm optimization (HPSO) combined with an improved low-energy adaptive clustering hierarchy (HPSO-ILEACH) for CH selection, to boost energy efficiency and to stabilize the network. The algorithm was implemented for data aggregation and energy conservation, and resulted in a longer network lifetime.

Abdulzahra, et. al. (2022) [2] discussed a Clustering-based routing approaches that had clear advantages in efficient communication, scalability, and extending network lifespan. They introduced a clustering method for WSN-based IoT systems using Fuzzy C-Means (FCM) to select the best Cluster Head (CH). Instead of replacing CHs in every cycle, they used an energy threshold to adjust CH roles based on current energy levels, which helped to extend the sensor network's lifespan.

Ci X., et.al (2021) [32] discussed the integration of the Internet of Things (IoT) with sensor, RFID and network technologies to monitor large areas efficiently. The study proposed a novel clustering method designed to decrease energy usage and increase the network's lifespan. This method used a strategic cluster head selection of nodes based on various factors including energy levels, node density, and communication capabilities for integrating ordinary nodes into the network, prioritizing residual energy and distance. Simulation results showed that the algorithm effectively reduced and balanced energy usage across the network.

Yousif, Z., et.al (2021) [33] illustrated Wireless Sensor Networks (WSNs) as a pivotal sensing technology, transforming data collection and utilization in numerous smart city applications. The rapid depletion of sensor battery power posed a significant challenge, attributed to intensive computational tasks and communication operations. To address the issue, a new CH selection enhanced with probabilistic cluster head selection (LEACH-PRO), was introduced to measure prolong WSN node lifetimes, including probabilistic cluster head node selection. Simulation results demonstrated the superiority of LEACH-PRO over LEACH protocol.

Loganathan, S., et.al (2021) [16] mentioned the significance of wireless sensor networks in monitoring various physical conditions like temperature and humidity in remote areas but lacks the performance because of limited battery lifetime. The authors proposed a neighborhood-dependent self-diagnosis fault detection and data prediction clustering algorithm to identify the cluster heads and to improve network lifetime significantly. The proposed approach nearly doubled lifetime compared to LEACH, surpassed QLEACH and ECH, and achieved a 51% enhancement over the temporal approach.

Bensaid R., et.al (2020) [25] discussed the challenge of energy efficiency in Wireless Sensor Network (WSN)-based Internet of Things (IoT) system, given the increasing complexity of IoT deployments. The paper presented a clustering algorithm for IoT enabled WSNs. applications. The algorithm employs an FCM approach to form the cluster and minimizes overall energy consumption in each cluster by optimally selecting Cluster Head (CH) in each round. A comparison with the LEACH protocol is conducted to analyze performance. Results indicate that the proposed FCM method improves network lifetime.

Bhatti, D. M. S., et.al. (2016) [18] developed a cluster-based cooperative spectrum sensing algorithm to minimize the energy wastage, utilizing fuzzy c-means (FCM) to make clusters members and cluster heads (CHs) based on location, signal quality, and energy. The method addressed the unreliability of sensor data due to environmental interference. Although cooperative sensing increases accuracy, it also raises energy use and can reduce efficiency.

3. The spectral classification of network model

The present paper works on spectral classification, the Laplacian matrix, the energy model, and the network model.

3.1 Wireless Sensor Network model

Wireless Sensor Network (WSN) consists of N sensor nodes uniformly distributed across an N*N square area to represent the distribution of sensor nodes [9]. This network is assumed to possess fundamental capabilities and characteristics.

- All nodes are homogenous and thus are uniformly distributed.
- Uses Received Signal Strength Indicator localization (RSSI), so every node is informed of its location.

- Since they are sensor nodes, all static nodes cannot replenish their batteries.

3.2. Energy Consumption model:

The energy used by the signal to send a 1-bit message is determined on the basis of the electricity consumption equation

$$E_{XT}(d,k)=\{l*(E_{elec}+E_{fs}S^2) \text{ and } l*E_{elec}+E_{amp}S^4\}, S<S_0 \text{ and } mS\geq S_0 \quad (1)$$

Moreover, the present work differentiate itself from the work referenced in [11] by exploring the relationship between the Cluster Head (CH)'s coverage radius (R) and the cluster area size (S). This research presents an updated models for energy usage where the optimal cluster count may fluctuate. It also specifies that the single bit packet energy is required to receive a packet is measured by

$$E_{XR}(l) = l*E_{elec} \quad (2)$$

E_{diss} is the dissipated energy by a single bit by receiving or the transmitter's circuit in equations (1) and (2). Researchers report the energy exhaustion per bit in the multi-path fading mode and the free storage model using $E_{fs}S^2$ and $E_{amp}S^4$. The distance between the sending device and the receiving device is represented by S. The threshold S0 is described as follows:

$$2S_0=\sqrt{\frac{E_{fs}}{E_{amp}}} \quad (3)$$

The model for free remaining space is applied to the transmitter's energy consumption where the distance d0 is greater than d; for greater distances [10]. The energy exhaustion for computation is not considered in our model, as data transmitted consumes significantly more energy than computation. Assuming each Cluster Head (CH) transmits l-bits of information during each data collection phase, the energy exhaustion by a CH in one round can be expressed as

$$E_{ch}=\frac{M}{H}*l*E_{elec}+\frac{M}{H}*l*E_{ad}+\frac{M}{H}*l*E_{fs}*s_{SB}^4 \quad (4)$$

The Cluster Head (CH) wastes the energy by accumulating signals across nodes, broadcasting a network generated by a series to the SB, and receiving alerts from nodes. The total count of clusters is denoted by H, with Ead representing the energy required for each cluster to process a single bit of data. SSB stands for the aggregate path between a new cluster and the base station. The energy expenditure of each element within the cluster, known as Cluster Enrich, during one data collection cycle, is

$$E_{nrich} = l*E_{elec} + l*E_{elec} * s_{ch}^2 \tag{5}$$

In which S_{ch} denotes the separation between a node and a cluster head.

4. Clustering Algorithm:

The most appropriate number of clusters in a clustering algorithm is the ideal or most suitable number of distinct groups into which a dataset should be partitioned [12]. Also, determining this optimal number is the most significant step in any clustering process as it directly impacts the effectiveness and interpretability of the clustering results. Finding the optimal number of clusters involves balancing the trade-off between achieving high intra-cluster similarity and low inter-cluster dissimilarity. It aims to ensure that each cluster captures meaningful patterns within the data while minimizing redundancy and overlap between clusters [14].

Our clustering methodologies include three main steps to determine the optimal number of clusters and clustering algorithm for both centralized and distributed nodes described as follows:

4.1 Determination of the optimal number of clusters:

Determining the ideal clusters, denoted as k. It is crucial since the inter-cluster communication is directly proportional to a rise in k. Conversely, decreasing k enhances intra-cluster communication frequency. Therefore, an energy model is being examined and described in next section to form the optimal number of clusters.

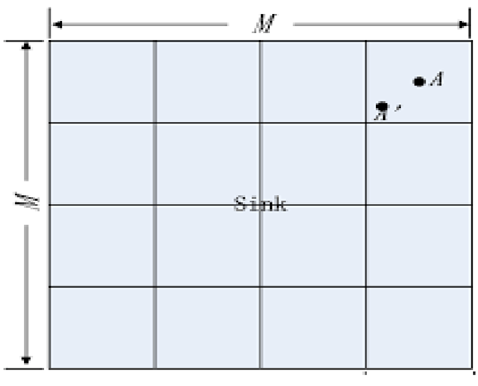

Figure.1 Cluster A uses cooperative node A to send data to the sink node.

Assume that each cluster area is an s*s square. Given the Poisson distribution and density μ, each Cluster contains, on average, μs^2 sensor nodes. Therefore, the number of clusters is $\frac{M}{\mu s2}$.

Once sensor nodes are grouped together to form appropriate number of clusters then every cluster member exclusively transmits sensory data to their respective cluster heads. Now the cluster heads start processing the data received from their clusters and sends it to peripheral nodes near to the sensor nodes, which have adequate remaining energy as shown in Figure 1, where the WSN is divided into 16 partitions turned as 16 clusters. Data from cluster A and it's sink node is transferred using cooperative node.

While narrowing the focus, our analysis explores the energy requirement within a cluster, as shown in Figure 2, specifically using cluster A as an example. Regarding intra-cluster communications, it's noted that cluster members are not in close proximity to the CH, leading researchers to posit that any of the node in the cluster can directly communicate with the CH. Based on Equation (5), the energy exhaustion of cluster members in each cluster is calculated. In Figure 2, the relationship between s and R is defined as s = 2R, where s denotes the cluster length. This allows for the calculation of the expected square distance between the CH and cluster members.

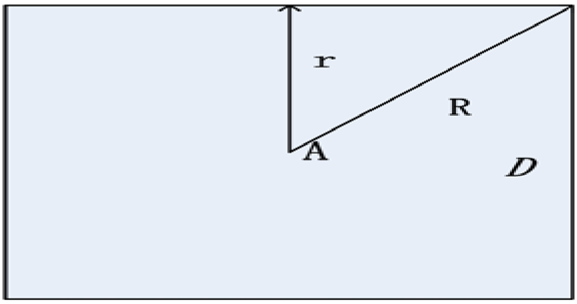

Figure .2 Energy consumption within a cluster.

The total energy loss during single data gathering by cluster A can be computed by

$$E_{cluster}=E_{ch}+(\frac{M}{K}-1)\ E_{nrich} \tag{6}$$

E_{ach} can be determined since the CH does not immediately send the information to another cluster head

$$E_{ch}=(\frac{M}{K}-1)*\ E_{diss}*l+\frac{M}{K}*\ E_{diss}*l*E_{ad}+l*\ E_{diss}+l*s_{ch}^{2}*E_{fs} \tag{7}$$

Where S_{cn} is the separation between the cooperating node and the cluster head. We can obtain depending on Eqs. (5) and (6), where the overall energy required for gathering one round of data is

$$E_{luster}=E_{ch}+\ \frac{M}{K}*E_{nrich}=\mu s2(\frac{M^{2}}{6K}*\ E_{diss}*l+l*E_{lec})+(\frac{M}{K}-1)*\ E_{diss}*l+E_{diss}*l*E_{ad}+l*\ E_{diss}+l*s_{ch}^{2}*E_{fs} \tag{8}$$

$$E_{overall}=k* E_{luster} = \frac{M^2}{s^2} + \mu s2 \left(\frac{M^2}{6K}* E_{diss} *l+l* E_{lec}\right) + k*\left(\frac{M}{K}-1\right)* E_{diss} *l+\frac{M}{K}* E_{diss} *l*E_{ad}+l* E_{diss} + l*s_{ch}^2*E_{fs} \quad (9)$$

Equation (9) is taken to form the ideal number of clusters with respect to k and we initially set it to 0. Therefore, the number of clusters, $k = \sqrt{\frac{M^2}{6s_{ch}^2}}$, can be obtained to get the lowest value of the Total.

4.2 Algorithm for centralized node clustering:

This research presents a strategy utilizing the spectral splitting method for centralized clustering, aiming to segment the network into an optimal configuration. We assume the sink node functions as a central node, grouping all Cluster Heads into k clusters and establishing connections among them.

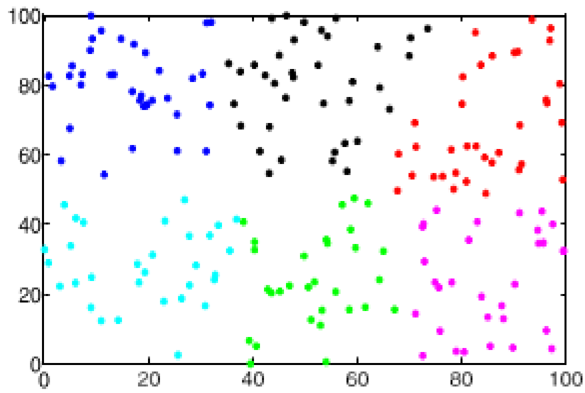

Figure.3 The sensor nodes are partitioned into six groups of suggested approach for spectrum partitioning.

To achieve the ideal bipartitions of a given WSN graph, we employ a two-phase spectral bisection clustering process as detailed in Algorithm 1. This process involves dividing the graph into two sub-networks and recursively applying the same method to the resulting sub-graphs, resulting in two disjoint graphs, G1 and G2, each with a nearly equal number of nodes. Through the spectral partitioning method of clustering, the ideal number of clusters can be identified, k. In general, the spectral partitioning approach efficiently organizes 200 sensor nodes into six clusters, each depicted in a unique color in Figure 3.

Algorithm1: Algorithm for spectral bisection
Required: For Graph G = (e,v); Verify: G1=(e1,v1),G2=(e2,v2);

```
calculate the v-eigenvector of Fiedler;
for each node i in G
if vi<0 then
Partition v1 should contain node i.
I++
else if vi>0, then
Node i should be in partition v2;
i++;
else
end if
end for
```

4.3 Algorithm for distributed node clustering:

In this section, a distributed clustering algorithm is discussed that proved to be more suited to WSNs, relying solely about the information of node's direct neighbors. In our simulations, a performance comparison of distributed algorithm against a centralized one is set as a benchmark. The rationale is that a centralized algorithm, with its access to complete network topology, is expected to achieve optimal performance. Initially, the sink is presumed to have a general understanding of the sensing area but does not require precise locations of the sensor nodes. After applying the fuzzy C-means (FCM) method to segment the field into c cluster zones, each cluster's CH is selected from the range of its sensor network based on geographical cluster center. Subsequently, the chosen CHs broadcast invitations to the sensor nodes, so that they can join their respective clusters.

The sink is tasked with determining the number of cluster areas within a given sensing environment and establishing n number of optimal clusters. To accomplish this, we adopt the clustering method outlined in [34], which divides the survey area into smaller grids, each represented by a virtual node positioned at the center. According to Algorithm 2, the nodes in set V approximate the center points of the sensing field's c cluster areas, where V represents the collection of nodes within these grids.

After identifying its geographical location, the nearest sensor nodes assume the role of Cluster Head. All nodes within a radius r of the center are eligible to vie for the Cluster Head position. They broadcast a message containing their identity and location to participate in the election. Following a designated delay, the candidate positioned nearest to the cluster's center is elected as the Cluster Head.

Once, the selection of a Cluster Head (CH) is done, the CH extends invitations to other sensors to become members of the cluster by broadcasting a specific signal. During this phase, each sensor not designated as a CH identifies and aligns with the cluster led by the nearest CH.

Algorithm 2: Algorithm for distributed node clustering

```
Required
Graph G = (υ,e) is a supporting graph
G’ is the subset of nodes (υ1,e1)
υ’’=e
```

Verify
1. Initialize K clusters
2. For each node j ϵ ʋ1:
For each node belonging to I, node j receives the component ʋij, which is equal to:

$$u_{ij} = \frac{1}{\sum_{k-1}^{c} (\frac{S_{ij}}{S_j})^{2(M-1)}}$$

3. End for

Repeat
1. For j=1 to k:
Compute each cluster's centroid position:

Center = $\frac{\sum_{j-1}^{M} \quad \vartheta_j^M \text{ . Position(mode)}}{\sum_{j-1}^{K} \quad \vartheta_j^M}$

Continue until the cluster does not change

2.End for

For Each Node

1. For I = 1 to M:
Calculate the stoch (distance) between i and ʋ''.
If stoch < r, then transmit a CH election message to i .
Select the node closest to ʋ'' as the cluster head.
2. End for

For Each Cluster

1. For j = 1 to k:
Send promotional message to the sensor field.
Nodes i through n choose the cluster they want to join.
2. End if

5. CH selection and Cooperative nodes

The selection of Cluster Heads (CHs) and the use of cooperative nodes are crucial for enhancing the energy efficiency of Wireless Sensor Networks (WSNs). CH selection is aimed at reducing energy use and extending network life by choosing nodes based on their energy levels, location, and communication costs to manage cluster communications. Cooperative nodes are instrumental in ensuring energy-efficient data transmission throughout the network by aiding in the relay of data using optimized routing protocols. Together, CHs and cooperative nodes significantly boost energy savings, network scale, and performance, leading to more resilient and long-lasting WSNs.

5.1 Method for picking cooperative sensor nodes

The network lifetime is significantly impacted by Cooperative Nodes (CNs) choice. Therefore, it is necessary to develop a suitable selection technique. The sink node periodically sends several signals transmitted at a predetermined level of power to all sensor nodes during in the initialization stage. The messages are picked up by the nodes close to the sink node and Flooding the remainder of the system with them. The sensor nodes with the best potential CNs have obtained enough remaining energy n_a and more messages. Whenever a sensor node transmission receives a transmitter, it adds one to the transmitter counter. The sensor node then determines chance, a probability that it will be chosen as a CN. Using the counter n_a, the transmitter communications' typical transmission power and its remaining energy, specifically, u_{chance}, could be determined by

$$u_{chance}=b1\frac{E_{response}}{E_{Max}}+b2na+b3\frac{\Sigma V}{n_a} \tag{10}$$

The weight factors for the remaining energy, the n_a values, and the average signal intensity of the transmitted transmitter messages are, represented by letters b1, b2, and b3, respectively.

The node v then transmits a potential signal to the Cluster Head (CH)with its identifier and probability value in it. It is more likely that the node with greater probability values will be chosen as CNs. When a CN's energy below a specific threshold, then the CNs is circulated amongst the cluster members.

5.2 Selection of CH:

The use of heuristic clustering and its optimization, particularly for categorizing cluster strength, is well-suited to fuzzy logic due to its ability to operate securely even with imprecise and noisy inputs, showcasing its inherent robustness. The fuzzy logic control model comprises fuzzifiers, fuzzy inference engines, and defuzzifiers. In this research, the likelihood of a node to become a Cluster Head (CH) is assessed using a fuzzy inference system (FIS). The initial input for fuzzy logic, as shown in Table 1, employs a trapezoidal membership function for 'near' and 'far' values, while the sensor node's residual energy is categorized into ranges with triangular membership functions for 'low,' 'medium,' and 'high' energy levels. Table 1 also outlines the fuzzy logic rules based on two input parameters, allowing for the determination of a fuzzy output for the probability. This fuzzy output is then translated into a precise value through defuzzification, a process that calculates the center of area (COA) using the method detailed in Equation 11.

Table .1 Rules for fuzzy mapping

Probability	proximity to CN	Energy of Residual
Very high	Close	High
Medium	Close	High

Very low	Close	Low
Medium	Medium	Medium

$$\text{Output} = \frac{\int x*\mu_{\text{chance}}(x)d(x)}{\int xd(x)} \tag{11}$$

The membership function of a fuzzily defined set of probabilities is denoted by μ_{chance}. A node's likelihood of becoming a Cluster Head (CH)is higher if it is close to the cooperating sensor node and stores more incredible residual energy.

6. Results of Simulation:

The simulation findings are outlined in this section to assess our suggested algorithms. It is presumed that the WSN nodes are dispersed over a 101 m* 101 m area when the network model is established. A central location houses the sink node (51, 51). According to the simulation considerations, the number of sensor nodes changes. Table 2 describes the simulation's input settings. Our approaches are compared to the HEED algorithm in three different ways: the frequency of cycles till the first one node become dead and the number of sensor networks that stay active throughout time, the growth the required energy supply on performance. This is because, like our approach, hybrid energy-efficient distributed algorithm also takes residual energy into account when choosing the cluster head. We disregard the impact of wireless channel interference and transmission collisions.

Table.2 Parameters of Configuration

Value	Parameters
.6j	Energy Initial
51nj/bit	Select
3999bit	Message size
11pj/bit/m^2	Efs
6nj/bit/message	E_d
.044pj/bit/m^2	Eamp

6.1 Measuring the number of rounds for the first node to die:

Maintaining the operational status of all sensor nodes for as long as possible is crucial since the death of a single node can significantly impact network performance. Thus, pinpointing the moment the first node fails is essential. Figures 4 and 5 depict the simulation durations for various comparison techniques and the timing of the first node's failure. Despite differences in network topology in Figure 4, with a constant starting energy and consistent sensor nodes, Figure 5 varies the initial energy levels. This allows us to separately assess the impact of network structures and energy models on our algorithm's efficiency. Figure 4 demonstrates the performance of different clustering strategies as the sensor count N increases from 101 to 301, with our distributed method marking the first node's failure around 701, compared to the hybrid energy-efficient distributed method, which records it at about 601. Figure 5 further highlights the superiority of our methods, especially as the initial energy of sensor nodes is raised from 0.6 J to 1.6 J and then to 2.6 J, showing a more pronounced advantage over the hybrid energy-efficient distribution.

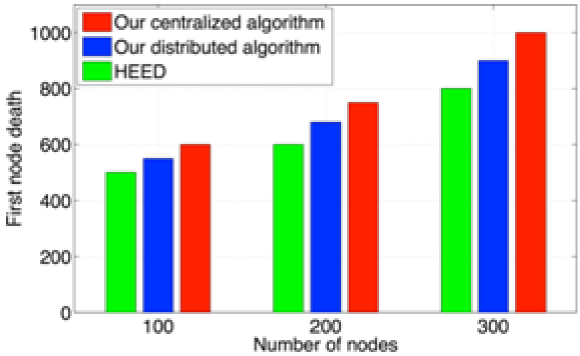

Figure.4 The effect of node density by growing the number of sensors N from 101 to 301

While the hybrid energy-efficient distributed approach may allow for more rounds post the initial node failure compared to our proposed algorithm, enhancing the network's overall lifetime, it is clear from the data that our algorithm performs effectively across varying initial energy levels and node densities. The timing of the first node's failure serves as an indicator of our algorithm's resilience, showcasing the robustness and efficiency of our approach. The optimal number of clusters plays a crucial role in achieving network scalability, longevity, and the strategic positioning of cluster heads. The power model (12) is instrumental in determining the ideal cluster count. By comparing our algorithm with the hybrid energy-efficient distribution, we can identify the most efficient cluster configuration and the strategy of cluster head selection to minimize energy consumption. Our centralized algorithm demonstrates superior performance over the distributed model, as evidenced in Figures 4 and 5, due to the base station's (BS) comprehensive energy levels of all other nodes, optimizing energy conservation for data transmission. Additionally, this methodology simplifies the clustering process by dividing matrices according to their eigenvalues, further reducing energy expenditure during the cluster formation phase.

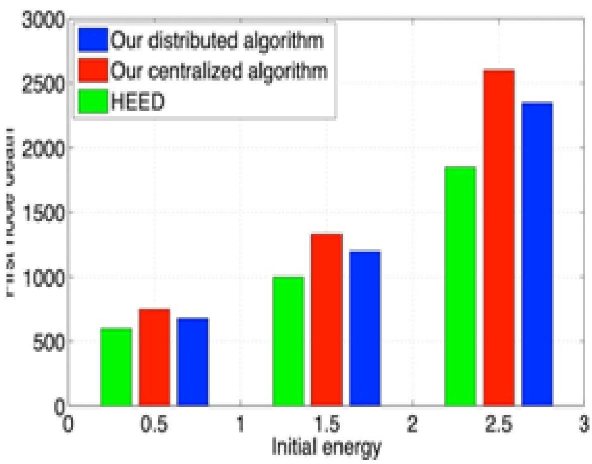

Figure.5 The advantages of our algorithms over hybrid energy-efficient distributed

6.2 Measuring of the nodes' lifetime:

We configured 201 sensor nodes in this section, each with an initial energy of 0.6 J. Figure 6 shows the number of nodes in network over time for each algorithm compared. It is evident that our techniques can increase network lifespan when compared to hybrid energy-efficient distributed. The leftover energy of the exploratory CH determines its random selection in a hybrid energy-efficient distribution.

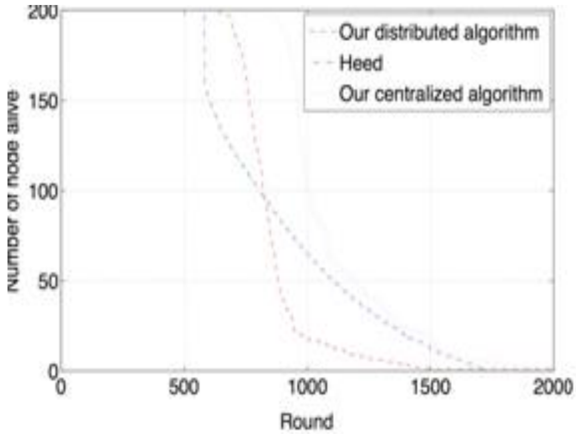

Figure .6 The evolution of the number of active nodes, compared to HEED

As a result, since they choose the ultimate cluster heads based on the cost of intra-cluster communications, the sensor with low residual energy can become CH. CH also consume more energy than it should, which is unbalanced. We employ two ways in our algorithms to regulate the energy consumption of the nodes. Prior to choosing cluster heads depending on FIS, we simply select collaborative nodes to lower the energy usage of CH. Low energy nodes have a decreased probability of selection thanks to cluster creation depending on the residual energy of nodes. As a result, we can effectively extend the network's lifespan.

6.3 Comparing the left-over energy:

The channel's overall residual energy since many data collection rounds is shown in Figure 7. In contrast to our techniques, hybrid energy-efficient distribution rapidly reduces the network residual energy. We may observe that after 1000 data collecting rounds, the hybrid energy-efficient distributed uses about 95% of the amount of energy. Nevertheless, our centralized algorithm just uses 82% of the total energy. Such an increase is related to taking overall distances to the central nodes into account and using a collaborative node selection technique. It effectively lowers the energy used during intra-class and inter-class creation. It is relatively simple to determine the length between central nodes and candidate nodes because we were able to decide on the centralized database of each Cluster before cluster creation. In actuality, the Cluster Head (CH)selection criteria can guarantee that nodes around the center nodes have a better likelihood of becoming CH. The suggested algorithm help distribute the nodes across the various clusters.

Additionally, the cooperative nodes are successfully rotated using the distance to BS remaining energy in our suggested approach. It is evident that CNs can reduce the energy required for inter-class data transfer. The simulation findings show that, when compared to hybrid energy-efficient distributed procedures, our algorithms significantly outperform in terms of network's lifespan and exhausted energy.

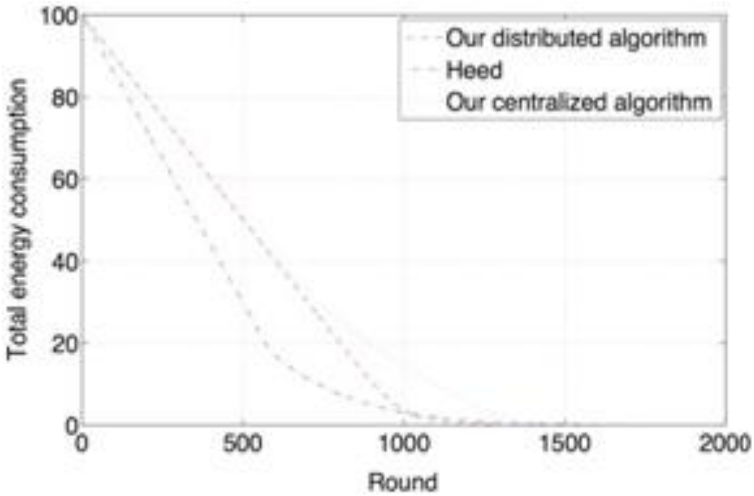

Figure.7 Channel's overall residual energy

7. Conclusions:

The present paper addresses the challenges of big data and proposed a viable solution to increase the lifespan of IoT equipped wireless sensor networks. By utilizing collaborative nodes for efficient data transmission and managing the energy consumption of cluster heads, we assess the energy demands of both intra- and inter-cluster communications through the proposed energy model. Moreover, the paper introduced an alternative algorithm alongside a centralized strategy for cluster formation within the network, thereby enhancing its longevity. The proposed method for selecting collaborative nodes (CN) and cluster heads (CH) is designed to balance energy consumption across sensor nodes effectively.

Simulations reveal that the proposed algorithm markedly enhances network efficiency, yet it is not without its constraints. One limitation is the prerequisite for sensor nodes to be uniformly distributed, with the network topology staying static over time. Additionally, the algorithm assumes homogeneity among nodes, each with self-awareness of their location via RSSI localization. Additionally, the heuristic approach of clustering algorithm may occasionally lead us to clustering inefficiencies.

References

1. A. Bazregar, A. Movaghar, A. Barati, MRE Nejhad, and H. Barati, "A new automatic clustering algorithm via deadline timer for wireless ad-hoc sensor networks," in Proc. 3rd International Conference on Information and Communication Technologies: From Theory to Applications (ICTTA 2008), Damascus, 2008, pp. 1–6.

2. Abdulzahra, A., & Al-Qurabat, A. (2022). "A Clustering Approach Based on Fuzzy C-Means in Wireless Sensor Networks for IoT Applications." Karbala International Journal of Modern Science, vol. 8, pp. 579-595, doi: 10.33640/2405-609X.3259.
3. F. Ahamad and R. Kumar, "Energy efficient region-based clustering algorithm for WSN using fuzzy logic," in Proc. 2016 IEEE International Conference on Recent Trends in Electronics, Information Communication Technology (RTEICT), 2016, pp. 1020–1024.
4. F. AL-Obaidy, H. Zereshkian, and FA Mohammadi, "An energy-efficient routing algorithm in Zigbee-based cluster tree wireless sensor networks," in Proc. 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE), Windsor, 2017, pp. 1–5.
5. G. Asada, M. Dong, TS Lin, F. Newberg, G. Pottie, and WJ Kaiser, "Wireless integrated network sensors: low power systems on a chip," in Proc. 24th European Solid-State Circuits Conference, The Hague, 1998, pp. 9–16.
6. HE Alami and A Najid, "CFL: Cluster formation using fuzzy logic for wireless sensor networks," in Proc. 2015 IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA), Marrakech, 2015, pp. 1–6.
7. Hussein, A., & Abdulrazzak, H. (2024). "Fuzzy C-Means Clustering Approach for Green IoT in Smart Cities." pp. 588-593, doi: 10.1109/ICPC2T60072.2024.10474897.
8. Lahmar, A. Zaier, M. Yahia, J. Lloret, and R. Bouallegue, "Optimal data transmission for decentralized IoT and WSN based on Type-2 Fuzzy Harris Hawks Optimization," Internet of Things, vol. 25, 2024, Art. no. 101028, ISSN 2542-6605. [Online]. Available: https://doi.org/10.1016/j.iot.2023.101028.
9. J. Lee and W. Cheng, "Fuzzy logic-based clustering approach for wireless sensor networks using energy predication," IEEE Sensors J., vol. 12, no. 9, pp. 2891–2897, 2012.
10. JC Bezdek, "Pattern recognition with fuzzy objective function algorithms," Adv. Appl. Pattern Recog., vol. 22, no. 1171, pp. 203–239, 1981.
11. JK Parker, LO Hall, and JC Bezdek, "Comparison of scalable fuzzy clustering methods," in Proc. 2012 IEEE International Conference on Fuzzy Systems (IEEE FUZZ), Brisbane, 2012, pp. 1–9.
12. JS Lee and TY Kao, "An improved three-layer low-energy adaptive clustering hierarchy for wireless sensor networks," IEEE Internet Things J., vol. 3, no. 6, pp. 951–958, 2016.
13. Kumar, P. S., & Barkathulla, A. (2023). "Fuzzy clustering and optimal routing in wireless sensor network." J. Intell. Fuzzy Syst., vol. 45, pp. 6863-6873. [Online]. Available: https://api.semanticscholar.org/CorpusID:259915656.
14. LB Bhajantri and AV Sutagundar, "Fuzzy logic-based Cluster Head (CH) selection and data processing in distributed sensor networks," in Proc. 2016 International Conference on Computing, Analytics and Security Trends (CAST), Pune, 2016, pp. 282–288.
15. Liu, S. Guo, Y. Shi, and Y. Yang, "Deterministic binary matrix based compressive data aggregation in big data WSNs," Telecommunication Systems, vol. 66, no. 3, pp. 345–356, 2017.
16. Loganathan, S., Arumugam, J., & Chinnababu, V. (2021). An energy-efficient clustering algorithm with self-diagnosis data fault detection and prediction for wireless sensor networks. Concurrency and Computation:. https://doi.org/10.1002/cpe.6288.
17. M Li, L Zhang, Z Xiang, E Castillo, and T Guerrero, "An improved fuzzy C-means algorithm for brain MRI image segmentation," in Proc. 2016 International Conference on Progress in Informatics and Computing (PIC), Shanghai, 2016, pp. 336–339.
18. M. S. Bhatti, N. Saeed, and H. Nam, "Fuzzy C-Means Clustering and Energy Efficient Cluster Head Selection for Cooperative Sensor Network," Sensors, vol. 16, no. 9, p. 1459, 2016. [Online]. Available: https://doi.org/10.3390/s16091459
19. Manjeshwar and D.P. Agrawal, "Teen: a routing protocol for enhanced efficiency in wireless sensor networks," in Proc. 15th IEEE International Conference on Parallel and Distributed Processing Symposium, San Francisco, 2001, pp. 2009–2015.
20. Manjeshwar and D.P. Agrawal, "Umpteen: a hybrid protocol for efficient routing and

comprehensive information retrieval in wireless," in Proc. International Conference on Parallel and Distributed Processing Symposium (IPDPS 2002), Ft. Lauderdale, 2002, pp. 8–19.
21. M-C Hung and D-L Yang, "An efficient fuzzy C-means clustering algorithm," in Proc. IEEE International Conference on Data Mining (ICDM 2001), San Jose, 2001, pp. 225–232.
22. M-W Park, J-Y Choi, Y-J Han, and T-M Chung, "An energy efficient concentric clustering scheme in wireless sensor networks," in Proc. INC, IMS and IDC, 2009. NCM '09. Fifth International Joint Conference On, Seoul, 2009, pp. 58–61.
23. Naderloo, S. A. Fatemi Aghda, and M. Mirfakhraei, (2023) "Fuzzy-based cluster routing in wireless sensor network," Soft Comput., vol. 27, no. 10, pp. 6151-6158, Mar. 20, doi: 10.1007/s00500-023-07976-6.
24. P. Nayak and B. Vathasavai, "Energy efficient clustering algorithm for multi-hop wireless sensor network using type-2 fuzzy logic," IEEE Sensors J., vol. 17, no. 14, pp. 4492–4499, 2017.
25. R. Bensaid, M. Ben Said, and H. Boujemaa, (2020)"Fuzzy C-Means based Clustering Algorithm in WSNs for IoT Applications," in International Wireless Communications and Mobile Computing (IWCMC), 2020, pp. 126-130. [Online]. Available: https://api.semanticscholar.org/CorpusID:220887061
26. S. Cheng, Z. Cai, J. Li, H. Gao, "Extracting kernel dataset from big sensory data in wireless sensor networks," IEEE Trans. Knowl. Data Eng., vol. 29, no. 4, pp. 813–827, 2017.
27. Sharmin, S.; Ahmedy, I.; Md Noor, R. An Energy-Efficient Data Aggregation Clustering Algorithm for Wireless Sensor Networks Using Hybrid PSO. Energies 2023, 16, 2487. https://doi.org/10.3390/en16052487.
28. Singh, S., Anand, V., (2023) "Load balancing clustering and routing for IoT-enabled wireless sensor networks" Journal of Network Management, Volume33, Issue5,September, https://doi.org/10.1002/nem.2244.
29. Tumula, S., Ramadevi, Y., Padmalatha, E., Kumar, G. K., Gopalachari, M. V., Abualigah, L., Chithaluru, P., & Kumar, M. (2023). An opportunistic energy-efficient dynamic self-configuration clustering algorithm in WSN-based IoT networks. International Journal of Communication Systems. https://doi.org/10.1002/dac.5633.
30. V. Cherappa, T. Thangarajan, S. S. Meenakshi Sundaram, F. Hajjej, A. K. Munusamy, and R. Shanmugam, (2023) "Energy-Efficient Clustering and Routing Using ASFO and a Cross-Layer-Based Expedient Routing Protocol for Wireless Sensor Networks," Sensors (Basel), vol. 23, no. 5, Art. no. 2788, Mar. 2023, doi: 10.3390/s23052788. PMID: 36904993; PMCID: PMC10007087.
31. X. Ci, K. Wen, Y. Sun, W. Sun, and W. Deng, "An Energy Efficient Clustering Algorithm in Wireless Sensor Networks for Internet of Things Applications," Journal of Physics: Conference Series, vol. 1881, no. 4, Art. no. 042035, Apr. 2021. [Online]. Available: https://dx.doi.org/10.1088/1742-6596/1881/4/042035.
32. Yousif, Z.; Hussain, I.; Djahel, S.; Hadjadj-Aoul, Y. A Novel Energy-Efficient Clustering Algorithm for More Sustainable Wireless Sensor Networks Enabled Smart Cities Applications. J. Sens. Actuator Netw. 2021, 10, 50. https://doi.org/10.3390/jsan10030050.
33. Bhatti, D. M. S., Saeed, N., & Nam, H. (2016). Fuzzy C-Means Clustering and Energy Efficient Cluster Head Selection for Cooperative Sensor Network. Sensors, 16, 1459. https://doi.org/10.3390/s16091459