



Gaussian Mixture Model-Based Clustering for Energy Saving in WSN

Mohammed Saleh Mutar¹ , Dalal Abdulmohsin Hammood¹ ,
and Seham Ahmed Hashem² 

¹ Electrical Engineering Technical College, Department of Computer Technical Engineering,
Middle Technical University (MTU), Al Doura, 10022 Baghdad, Iraq

dalal.hammood@mtu.edu.iq

² Technical Instructors Training Institute, Middle Technical University, Baghdad, Iraq

dr.seham.ahmed@mtu.edu.iq

Abstract. In wireless sensor networks, nodes have limited access to energy sources and must make efficient use of what they have. Energy consumption may be decreased and network life can be prolonged via the process of clustering. To reduce the network's power consumption and increase its lifespan, we used a new clustering technique in this work. Centralized cluster formation and decentralized cluster heads form the basis of this stage of clustering. Clusters are determined via a centralized Gaussian mixture model (GMM) technique, and once they are generated, they don't change. After that, it chooses which cluster heads (CHs) should spin. Inside those clusters to minimize energy consumption prior to the data transmission phase to the base station (BS), taking into account the varying quantities of energy in the nodes. Thus, the proposed approach not only effectively addresses the energy consumption problem, but also significantly lengthens the lifespan of the network. The results demonstrate the following ways in which the suggested method lessens the burden on network resources. It increases network lifetime by 301%, 131%, and 122%, decreases energy consumption by 20.53%, 6.14%, and 5%, and increases throughput by 47%, 9%, and 4% when compared to the Flat, FUCA, and FCMDE protocols.

Keywords: Wireless sensor networks · Gaussian mixture model · clustering · energy consumption

1 Introduction

Wireless sensor networks, which are deployed in big numbers to collect data about their surroundings and send it to a central location, are characterized by their low cost, small size, and constrained resource availability [1]. Numerous applications, including habitat these networks are used for surveillance, border surveillance, healthcare surveillance, etc. The wireless sensor nodes are often placed in an unfriendly or unmanaged environment. What's worse is that these nodes only have basic connection, computing, storage, and battery capabilities [2].

The battery that powers the remainder of the system's components (processing, sensing, receiving, and sending) is defined by its compact size and its constraints as a consequence of the sensor nodes' small size. Certain circumstances make it difficult, costly, or impossible to replace or recharge the battery [3]. In these networks, sensor nodes are placed densely to form data reading vectors that are geographically and temporally coupled. Multiple network resources are required for the processing and transmission of these vectors. The whole network's resources are impacted. Substantially while processing and sending these data vectors. These duplicate data vectors' transmission causes a number of issues for the network, including high bandwidth usage, energy consumption, and a number of overhead expenses related to data storage, processing, and communication [4].

Furthermore, a number of internal node processes, including sensing, processing, and data transfer, might negatively impact the sensor node's performance. The process that uses the most energy is data transmission [5]. As a result, it is necessary to prolong the lifespan of long-term applications like continuous monitoring systems. However, the pace of creating data for base station processing is often quite high. Energy dissipation reduction is a major issue in WSNs [6].

One practical solution for dealing with these problems and using the energy at hand is clustering. This is caused by clustering, which divides the network into clusters and requires each cluster's sensor nodes (SNs) to relay data to a cluster head (CH) [7]. Due to the sensors' close proximity to the CHs, they may reduce their transmission powers, which would save energy and lengthen the network's lifetime. CHs are chosen from the SNs to handle collecting data from sensors in their clusters, putting it all together, and sending it to the BS [8]. The popular, adaptable, and effective Gaussian-based mixture models (GMM) are used to describe both univariate and multivariate data. They've been put to use in a variety of applications, including machine learning, voice and image processing, pattern recognition, computer vision, and statistical data analysis. Using a limited mixture of Gaussian densities, it may handle issues like data analysis and grouping [9].

Similar to how k-means may be used to classify data, Gaussian mixture models can organize sensor nodes into groups. However, Gaussian mixture models provide several benefits that k-means cannot. In the first place, k-means does not include variation. The spread of a normal distribution, measured in terms of its variance, is what we mean when we talk about variance. The k-means model may be understood as if it were a set of circles, with the farthest distant point in each cluster defining the radius of the circle. When the sensor is circular, it functions as expected. Alternatively, Gaussian mixture models are capable of accommodating very elongated clusters. The second distinction is that k-means conducts hard classification whereas Gaussian mixture models do soft classification.

What follows is an explanation of the rest of the paper. The related works are included in Sect. 2. In Sect. 3, we provide a quick summary of paradigm of networks and energy use model. In Sect. 4, we cover the proposed procedure in detail. Discussions and results from the simulations are reported in Sect. 5. The last section of the article provides an overview of the main points.

2 Related Works

By introducing a model into a wireless sensor system, the energy efficiency of wireless systems will be increased. While discussing design difficulties and practical constraints, an effort is made to build an application-oriented system to enhance the functionality of wireless sensor networks while reducing their energy consumption.

Gupta et al. (2014) proposed a Gaussian mixture model (GMM), which is a collection of mixtures of multivariate Gaussian distributions and is a suitable model for clusters of various sizes that are correlated with one another. The GMM's clustering measures the associated posterior probability for each node and explains how each node connects to each cluster, i.e., the mean of the corresponding nodes. The clustering carried out by GMM is known as "soft clustering," since nodes are not limited to a single cluster [10].

Tsiligaridis et al. (2016) proposed a stochastic approximation (SA)-based distributed EM algorithm, which targets problems with sensor networks' dispersed clustering transmission cost minimization. Each node in the network in our configuration perceives an in which the world may be modeled as a collection of Gaussians, each of which corresponds to one of the basic needs. The distributed clustering issue is studied in terms of a whereby all of reality may be represented by a scattering of Gaussians, one for each of our fundamental desires. By reducing the need for network cycles and keeping calculations and communications local, The use of DEM-SA in a WSN reduces both traffic and contention [11].

Hojjatnia et al. (2021) offered a new method, namely GDECA. Which applies the premise that the distributions of nodes in the actual world are mixtures of the Gaussian distribution. Therefore, So That We Can Find out the Parameters of These Distributions by Fitting the Gaussian Mixture Model (GMM) To the Nodes, GDECA uses a distribution estimation technique that it has adopted from machine learning (ML). Additionally, the dispersion of nodes is used to calculate sinks' routing [12].

Al-Janabi et al. (2022) to enhance ES and lengthen the lifetime of sensor nodes, a k-means clustering strategy was presented. This method clusters the region of interest, which decreases the distance between the sensor nodes and the base station. Data is sent from each node to the cluster head, which in turn relays the information to the BS. Energy efficiency is improved, as well [13].

Chaubey et al. (2016) proposed a new hierarchical clustering algorithm. Some of the nodes in the proposed process must choose cluster heads that are further from the BS than they are. These nodes transmit their data to a different place, where it must travel a great distance before reaching the BS. These transmissions are referred to as "excess transmissions" and waste energy on the network. In the suggested approach, set up every sensor node in a distributed cluster environment and decide how many clusters to place there. Each sensor chooses one of the cluster's head nodes, and each cluster has its own cluster head, which communicates with the base station (BS) for communication purposes [14].

Moghadaszadeh et al. (2017) proposed a new algorithm in which expectation maximization (EM) is used. Make a suggestion for a fix to the issues the K-Means family, which is employed in many clustering algorithms, is now experiencing. The suggested approach, it uses the EM clustering technique as its foundation, increases network longevity and improves energy efficiency. Clustering is accomplished using a Gaussian mixture model using the EM technique. During the setup phase and seeks to build clusters based on the EM technique as it is specified for Gaussian mixture models (GMM). The GMM is used by the EM method to find clusters, and it gives the most likely parameters for each cluster [15].

Pancha et al. (2017) suggested a hierarchical low-energy clustering algorithm (LEACH). The first WSN clustering method is called LEACH. By using a clustering strategy with one CH in each cluster, LEACH lowers network power consumption / use. Once the sensor nodes are set up, the clustering process may begin. Here, the remaining network nodes choose the CH with the strongest received signal, which is the nearest CH. After selecting all of the CHs in the network at first using a probability-dependent threshold. Each and every CH served as a router for direct communication with the BS. Distributed clustering is carried out via LEACH [16].

Agrawal, D. and Pandey, S., (2018) proposed an approach called the “method for uneven clustering using fuzzy logic” to do just that: extend the lifespan of the network. Uneven clusters are formed using this procedure. The goal is to equalize heating and cooling needs. Fuzzy logic is used to determine which nodes in a cluster should serve as leaders. The density, the remaining energy, and Inputs consist of distance from the station’s home base. The two resulting fuzzy variables are the radius of the competition and the ranking. To do fuzzy inference, the Mamdani technique is used [17].

Abdulzahra, A. and Al-Qurabat (2022) introduced a novel fuzzy c-means based clustering methodology with distance- and energy-limited termed (FCMDE) for clustering to increase the lifespan of WSN. Although FCMDE uses the fuzzy c-means approach to cluster SNs, it does not choose the node closest to the fuzzy c-means centroid as the CH but instead selects the node closest to the majority of nodes in the network. The closeness criterion ensures that all nodes in a given cluster remain in close proximity to their CH, allowing them to maintain drastically reduced transmission powers [18].

3 Preliminaries

In this part, we detail the energy usage and network model.

3.1 Network Model

In this part, we provide a common monitoring environment for applications based on WSNs. To ensure the system’s low power consumption, we adopt a cluster-based layout. A square sensing field with N randomly spaced sensor nodes surround the BS. The nodes continually assess their surroundings and communicate their findings to the CH, who then periodically sends the information obtained to the BS (also known as the gateway (GW)). For our network model, we presumptively consider the following:

1. Due to the homogenous network that we have taken into consideration for the proposed study it is assumed that all of the nodes have the same initial energy and are hence static.
2. Based on the suggested packet routing scheme, each CH gets each CH sends data packets to its corresponding GW, and each GW receives data packets from its corresponding CMs. And a small number of CMs. The data aggregation procedure will thus be carried out by these CH and GHs for effective transmission and to save network energy.
3. Using the proposed method, all sensor nodes engage in single-hop communication.
4. All sensor node has a fixed initial energy and is energy-limited.
5. The GW should be unrestricted by energy, computation, and network coverage.

3.2 Energy Model

Energy is required by sensor nodes for a variety of purposes, including sensing, network maintenance, data processing, packet receipt, and packet transfer. The distance traveled and the size of the packet determines how much energy is needed to convey it [19, 20]. To broadcast a packet of k – bits across a distance of d , the transmitter has to expend a certain amount of energy, as follows:

$$E_{Tx}(k, d) = \begin{cases} k \times E_{elec} + k \times \epsilon_{fs} \times d^2 & \text{if } d < d_0 \\ k \times E_{elec} + k \times \epsilon_{mp} \times d^4 & \text{if } d \geq d_0 \end{cases} \quad (1)$$

Receiving a k – bits packet consumes the following amount of energy:

$$E_{Rx}(k) = k \times E_{elec} \quad (2)$$

E_{elec} in (1) and (2) stands for the energy spent per bit by the transmitter or reception circuits, respectively. We use ϵ_{fs} and ϵ_{mp} , transmission and receiving circuit power consumption, respectively, to characterize the energy expenditure of the amplifier for each bit in two different models: free space and multi-path fading. The letter d represents the separation between the transmitter and receiver. The d_0 threshold is formulated as having

$$d_0 = \sqrt{\epsilon_{fs} / \epsilon_{mp}} \quad (3)$$

Another component that is taken into account is the data aggregation power consumption, or E_{da} . We assume that each cluster member transmits k – bits to its CH during each data collection period, and that the energy used by a CH during one data collection period may be represented as

$$E_{CH} = \frac{N}{c} \times E_{elec} \times w + \frac{N}{c} \times E_{da} \times w + \epsilon_{mp} \times w \times d_{BS}^2 \quad (4)$$

Energy is wasted by the CH when it gathers packets from nodes, aggregates them, and sends the resultant packets to the BS. Provided is the typical separation between a CH and a BS by d_{BS} , while the number of clusters is given by c .

4 The Proposed Method

The suggested procedure may be carried out in three distinct stages. Selecting an appropriate cluster number is the first Phase. During Phase 2, a centralized clustering method is suggested using the Gaussian mixture model. The last stage involves data transmission between cluster nodes and CHs. Figure 1 shows the flowchart of the proposed system.

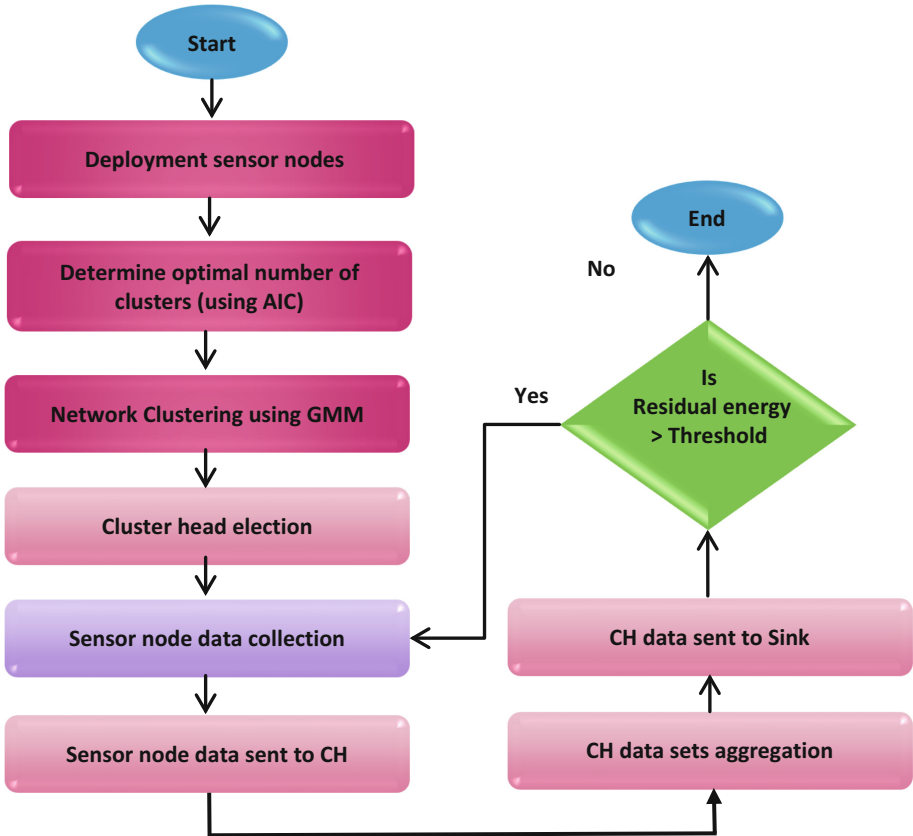


Fig. 1. The flowchart of the proposed system

4.1 Optimal Number of Clusters in a Gaussian Mixture

For clustering, a precise estimate of importance of cluster count. The quantity of clusters may be derived from the data and used as additional parameter. Akaike’s information criterion (AIC) and the Bayesian inference criterion (BIC). Is used to determine the ideal number of clusters [21, 22]. The AIC is

$$AIC(G) = -2\ln L(\theta(G)) + 2P(G) \tag{5}$$

$L(\theta(G))$ is the likelihood value computed at $\theta(G)$, and P is the total number of parameters to be evaluated. The vector containing the parameters' greatest likelihood estimates, in which G represents the Gaussian density function. The estimated C is the quantity of clusters (abbreviated C) with the lowest AIC value.

4.2 Gaussian Mixture Models Clustering

Several methods exist for clustering d -dimensional data sets into a predetermined size (say C). Popular clustering techniques like as model-based clustering and K -means may complete this task using a Gaussian mixture model. As previously discussed, these clustering algorithms may be divided into two categories: soft clustering algorithms and hard clustering algorithms. Mixture models make use of the probabilistic soft clustering technique. Data points provide samples from each cluster's probability distribution, which is represented as a cluster in d -dimensional space. Gaussian Mixture Models assume that each clusterable data point is selected simultaneously from a set of distributions whose parameters are unknown and is thus a mixture of Gaussian distributions. To determine the values of these unknowable factors and then create the various clusters, a learning method is used [23].

According to Eq. (6), the probability distribution $p(X)$ of a node in a network (which is denoted by a vector X) is the weighted sum of the probability distributions of the node in each of the node's component C clusters. The distribution of each component (denoted by $N(X|\mu_C, \Sigma_C)$) is a cluster represented by a Gaussian.

$$p(x) = \sum_{C=1}^C \pi_C N(X|\mu_C, \Sigma_C) \quad (6)$$

In Eq. (6), π_C is the coefficient of mixing for cluster C , which is one of C clusters; μ_C is the mean of the normal distribution for cluster C and Σ_C is the normal distribution's covariance measure for cluster C . The degree of a node's relationship with cluster C is indicated by the mixture coefficient π_C . The parameters that constitute a multivariate normal distribution that represents a cluster are mean and covariance. The value of variance or standard deviation is used in place of covariance for single-variable normal distributions.

They are probabilistic mixture models. Data samples are created using GMM clustering models. Each data point in these models, albeit to variable degrees, is a member of every cluster in the dataset. Being a part of a certain cluster has a chance of between 0 and 1, the actions listed below are done.

1. Set the starting values for μ , Σ and the mixing coefficient π , and then calculate L , the logarithm of the likelihood.
2. Analyze the accountability procedure with the current settings
3. Obtain new μ , Σ , and π using newly acquired obligations
4. Log-likelihood L should be calculated once again. Iterate through steps 2 and 3 until convergence is reached.

Since covariance and mean are both taken into account while creating clusters, GMM will not make any errors. These factors influenced our choice to use the GMM clustering method for the suggested approach.

In order to reduce the amount of energy needed to create clusters, clustering is performed before CH selection. A node's residual energy has to be greater than a threshold in order for it to be considered for CH selection. To avoid premature death and network disconnection, this criterion is essential. Additionally, the CH is selected as the node that is closest to the largest number of other nodes. The suggested approach does not prioritize picking the cluster's epicenter node above those that are farther out since doing so would waste energy.

5 Transmission of Data

When the CHs are recognized, the sensor nodes start sending data to them. The transmission power of nodes in a cluster is decreased because the Gaussian mixture modeling technique clearly achieves the shortest geographic distance to the CHs. The CHs lower the quantity of data by aggregating it, and then they transfer the resulting data to the BS.

6 Simulation and Performance Evaluation

The recommended strategy is simulated in Python. In order to show how well the proposed method works in simulations, a scenario is developed. 100 sensor nodes are selected for a $100 \times 100 M^2$ network. When first deployed, the BS is often located centrally inside the network as shown in Fig. 2. The settings for the simulation are shown in Table 1 below. The efficiency of the suggested approach is evaluated in comparison to that of both flat and clustered networks.

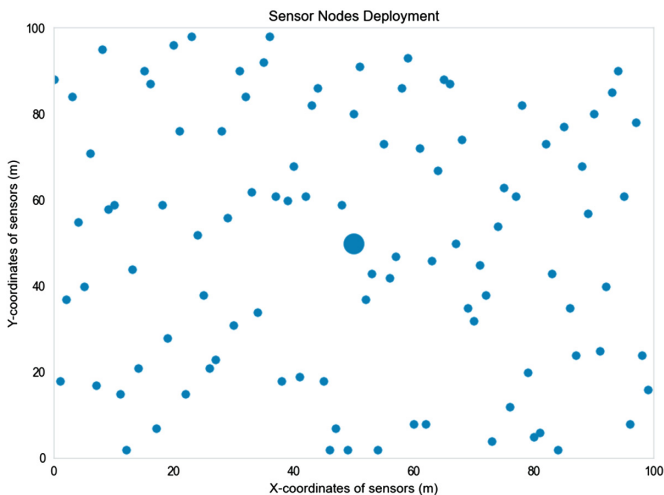


Fig. 2. Deploying sensor nodes in the target area.

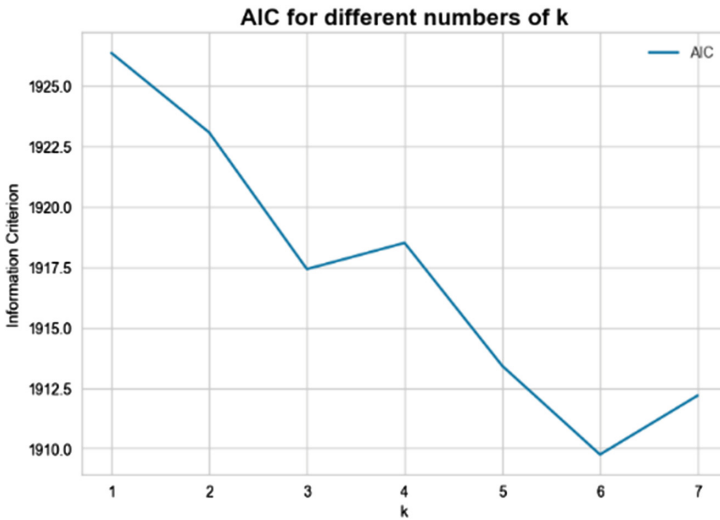
Since the number of clusters must be specified in advance for the GMM technique to work, we chose 6 as shown in Fig. 3 and calculated from Eq. 5. Our proposed method

Table 1. Parameters of Simulation.

Parameters	Values
<i>Network size</i> (m^2)	100×100
<i>Nodes deployment</i>	<i>Randomly</i>
<i>BS location</i>	<i>Center</i>
<i>Nodes count</i>	100
<i>Initial energy</i>	0.5 J
<i>Data packet</i>	100 readings * 64 bits
E_{elec}	50 nJ/bit
ϵ_{fs}	10 pJ/bit/ m^2
ϵ_{mp}	0.0013 pJ/bit/ m^4
E_{DA}	5 nJ/bit/signal
d_0	87 m

clusters 100 sensor nodes into six groups, each of which is illustrated by a distinct color (see Fig. 4).

The network's reliability is measured during a time interval called the stability period. Rounds till the first network node are tallied during the stability phase. (FND) completely loses power and is declared dead. When even a single node fails, the whole network is put at risk. The lifespan of a network is the number of cycles it takes for all of the nodes to run out of power.

**Fig. 3.** AIC score for optimal No. of clusters.

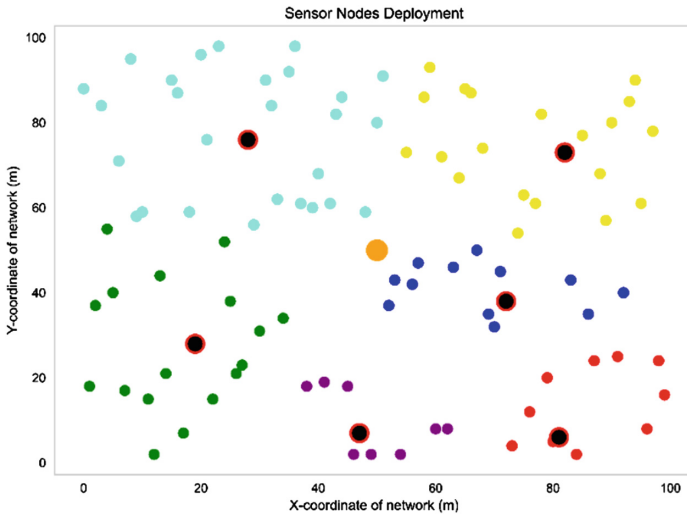


Fig. 4. Creation of Clusters using GMM.

The findings of a simulation research showing that the stability period for flat networks is 177 rounds and for clustered networks it is 534 rounds. The results suggest that the location of the BS and the overall sensor count deployed across the network are two crucial aspects that influence the lifetime of the sensor nodes (i.e., the density of the network). Clustering the nodes and placing the BS in the center of the network shortens the path data must travel, which in turn reduces the amount of energy needed to digest the data and extend the network's life.

The average amount of power used by nodes in the network will be the focus of our next investigation. The amount of energy a WSN uses is one of the most crucial metrics to evaluate. Figure 5 and 6 show the energy used by each SN in a flat network and a clustered network after 1000 rounds of the experiment.

For all scenarios, energy usage increases with network flatness. Where the energy required to get data packets to their destination depends largely on how far they must travel. We can see that after 1000 iterations of data collection, the clustered network uses less than 283% as much energy as the flat network.

The effects of dense sensor node deployment on network energy consumption and first node death (i.e. network stability) are studied. Table 2 below analyzes the results for average energy usage for 1000 rounds throughout the complete network to emphasize the impact on big and small networks. The simulation runs with a variable number of nodes, typically between 200 and 500. The impact of GW distance and network density on energy consumption is seen in all cases. As we have shown by analysis of the findings, and in accordance with Table 3, clustered networks need much less overall network energy consumption each round than flat networks. This research proves that clustering has a major impact on networks of all sizes.

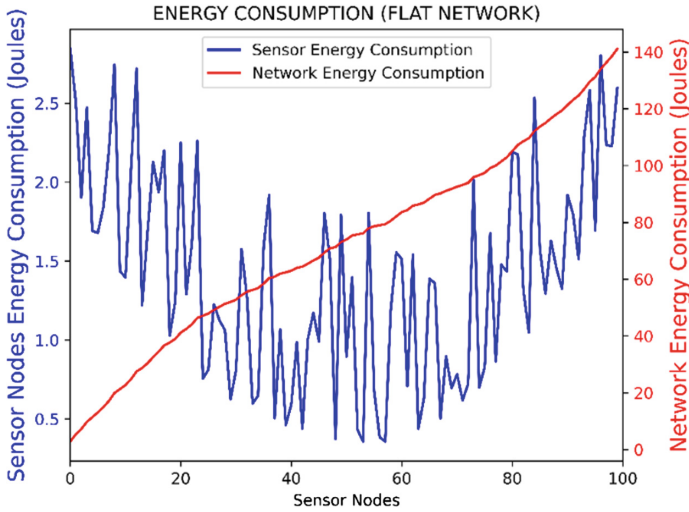


Fig. 5. The energy consumption of sensor nodes use in flat WSN.

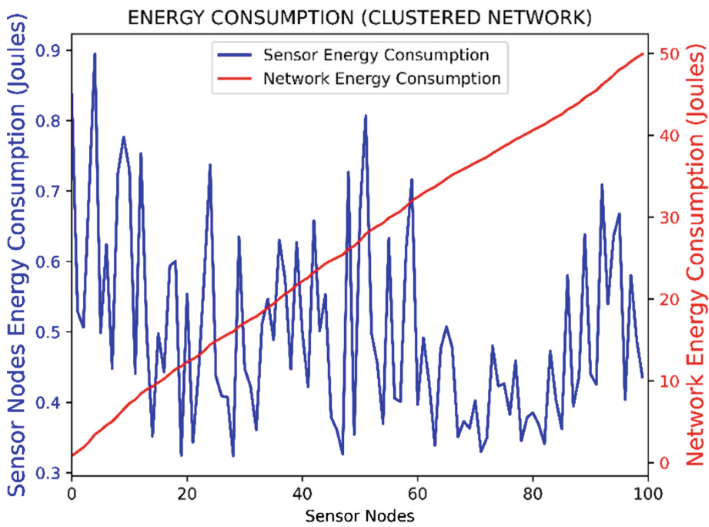


Fig. 6. The energy consumption of sensor nodes in clustered WSN.

Table 3 shows that the clustered network runs for more iterations than the flat network does. As a result of the shortened distance that data from sensors must travel, this helps to preserve the battery life of Increase the longevity of the network as a whole and the sensor nodes.

Table 2. The network energy consumption.

	Sensor Nodes No			
	200	300	400	500
Flat	924.7366	3612.09	9483.586	19834.87
Clustered	138.8371	272.6742	429.3441	613.3011

Table 3. The network stability (First Node Death).

	First Node Dead			
	200	300	400	500
Flat	34	14	8	5
Clustered	242	167	171	151

7 Comparison with Other Protocols

Furthermore, a comparative study using contemporary procedures, such as FUCA [17], and FCMDE [18], verified the efficacy of the proposed GMM. We evaluated the proposed GMM to the standard protocols in terms of network lifetime, stability, and the proportion of active to inactive nodes.

7.1 Lifetime Evaluation

It is crucial that as many sensor nodes as possible stay up for as long as possible, since node failures degrade overall network performance. Therefore, it is crucial to know when the first node will die. Network lifespan is measured from the moment when the first node in the network stops functioning.

In Fig. 7, we can see how GMM stacks up against the simulated outcomes from FUCA and FCMDE. We found that the lifetime of the recommended network has been greatly extended in comparison to other efforts; this is because our study takes into account both energy and distance. In Fig. 7, we can see that in comparison to the Flat, FUCA, and FCMDE protocols, the recommended GMM protocol improves first-SN by about 301%, 131%, and 122%.

7.2 Energy Consumption

The average Energy content that is wasted across the network is measured using the GMM protocol in this experiment. One of the most crucial criteria for evaluating a WSN's performance is its energy consumption. Energy consumption for the flat, FUCA, and FCMDE methods is also shown for comparison with the proposed GMM protocol in Fig. 8.

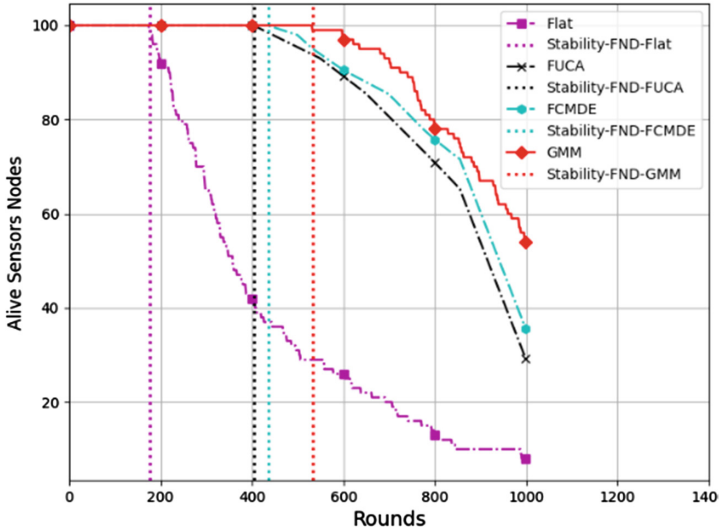


Fig. 7. FND stability time and lifespan of the network.

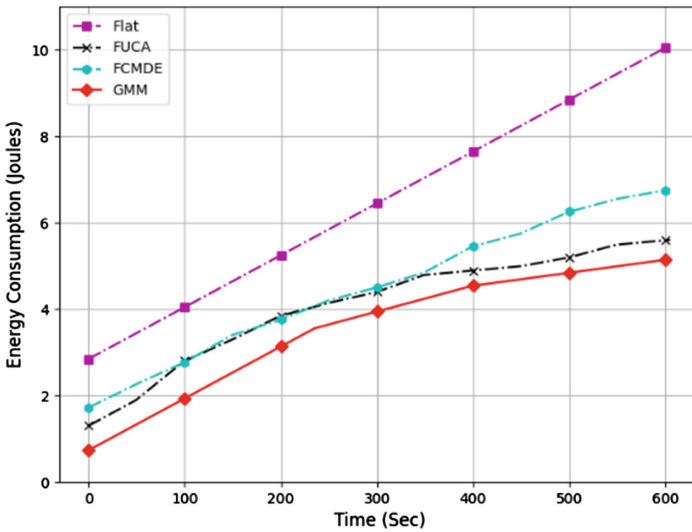


Fig. 8. The network's energy utilization level.

The experimental findings show that there was a decrease in energy use across all time periods. Compared to the flat, FUCA, and FCMDE protocols, the GMM protocol reduces energy consumption during data transmission by about 301%, 131%, and 122%, respectively, after 1000 rounds of the experiment. According to the results, the GMM procedure outperforms the other two and has the highest rate of energy savings.

7.3 Throughput Evaluation

To further evaluate the network's throughput, another simulated experiment was run. During transmission, throughput is measured as the proportion of acknowledged packets to the time spent communicating those packets between the CH and the sender.

$$\text{Throughput} = \frac{\text{total number of received packets by CH}}{\text{delay in process of communication}} \quad (7)$$

Throughput comparisons between the Flat, FUCA, and FCMDE protocols and the proposed GMM protocol are shown in Fig. 9. The proposed GMM protocol sends more packets to the CH in less time than the Flat, FUCA, and FCMDE protocols by margins of 47%, 9%, and 4%. Consequently, throughput measurement has evolved through time in comparison to earlier approaches.

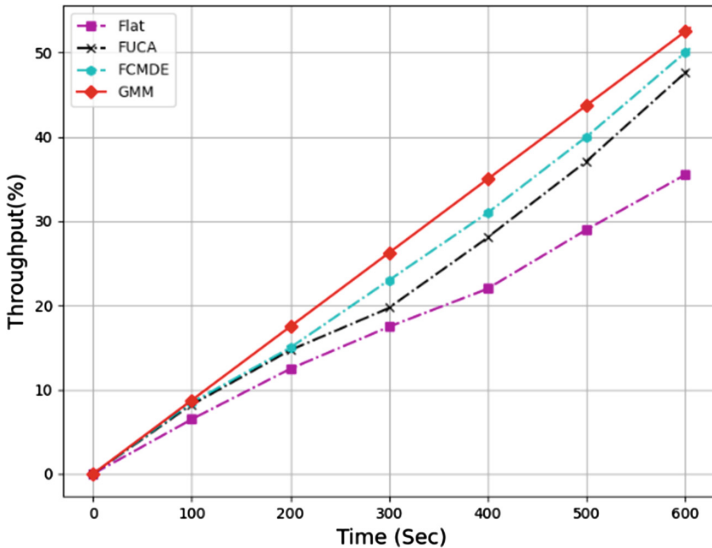


Fig. 9. A measure of the network's throughput.

8 Conclusions

In this research, a new clustering methodology based on Gaussian mixture models (GMM) for WSN was presented. The proposed approach minimizes expenses while maximizing efficiency and extending service life. During clustering, the approach selects a CH using a combination of GMM, node location, and residual power. The proposed method's efficacy has been shown via extensive simulation employing a wide range of evaluation performance metrics. One example is the average amount of energy used, while others include the stability and lifespan of the network. The suggested method

was shown to be better by a comparative study of current techniques. Future projects, In order to choose the cluster head, we want to utilize an optimization approach. Also, we plan to propose a new method for scheduling the sensor node's work based on the spatial correlation.

References

1. Saeedi, I.D.I., Al-Qurabat, A.K.M.: An energy-saving data aggregation method for wireless sensor networks based on the extraction of extrema points. In: Proceeding of the 1st International Conference on Advanced Research in Pure and Applied Science (Icarpas2021): Third Annual Conference of Al-Muthanna University/College of Science, vol. 2398, no. 1, p. 050004 (2022)
2. Abdulzahra, S.A., Al-Qurabat, A.K.M.: Data aggregation mechanisms in wireless sensor networks of IoT: a survey. *Int. J. Comput. Digit. Syst.* **13**(1), 1–15 (2023)
3. Al-Qurabat, A.K.M., Abdulzahra, S.A.: An overview of periodic wireless sensor networks to the internet of things. In: IOP Conference Series: Materials Science and Engineering, vol. 928, no. 3, p. 32055 (2020)
4. Saeedi, I.D.I., Al-Qurabat, A.K.M.: A systematic review of data aggregation techniques in wireless sensor networks. In: *Journal of Physics: Conference Series*, vol. 1818, no. 1, p. 12194 (2021)
5. Al-Qurabat, A.K.M., Mohammed, Z.A., Hussein, Z.J.: Data traffic management based on compression and MDL techniques for smart agriculture in IoT. *Wirel. Pers. Commun.* **120**(3), 2227–2258 (2021)
6. Al-Qurabat, A.K.M.: A lightweight Huffman-based differential encoding lossless compression technique in IoT for smart agriculture. *Int. J. Comput. Digit. Syst.* **11**(1), 117–127 (2021)
7. Panchal, A., Singh, R.K.: EHCR-FCM: energy efficient hierarchical clustering and routing using fuzzy C-means for wireless sensor networks. *Telecommun. Syst.* **76**(2), 251–263 (2020). <https://doi.org/10.1007/s11235-020-00712-7>
8. Naeem, A., Gul, H., Arif, A., Fareed, S., Anwar, M., Javaid, N.: Short-term load forecasting using EEMD-DAE with enhanced CNN in smart grid. In: Barolli, L., Amato, F., Moscato, F., Enokido, T., Takizawa, M. (eds.) WAINA 2020. AISC, vol. 1150, pp. 1167–1180. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-44038-1_107
9. Najar, F., Bourouis, S., Bouguila, N., Belghith, S.: A comparison between different Gaussian-based mixture models. In: 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA), pp. 704–708. IEEE (2017)
10. Gupta, S., Bhatia, V.: GMMC: Gaussian mixture model based clustering hierarchy protocol in wireless sensor network. *Int. J. Sci. Eng. Res. (IJSER)* **3**(7), 2347–3878 (2014)
11. Tsiligaridis, J., Flores, C.: Reducing energy consumption for distributed em-based clustering in wireless sensor networks. *Procedia Comput. Sci.* **83**, 313–320 (2016)
12. Houriya, H., Mohsen, J., Saeedreza, S.: Correction to: improving lifetime of wireless sensor networks based on nodes' distribution using Gaussian mixture model in multi-mobile sink approach. *Telecommun. Syst.* **77**(1), 269 (2021)
13. Al-Janabi, D.T.A., Hammood, D.A., Hashem, S.A.: Extending WSN life-time using energy efficient based on K-means clustering method. In: Chaubey, N., Thampi, S.M., Jhanjhi, N.Z. (eds.) COMS2 2022. CCIS, vol. 1604, pp. 141–154. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-10551-7_11
14. Chaubey, N.K., Patel, D.H.: Energy efficient clustering algorithm for decreasing energy consumption and delay in wireless sensor networks (WSN). *Energy* **4**(5), 8652–8656 (2016)

15. Moghadaszadeh, M., Shokrzadeh, H.: An efficient clustering algorithm based on expectation maximization algorithm in wireless sensor network. In: 10th International Conference on Innovations in Science, Engineering, Computers and Technology (ISECT 2017) Dubai (UAE), pp. 19–25 (2017)
16. Engineering, T., Panchal, A., Singh, A.K.: LEACH based clustering technique in wireless sensor network. *Test Eng. Manag.* **82**, 4185–4188 (2020)
17. Agrawal, D., Pandey, S.: FUCA: fuzzy-based unequal clustering algorithm to prolong the lifetime of wireless sensor networks. *Int. J. Commun. Syst.* **31**(2), e3448 (2018)
18. Abdulzahra, A.M.K., Al-Qurabat, A.K.M.: A clustering approach based on fuzzy C-means in wireless sensor networks for IoT applications. *Karbala Int. J. Mod. Sci.* **8**(4), 579–595 (2022)
19. Bagci, F.: Energy-efficient communication protocol for wireless sensor networks. *Ad-Hoc Sens. Wirel. Netw.* **30**(3–4), 301–322 (2016)
20. Wang, N., Zhu, H.: An energy efficient algorithm based on LEACH protocol. In: 2012 International Conference on Computer Science and Electronics Engineering, vol. 2, pp. 339–342. IEEE (2012)
21. Liu, Z., Song, Y.-Q., Xie, C.-H., Tang, Z.: A new clustering method of gene expression data based on multivariate Gaussian mixture models. *SIViP* **10**(2), 359–368 (2015). <https://doi.org/10.1007/s11760-015-0749-5>
22. Kim, H.-J., Cavanaugh, J.E., Dallas, T.A., Foré, S.A.: Model selection criteria for overdispersed data and their application to the characterization of a host-parasite relationship. *Environ. Ecol. Stat.* **21**(2), 329–350 (2013). <https://doi.org/10.1007/s10651-013-0257-0>
23. Vashishth, V., Chhabra, A., Sharma, D.K.: GMMR: a Gaussian mixture model based unsupervised machine learning approach for optimal routing in opportunistic IoT networks. *Comput. Commun.* **134**, 138–148 (2019)