(12) PATENT APPLICATION PUBLICATION
(19) INDIA
(22) Date of filing of Application :27/06/2023

(21) Application No.202341043083 A
(43) Publication Date : 01/09/2023

(54) Title of the invention : A MACHINE LEARNING BASED SYSTEM TO OPTIMIZED CLUSTER HEAD SELECTION IN WIRELESS SENSOR NETWORKS AND METHOD THEREOF

(51) International classification	:H04W0084180000, H04W0052020000, H04W0040320000, G06N0020000000, H04W0040100000
(86) International Application No Filing Date	:PCT// :01/01/1900
(87) International Publication No	: NA
(61) Patent of Addition to Application Number Filing Date	:NA :NA
(62) Divisional to Application Number Filing Date	:NA :NA

(71)**Name of Applicant :**
1)Andhra University
Address of Applicant :Visakhapatnam, Andhra Pradesh, India. Pin Code: 530003 ----------- -----------
Name of Applicant : NA
Address of Applicant : NA
(72)**Name of Inventor :**
1)Prof.James Stephen Meka
Address of Applicant :Dr.B.R. Ambedkar Chair Professor, Dean, A.U. TDR-HUB, Andhra University, Visakhapatnam, Andhra Pradesh, India ----------- -----------
2)Mr.Seelamanthula Sreenivasu
Address of Applicant :Research Scholar, Department of Computer Sciences & Systems Engineering, A. U. College of Engineering, Andhra University, Visakhapatnam, Andhra Pradesh, India ----------- -----------

(57) Abstract :
[055] The present invention discloses a machine learning based system to optimized cluster head selection in wireless sensor networks and method thereof. In the present invention, the system comprises an Energy focussed protocol designed in which initially the nodes are divided into clusters and for each cluster, a cluster head shall get selected. Thereafter data has been transmitted between various nodes to base station and the energy consumption is computed. In this paper, Clustering is done using various machine learning algorithms and the energy consumption results are compared. Accompanied Drawing [FIGS. 1-7]

No. of Pages : 29 No. of Claims : 2

“FORM 1 THE PATENTS ACT 1970 (39 of 1970) and THE PATENTS RULES, 2003 **APPLICATION FOR GRANT OF PATENT** (See section 7, 54 and 135 and sub-rule (1) of rule 20)	(FOR OFFICE USE ONLY)

	Application No.	
	Filing date:	
	Amount of Fee paid:	
	CBR No:	
	Signature:	
1. APPLICANT'S REFERENCE / IDENTIFICATION NO. (AS ALLOTTED BY OFFICE)		

2. TYPE OF APPLICATION [Please tick (✓) at the appropriate category]

Ordinary (✓)		Convention ( )		PCT-NP ( )	
Divisional ( )	Patent of Addition ( )	Divisional ( )	Patent of Addition ( )	Divisional ( )	Patent of Addition ( )

3A. APPLICANT(S)

Name in Full	Nationality	Country of Residence	Address of the Applicant
Andhra University	Indian	India	Visakhapatnam, Andhra Pradesh, India. Pin Code: 530003

3B. CATEGORY OF APPLICANT [Please tick (✓) at the appropriate category]

Natural Person ()	Other than Natural Person		
	Small Entity (✓)	Startup ()	Others ()

4. INVENTOR(S) [Please tick (✓) at the appropriate category]

Are all the inventor(s) same as the applicant(s) named above?	Yes ()	No (✓)

If "No", furnish the details of the inventor(s)

Name in Full	Nationality	Country of Residence	Address of the Inventor
1. Prof.James Stephen Meka	Indian	India	Dr.B.R. Ambedkar Chair Professor, Dean, A.U. TDR-HUB, Andhra University, Visakhapatnam, Andhra Pradesh, India
2. Mr.Seelamanthula Sreenivasu	Indian	India	Research Scholar, Department of Computer Sciences & Systems Engineering, A. U. College of Engineering, Andhra University,

			Visakhapatnam, Andhra Pradesh, India

5. TITLE OF THE INVENTION

"A MACHINE LEARNING BASED SYSTEM TO OPTIMIZED CLUSTER HEAD SELECTION IN WIRELESS SENSOR NETWORKS AND METHOD THEREOF"

6. AUTHORISED REGISTERED PATENT AGENT(S)	IN/PA No.	
	Name	
	Mobile No.	
7. ADDRESS FOR SERVICE OF APPLICANT IN INDIA	Name	Prof.James Stephen Meka
	Postal Address	Dr.B.R. Ambedkar Chair Professor, Dean, A.U. TDR-HUB, Andhra University, Visakhapatnam, Andhra Pradesh, India
	Telephone No.	
	Mobile No.	9542354100
	Fax No.	
	E-mail ID	reenivasusept30@gmail.com jamesstephenmm@gmail.com

8. IN CASE OF APPLICATION CLAIMING PRIORITY OF APPLICATION FILED IN CONVENTION COUNTRY, PARTICULARS OF CONVENTION APPLICATION

Country	Application Number	Filing date	Name of the applicant	Title of the invention	IPC (as classified in the convention country)

9. IN CASE OF PCT NATIONAL PHASE APPLICATION, PARTICULARS OF INTERNATIONAL APPLICATION FILED UNDER PATENT CO-OPERATION TREATY (PCT)

International application number	International filing date

10. IN CASE OF DIVISIONAL APPLICATION FILED UNDER SECTION 16, PARTICULARS OF ORIGINAL (FIRST) APPLICATION

Original (first) application No.	Date of filing of original (first) application

11. IN CASE OF PATENT OF ADDITION FILED UNDER SECTION 54, PARTICULARS OF MAIN APPLICATION OR PATENT

~~Main application/patent No.~~	~~Date of filing of main application~~

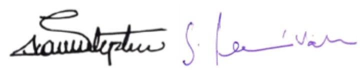
12. DECLARATIONS

i) Declaration by the inventor(s)

(**In case the applicant is an assignee**: the inventor(s) may sign herein below or the applicant may upload the assignment or enclose the assignment with this application for patent or send the assignment by post/electronic transmission duly authenticated within the prescribed period).

I/We, the above named inventor(s) is/are the true & first inventor(s) for this Invention and declare that the applicant(s) herein is/are my/our assignee or legal representative.

(a) Date 26/06/2023

(b) Name	(c) Signature
1. Prof.James Stephen Meka 2. Mr.Seelamanthula Sreenivasu	

~~**(ii) Declaration by the applicant(s) in the convention country**~~

~~**(In case the applicant in India is different than the applicant in the convention country:** the applicant in the convention country may sign herein below or applicant in India may upload the assignment from the applicant in the convention country or enclose the said assignment with this application for patent or send the assignment by post/electronic transmission duly authenticated within the prescribed period)~~

~~I/We, the applicant(s) in the convention country declare that the applicant(s) herein is/are my/our assignee or legal representative.~~

~~(a) Date~~

~~(b) Signature(s)~~

(c) ~~Name(s) of the signatory~~

(iii) Declaration by the applicant(s)

I/We the applicant(s) hereby declare(s) that: -

- ☐ ~~I am/~~ We are in possession of the above-mentioned invention.
- ☐ The ~~provisional~~/complete specification relating to the invention is filed with this application.
- ~~☐ The invention as disclosed in the specification uses the biological material from India and the necessary permission from the competent authority shall be submitted by me/us before the grant of patent to me/us.~~
- ☐ There is no lawful ground of objection(s) to the grant of the Patent to me/us.
- ☐ ~~I am~~/we are the true & first inventor(s).
- ~~☐ I am/we are the assignee or legal representative of true & first inventor(s).~~
- ~~☐ The application or each of the applications, particulars of which are given in Paragraph-8, was the first application in convention country/countries in respect of my/our invention(s).~~
- ~~☐ I/We claim the priority from the above mentioned application(s) filed in convention country/countries and state that no application for protection in respect of the invention had been made in a convention country before that date by me/us or by any person from which I/We derive the title.~~
- ~~☐ My/our application in India is based on international application under Patent Cooperation Treaty (PCT) as mentioned in Paragraph-9.~~
- ~~☐ The application is divided out of my /our application particulars of which is given in Paragraph-10 and pray that this application may be treated as deemed to have been filed on DD/MM/YYYY under section 16 of the Act.~~
- ☐ ~~The said invention is an improvement in or modification of the invention particulars of which are given in Paragraph-11.~~

13. FOLLOWING ARE THE ATTACHMENTS WITH THE APPLICATION

(a) Form 2

Item	Details	Fee	Remarks
Complete/ ~~Provisional~~ specification) #	No. of pages: 23		
No. of Claim(s)	No. of claims: 02 No. of pages: 02		
Abstract	No. of pages: 01		
No. of Drawing(s)	No. of drawings: 07 No. of pages: 03		

In case of a complete specification, if the applicant desires to adopt the drawings filed with his provisional specification as the drawings or part of the drawings for the complete specification under rule 13(4), the number of such pages filed with the provisional specification are required to be mentioned here.

(b) Complete specification (in conformation with the international application)/as amended before the International Preliminary Examination Authority (IPEA), as applicable (2 copies).

(c) Sequence listing in electronic form

(d) Drawings (in conformation with the international application)/as amended before the International Preliminary Examination Authority (IPEA), as applicable (2 copies).

(e) Priority document(s) or a request to retrieve the priority document(s) from DAS (Digital Access Service) if the applicant had already requested the office of first filing to make the priority document(s) available to DAS.

(f) Translation of priority document/Specification/International Search Report/International Preliminary Report on Patentability.

(g) Statement and Undertaking on Form 3

(h) Declaration of Inventorship on Form 5

(i)Power of Authority

(j)**Total fee ₹……….in Cash/ Banker's Cheque /Bank Draft bearing No.......... Date on …………. Bank.**

I/We hereby declare that to the best of my/our knowledge, information and belief the fact and matters slated herein are correct and I/We request that a patent may be granted to me/us for the said invention.

Dated this 26th day of June 2023

Applicant: Andhra University

To,

The Controller of Patents

The Patent Office, at Chennai

Note: -

* Repeat boxes in case of more than one entry.
* To be signed by the applicant(s) or by authorized registered patent agent otherwise where mentioned.
* Tick ()/cross (x) whichever is applicable/not applicable in declaration in paragraph-12.
* Name of the inventor and applicant should be given in full, family name in the beginning.
* Strike out the portion which is/are not applicable.
* For fee: See First Schedule”;

FORM 2

THE PATENTS ACT, 1970

(39 of 1970)

&

The Patent Rules, 2003

COMPLETE SPECIFICATION

(See section 10 and rule 13)

TITLE OF THE INVENTION

"A MACHINE LEARNING BASED SYSTEM TO OPTIMIZED CLUSTER HEAD SELECTION IN WIRELESS SENSOR NETWORKS AND METHOD THEREOF"

Applicant:

Andhra University,

Visakhapatnam, Andhra Pradesh, India. Pin Code: 530003.

The following specification particularly describes the nature of the invention and the manner in which it is performed:

FIELD OF THE INVENTION

[001] The present invention relates to the field of the using AI and data analytics with novel techniques, methods, devices and apparatus. The invention more particularly relates to a machine learning based system to optimized cluster head selection in wireless sensor networks and method thereof.

BACKGROUND OF THE INVENTION

[002] The following description provides the information that may be useful in understanding the present invention. It is not an admission that any of the information provided herein is prior art or relevant to the presently claimed invention, or that any publication specifically or implicitly referenced is prior art.

[003] Further, the approaches described in this section are approaches that could be pursued, but not necessarily approaches that have been previously conceived or pursued. Therefore, unless otherwise indicated, it should not be assumed that any of the approaches described in this section qualify as prior art merely by virtue of their inclusion in this section.

[004] In recent years, the emergence of novel applications for the Wireless Sensor Network (WSN) technology has led to an increase in their prevalence. Usually, WSNs are those networks having power from a limited capacity battery. Battery swaps or recharging may not mostly be possible due to the reason that there shall be numerous sensors in the network and manually recharging every node is highly impossible. At this juncture it is to mention that the entire wireless sensor network depends on the energy constraints. The energy consumption must be optimal and there must be an efficient strategy that reduce consumption. One strategy that aids in reducing energy consumption is a competent cluster head selection technique which enhances lifetime of a network. The focus of this paper is towards machine learning techniques in cluster head selection. Three techniques k means, hierarchical and Gaussian mixture clustering techniques are of the subject and among these Gaussian Mixture technique seemed to be an effective clustering technique.

[005] The term "Wireless Sensor Network" (WSN) refers to a collection of sensor nodes connected to one another in wireless communication mode. There are a wide variety of WSN applications, some of which are well established while others are in their infancy. Recent years have seen the rise of sensor networks as a viable infrastructure for numerous crucial monitoring and command functions. Sensor networks differ from mobile ad hoc networks (MANETs) in many ways, including their restricted mobility, capabilities, and dense deployment. In order to solve issues such as routing, medium access control, bandwidth allocation, security and self-organization that have swamped wireless networks, novel energy-efficient methods must be developed. To keep a network operational for long time, it's helpful to use strategies like hierarchical (tiered) topologies and the exploitation of trade-offs among latency, energy, and accuracy.

[006] As there is a difficulty in dealing all transactions in one – one mode between base station and the sensor network nodes, clustering process plays a vital role in preserving the efficiency of any WSN. Selection of a few heads from among the network's nodes and the subsequent clustering of the other nodes around these heads is the primary objective of the clustering process.

The cluster heads must coordinate the activities of their nodes and aggregate data from them (intra-cluster coordination), as well as represent their clusters in communications with other clusters and outside observers (inter-cluster communication). Sensors periodically report their findings to an off-site monitor (such as a base station). Clustering could be used in single-hop as well as multi-hop networks. Restructuring clusters on a regular basis allows for the selection of nodes with the largest amount of leftover energy to take on the role of cluster leads.

[007] The majority of clustering applications use K-means clustering. It is the simplest technique of unsupervised learning and is based on the concept of centroid. The strategy seeks to consolidate data inside a cluster. Due to its iterative nature, K-means works best with less number of nodes. In case of more number of nodes are to be clustered the process shall take long time. Density-based spatial cluster analysis (DBSCAN) is a technique for categorising data based on density, as opposed to the k-means algorithm. This method does well when looking for any outliers. It locates clusters of arbitrary forms by using the number of data points that are present in different locations. To identify anomalies among the clusters of high density, it divides the space into high-density and low-density sectors.

[008] Not all types of data could be processed by all the algorithms. Every technique has its own significance in clustering the network nodes. Multiple Gaussian distributions are used in the Gaussian mixture model to suit irregularly shaped data. This multilayer model is composed of multiple independent Gaussian sub models. The model determines where to place a data point based on the chance that it follows a certain Gaussian distribution.

In comparison to the k-means technique, BIRCH (Balance Iterative Reducing and Clustering utilising Hierarchies) performs better on huge data points. It takes the raw data points and transforms them into clusters of summaries. The summaries provide as much detail about the data points' distributions as feasible. BIRCH is often used with other clustering algorithms since the summaries it produces may be used as inputs to those other algorithms.

[009] When compared to previous clustering algorithms, this technique groups data in an entirely new manner. Clusters in the data become apparent when each data point shares its degree of similarity with every other data point. Ordering Points in the Identification of Cluster Structure (OPTICS) is another technique which is very much useful when the densities are fluctuating. Hierarchical clustering algorithm is useful when the data points or the nodes are to be hierarchically clustered.

Problem Specification:

[010] In data transmission, power requirements are proportional to data size and transmission distance. Normal sensor nodes and cluster head sensor nodes are the two main types of sensor nodes used in a WSN. Normal sensor nodes provide data to respective CH, whereas CH is required to send the data either to the sink or to the other CH. The vital research in addressing the problem throughout the years has been in the form of efficient techniques proposed one after another. As Machine learning and deep learning techniques are adding extraordinary flavour in several domains, in this research exploring ML techniques that suit efficient clustering is the major problem which is going to be addressed in this research. Also the energy efficiency computation of the identified techniques is part of the work to be done.

In short, the planned work addresses the following issues:

1) Appropriate clustering and choice of CH.

2) Effective packet routing.

3) Energy efficiency computation

[011] In the next section the contributions of other researchers are discussed and in the third section proposed work is elaborated. In the fourth section results are discussed and the final section concludes the paper.

Related Contributions:

[012] Mandli et.al. [1] proposed a novel algorithm namely EECHIGWO-based clustering protocol, an improved version of GWO, so as to address the original GWO's problems which are a result of uneven exploitation and exploration, insufficient population variety, and too-rapid convergence. Average intra cluster distance, CH balancing factor, average throughput, number of rounds of operation, energy consumption, and number of failed nodes are all metrics for the proposed EECHIGWO. Minimising energy usage, preventing early convergence, and extending the lifespan of WSN networks are all validated by the simulation findings. When compared to the FIGWO, SSMOECHS, HMGWO, LEACH-PRO, and FGWSTERP protocols, the proposed method improves network stability by 169.29%, 19.03%, 253.73%, 307.89%, and 333.51%.

[013] Nabavi et.al. [2] Combined the multi-objective evolutionary algorithm with the gravitational search technique which seems to be a unique optimisation approach that the authors contributed towards wireless sensor networks clustering techniques. Routing is performed using the Gravitational search algorithm besides the multi objective genetic algorithm for clustering. This seemed to be a better combination which resulted performance in terms of energy efficiency, data delivery rate, and information packet transmission rate.

[014] Kusla et.al. [3] in their proposed AVOA-based energy-efficient CH selection technique, considered to prove the longevity of the network to be an essential component of a WSN. Monitoring the amount of energy used is a challenging process. Their technique with a fitness function that takes into consideration intra-cluster distance, residual energy, and cluster compactness is well presented and the results have been compared to the results acquired using other, more well-known methodologies, such as ACO, ABC, WHO, GTO, ASO, HS, PSO, and BBO. The efficacy of proposed technique is well experimented using two WSN scenarios. Their experimentation resulted not only in the energy usage but also the increase in lifetime of the network. The future research insights they had given are other elements affecting in parallel with residual energy, sensor range and gearbox energy.

[015] Syed Asif Ali et.al. [4] in their work, used K-medoid clustering approach, with a focus placed on the SeNs' energy reduction. Second, a CH is chosen by utilising a weighted cluster head selection approach. This method involves striking a balance between the energy of the source node (SeN), the distance between the source node and the sink node (SN), and the distance between the source node and the CH. A constant weight is assigned to each node, and this weight is based on the node's energy as well as its distance from the SN and the cluster core. Their observation seems to be that the relative energy efficiency (RE) of the sensor nodes (SeN) is very necessary for the continued operation of wireless sensor networks (WSNs). The findings of their simulation work, when compared to the approaches proposed earlier indicate an increase in the lifetime of the proposed network.

[016] Sengathir et.al. [5] proposed a hybrid algorithm HMABCFA which is a combination of Artificial Bee Colony and Firefly algorithms. Exploration of the potentiality of WSNs based on their frequency range is possible because of their ability to achieve data transmission between them. Recharging sensors is a complex process, especially in challenging environments. Short-lived networks, limited coverage, inflexible schedules, and a lack of centralised data storage are some of WSN's biggest problems. The recommended HMABCFA utilises the benefit of the Firefly optimisation approach in order to develop a new position that might potentially replace the position that isn't updated during the earlier phase of ABC. The clustering process is made better by including the Firefly optimisation technique into the ABC algorithm. This gets rid the issues of premature convergence, slow convergence, and being stranded at a local point of optimality. The findings of the HMABCFA include the network's lifetime enhancement by 23.21%, energy stability raise by 19.84%, and network latency decrease by 22.88%.

[017] Panimalar et.al. [6] in their investigation combined Sparrow search technique and the Differential Evolution approach to solve the energy efficiency issue in selection of cluster heads in wireless sensor networks. They made the individual nodes to be optimally utilised by combining high level search efficiency using sparrow search methodology. This hybrid model seems to be optimising the alive/dead node count, throughput, and residual energy to the greatest extent possible in order to reach such outstanding outcomes. In terms of both residual power and throughput, the recommended Improved Sparrow search approach resulted in better performance than other algorithms proposed earlier. This method uses the Differential evolution model to choose the best possible cluster head.

[018] Tay et.al. [7] in their research, suggested a novel clustering technique for WSNs which played a crucial role in saving power and lasted longer. The Cluster Centred Cluster Head Selection Algorithm (C3HA), designed to achieve this goal, offers an incredible outcome in selecting CH while also producing a more efficient WSN than the widely used PEGASIS and LEACH clustering algorithms. When compared to other common algorithms, this one seemed to be faster and more accurate in choosing a CH. C3HA has a much higher residual energy compared to other algorithms. The results proved to be commendable demonstrating that, on average, there are 14.68 percent more surviving nodes in the network using the suggested approach. There were 187.5% more packets than with LEACH. It is possible to determine whether or not a node is awake or sleep based on the distance that separates it from other nodes. Their recommendation is to make use of algorithms that are based on machine learning to speed up the CH selection process.

[019] Panchal et.al. [8] proposed a protocol “Energy Aware Distance-based Cluster Head selection and Routing” (EADCR) aimed at improving the lifetime of the nodes consequently the WSN’s longevity. An innovative clustering approach is devised in which the CH selection is based on a novel fitness function. This is because the nodes use a significant amount of power while the clustering phase is in progress. In addition to the above clustering technique, a unique packet-routing system is used to reduce the amount of energy needed for communication between CHs by using the shortest-path approach. The proposed approach not only enhances network longevity but also helps in energy saving and reducing the unnecessary movement on the network.

[020] Kale et.al. [9] To improve network longevity and power savings, their research proposed a novel cluster head selection mechanism which is a combination of three algorithms - Fitness-based Glowworm Swarm Optimisation (GSO), the Fruitfly Optimisation Algorithm (FFOA), and Fruitfly Algorithm (FGF). The proposed technique outperformed GOA, FFOA, GALLF, CS, GSO, and ABC by a factor of 90.08%, 63.37%, 58.64%, 68.02%, 90.84%, and 72.67% in terms of the number of surviving nodes.

[021] Priyadarshini et.al. [10] The proposed MSDS-MI system demonstrates a maximum decrease in network size of up to 50%, an improvement in network longevity of up to 60%, and a reduction in maximum residual energy consumption of up to 47.6%. When compared to other approaches such as Pseudo Dominating Sets (PDS), Connected Dominating Sets (CDS), Distributed Self-Healing Approach (DSHA), Dynamic Cluster Head Genetic Algorithm (DCH-GA), the recommended strategy shows more promise in terms of its ability to provide desirable results.

[022] Liang Zhao et.al. [11] A modified form of the LEACH cluster-head selection approach referred to as LEACH-M was proposed which uses a dynamic method known as distributed address assignment method (DAAM) that ZigBee uses. Node’s residual energy, network addresses and other such vital parameters are taken into account in improving the cluster head selection threshold equation. In addition, by using a cluster-head competitive mechanism, LEACH-M was able to successfully balance the energy load of the network and significantly improve the efficiency.

[023] Jin-Gu Lee et.al. [12] A spider monkey optimisation and energy-efficient cluster head selection (SSMOECHS) technique which is based on random samples was proposed to extend the life and maintain network integrity in WSNs. Methods for selecting cluster leads using sampling-based SMO including their use in basic SMO are outlined.

[024] Pawan Singh Mehra et.al [13] in their research, provided a fuzzy-based balanced cost CH selection algorithm (FBECS) that takes into consideration the residual energy, the distance from the sink, and the node density in the region. The potential CH function of each node is evaluated by the eligibility index of that node. This protocol is able to provide load balancing by picking the most qualified candidate for the position of cluster coordinator. It does this by taking into consideration the probabilities that have been supplied for each sensor node. When compared to those of its equivalents, BCSA and LEACH, the performance of FBECS has been empirically proven as being better due to its longer time of greater stability, longer lifespan with load balancing, and considerable information forwarding to sink.

[025] Ashwin. R. Jadhav et.al. [14] presented WOA-Clustering (WOA-C), an energy-efficient cluster head selection method inspired by the Whale Optimisation method (WOA). The technique proposed helps in choosing energy-aware cluster heads based on a fitness function that takes into consideration the node's residual energy as well as the sum of the energies of its neighbours. This helps to ensure that the cluster heads have a lower overall energy consumption.

[026] Puneet Azad et.al. [15] Choosing competent cluster heads is a significant task with WSNs. The method proposed uses fuzzy logic to identify the CH. A fuzzy multiple attribute decision-making (MADM) technique is utilised in selection of CHs. This method had taken into consideration the residual energy, the distance of the nodes and the number of neighbours, from the base station. The results of the simulation demonstrate that in contrast to the distributed hierarchical agglomerative clustering (DHAC) protocol, this strategy is preferable when it comes to prolonging the lifetime of the network in environments that include a mix of different types of nodes.

[027] Ossama Younis et.al. [16] proposed a technique HEED (Hybrid Energy-Efficient Distributed clustering) which is an effective Cluster head selection algorithm in which the CHs are selected at regular intervals. The proposed technique employs a parameter combination of the residual energy at each node and nodes' degrees. HEED completes in O(1) iterations, has a little message cost, and leads to a very even distribution of cluster heads over the network. The simulation results simulations show that the suggested method is successful in extending the lifespan of networks and allowing for scalable data aggregation.

[028] Heinzelman et.al. [17] proposed a motivational technique namely LEACH (Low-Energy Adaptive Clustering Hierarchy) which became a lead to the current research on Wireless Sensor Networks. It randomly rotates the cluster head role so that the network’s energy gets distributed effectively and the network gets into a balance. The technique by adopting certain measures reduced the amount of data which is to be transmitted. In addition, LEACH offered scalability, networks dynamism and resiliency Energy consumption is reduced by a factor of 8 compared to standard protocols in simulated deployments.

[029] Lindsey et.al. [18] introduced PEGASIS, Power-Efficient GAthering in Sensor Information Systems, which is a chain-based technique; in comparison to LEACH, it is almost optimal. Because each node in PEGASIS only communicates with its near neighbours and takes turns delivering data to the base station, the total amount of power that is used during each cycle is greatly reduced. According to simulation data, PEGASIS outperforms LEACH by a factor of 100-300% across a variety of network sizes and topologies when 1, 20, 50, and 100% of nodes die.

[030] Although mentioned few contributions, it is understood that a significant and vast research was done in wireless sensor networks till now and yet there is a lot of scope for improvement. The above research contributions include the coverage over the past two decades. Numerous techniques have been deployed in wireless sensor networks to improve the network lifetime, clustering process, cluster head selection, energy efficiency and so on. Among the techniques used in the contributions, few are general in nature, and some other algorithms are nature inspired, bio inspired and genetic algorithms. The contributions however are of tremendous performance. With the advent of Machine learning, the research accelerated notably in all the areas where Wireless Sensor Networks are not an exception. As the focus of research seems to be less usage of Machine Learning techniques in WSNs, our research is inclined towards using the Machine Learning in this field.

SUMMARY OF THE PRESENT INVENTION

[031] In view of the foregoing disadvantages inherent in the known types of conventional systems, methods and techniques, are now present in the prior art, the present invention provides a machine learning based system to optimized cluster head selection in wireless sensor networks and method thereof, which has all the advantages of the prior art and none of the disadvantages.

[032] It is an object of the present invention, the system comprises an Energy focussed protocol designed in which initially the nodes are divided into clusters and for each cluster, a cluster head shall get selected. Thereafter data has been transmitted between various nodes to base station and the energy consumption is computed. In this paper, Clustering is done using various machine learning algorithms and the energy consumption results are compared. For this a network model has to be planned.

[033] In this respect, before explaining at least one object of the invention in detail, it is to be understood that the invention is not limited in its application to the details of set of rules and to the arrangements of the various models set forth in the following description or illustrated in the drawings. The invention is capable of other objects and of being practiced and carried out in various ways, according to the need of that industry. Also, it is to be understood that the phraseology and terminology employed herein are for the purpose of description and should not be regarded as limiting.

[034] These together with other objects of the invention, along with the various features of novelty which characterize the invention, are pointed out with particularity in the disclosure. For a better understanding of the invention, its operating advantages and the specific objects attained by its uses, reference should be made to the accompanying drawings and descriptive matter in which there are illustrated preferred embodiments of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

[035] When considering the following thorough explanation of the present invention, it will be easier to understand it and other objects than those mentioned above will become evident. Such description refers to the illustrations in the annex, wherein:

[036] FIG. 1-, illustrate various representations for a machine learning based system to optimized cluster head selection in wireless sensor networks and method thereof in accordance with an embodiment of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

[037] The following sections of this article will provide various embodiments of the current invention with references to the accompanying drawings, whereby the reference numbers utilised in the picture correspond to like elements throughout the description. However, this invention is not limited to the embodiment described here and may be embodied in several other ways. Instead, the embodiment is included to ensure that this disclosure is extensive and complete and that individuals of ordinary skill in the art are properly informed of the extent of the invention. Numerical values and ranges are given for many parts of the implementations discussed in the following thorough discussion. These numbers and ranges are merely to be used as examples and are not meant to restrict the claims' applicability. A variety of materials are also recognised as fitting for certain aspects of the implementations. These materials should only be used as examples and are not meant to restrict the application of the innovation.

[038] Referring now to the drawings, these are illustrated in FIG. **1**, the present invention discloses a machine learning based system to optimized cluster head selection in wireless sensor networks and method thereof.

Network Model:

In this following subsection the details of Network model for the proposed work are presented and the model is based on few assumptions mentioned below.

- The sensor nodes of wireless network are randomly spread in a square area.
- There is a stationary sink (Base Station) to which the sensor nodes communicate.
- Sensor nodes are equipped with same energy, able to sense and they could aggregate the data
- Sensors always have data to transfer, and they are firstly clustered according to the k-means algorithm.
- Links between the nodes are symmetric
- Energy consumption depends on the capability of the node as well the messages
- Nodes are not aware of the location and their capabilities are similar in terms of processing, communication
- Nodes once deployed are left unattended due to which battery recharge etc… are not possible.
- Energy consumption at a node takes place both to transmit as well receive data.

- Battery depletion which happens due to the data transfer and other activities finally leads to the death of sensor node.
- Here we are concerned only within a local network comprising base station and the sensor nodes. The communication between base station to base station is not addressed in this paper.

[039] Once the network is deployed the data communication persists between respective nodes and base station. The data communication takes the two step procedure and the energy consumption also is figured out.

1. Let the source node be ‘A’ and the energy required to transmit a ‘b’ bit block of message to Base station ‘BS’ at a distance d_{ABS} be E_{ABS}. E_{tx} & E_{rx} are the energy consumptions for transmission and reception of one bit for a unit distance.

$$E_{ABS} = b * d_{ABS} * E_{tx} + b * E_{rx} \quad \text{---------} \quad (1)$$

2. The above communication may happen directly from node to base station. The advantage in this type of transfer is that there will not be any delay in the communication while the constraint in this mode of transfer is the energy requirements of a node. Due to multiple transmissions and other such constraints most of the times the energy requirements may not be met. Alternatively, if the node could send the information in relay mode to the base station, the energy consumption at the intermediate nodes will be less at each node besides the propagation delay becomes more due to multiple hops.

   The energy consumption in the above scenario is as follows:

$$E_{ABS} = \sum_{i=0}^{n-2}(b * d_{N_i N_{i+1}} * E_{tx}) + \sum_{i=1}^{n-1}(b * E_{rx}) \quad \text{-------} \quad (2)$$

[040] Where the energy for transmission of data is the sum of the intermediate transmissions and the energy for reception is the sum of intermediate receptions of data.

[041] The above scenarios have their own advantages and limitations and are as shown in the following figures 1 (a) & (b).

[042] If the network is small then the scheme in 1(b) may result in good performance. But in reality the networks are large and in such cases, the third scenario is grouping the nodes into clusters and for each cluster a node could be selected as cluster head. The cluster heads communicate with the nodes as well base station. If necessary a cluster head may communicate with other cluster heads also. The importance of clustering is narrated below.

[043] Clustering is distributed in nature which gets terminated in limited number of iterations. Once clustering is performed the nodes are either designated to be a cluster head node or regular node. A regular node belongs to either of the clusters and basing on the local information it has the decisions may be taken independently. Clustering is said to be efficient when the number of clusters formed are reasonable and the performance of clustering is effective in terms of energy consumption. Moreover if the number of clusters formed are limited, the respective cluster heads may be given additional capabilities like enhancing the battery power, providing alternate sources of energy and so on.

[044] *<u>Step wise procedure of the work flow:</u>*

In the proposed work, a network is simulated comprising 100 nodes and a sink node where the nodes are randomly spread over the area. Here the 100 nodes are having equal capability. The energy possessed by each node is 1 J. Then the 100 nodes are divided into clusters using three different clustering algorithms. There are various clustering algorithms that include general clustering algorithms, genetic algorithms and machine learning algorithms. In this paper machine learning based clustering algorithms viz. k-means (KMC), Agglomerative hierarchical (HC) and Gaussian Mixture Model (GMMC) are opted and their performance is compared. After performing clustering in each method four clusters are obtained. Also there are various cluster formation optimization methods are available to test whether the number of clusters formed is reasonable or not. Silhouette method is one such algorithm used to test the cluster formation.

[045] After cluster formation, the energy efficiency is computed by transmitting messages from various nodes to base station. The performance with respect to the message transmission is compared for the three algorithms. The overall procedure is depicted in the below flow chart figured in 2.

PHASE I operation:

Network creation, cluster formation using the three clustering algorithms:

[046] The Network is simulated with 100 nodes randomly spread with initial energy of each node 5 J. Subsequently Clustering is performed using the three clustering algorithms namely k-means, agglomerative hierarchical and Gaussian Mixture Modeling for which the plotting of the nodes in the network could be seen in figures 3, 4 & 5 respectively.

[047] Once the clusters are formed the clustering is validated using Silhouette method where the number of clusters and Silhouette score seems to be similar for all the three algorithms. In the above process, for each clustering scheme after the random nodes generation, clustering is performed. Then each node shall be assigned a cluster label and also the Cluster Heads are obtained. Cluster Head (CH) is designated basing on the centroid distance to the CH node from all the nodes in the respective cluster.

PHASE II:

Message transfer and Energy Efficiency calculation:

[048] In this experimentation, ten messages are considered to be transmitted between the source node to sink. So as to transfer first the path of message transmission is to be identified. For this the dijkstra's algorithm is chosen. The algorithm results in standard and shortest path. Once the path is identified, the messages are transmitted. Then the energy involved in this message transmission are computed and finally the algorithm with high energy efficiency has been identified. The pseudocode representing the above procedure is depicted below.

(i) Generating random nodes:

for each node N_i

x,y ← get_random (x- coordinate, y- coordinate)

N_i ← (x,y)

(ii) Cluster formation (alg):

fit_nodes(alg)

for each node N_i

N_{jk} ← $N_i(n_jC_k)$ //Assigning cluster label to each node where N_{jk} is j^{th} node in k^{th} cluster

$CH_k \equiv N_{jk}$ (where $d_{N_{jk}} = \frac{\sum_{t=1}^{p} d_t}{p}$) //Assigning Cluster Head – Node having near to Mean distance from all the nodes)

(iii) Finding path – build_path(N_{jk},BS):

for each node N_{jk}:

$dist_{N_{jk}}$ ← euclidean(N_{jk},CH_k) //Distance between node and the cluster head

node_dist_list.append($dist_{N_{jk}}$)

for each cluster head CH_k:

$dist_{CH_k}$ ← euclidean(CH_k , BS) //Distance between cluster head and Base station

CH_dist_list.append($dist_{CH_k}$)

G ← generate_graph(node_dist_list, CH_dist_list) //Graph generation for nodes, cluster head and base station

path ← Dijkstra(G,N_{jk})

(iv) Computing Energy consumption for transmitting message M_i:

E_{M_i} ← 0

Source ← N_{jk}

Destn ← BS

P ← build_path(Source, Destn)

b ← num_bits(M_i)

for each distance d_t in P:

$E_{tx} = b * d_t * E_T$

$E_{rx} = b * E_R$

$E_{M_i} = E_{M_i} + E_{tx} + E_{rx}$

(v) Computing Energy consumption for transmitting all the messages:

for each message M_i:

compute E_{M_i}

$$E_{Total} = \sum_{i=1}^{n} E_{M_i}$$

(vi) Computing Energy consumption for transmitting all the messages for each algorithm:

$E_{Total_{KMC}}$ – Total energy consumed for transmitting all the messages by adopting K Means Clustering

$E_{Total_{AHC}}$ – Total energy consumed for transmitting all the messages by adopting Agglomerative Hierarchical Clustering

$E_{Total_{GMMC}}$ – Total energy consumed for transmitting all the messages by adopting Gaussian Mixture Modeling Clustering

(vii) Energy Efficiency (Average Energy consumption) for transmitting messages for each algorithm:

$$E_{Avg_{kmc}} = \frac{1}{n} * E_{Total_{KMC}}$$

$$E_{Avg_{ahc}} = \frac{1}{n} * E_{Total_{AHC}}$$

$$E_{Avg_{gmmc}} = \frac{1}{n} * E_{Total_{GMMC}}$$

Results and Discussion:

[049] The model is built using Python Programming, in which the base station BS is located and fixed at the centre throughout the simulation, while the sensor nodes are randomly spread throughout the defined space. Moreover, each sensor node has the information related to couple of local nodes.

[050] During the phase I operation, Clustering is performed and necessarily it must be validated. There are numerous techniques to validate the clustering i.e., in terms of number of clusters formed. Amongst Silhouette score is one measure of validation. The silhouette scores for the respective clustering methods are as mentioned below in table 1. As the silhouette scores and the number of clusters into which the network nodes are divided are equivalent, it could be said that the clustering performed seems to be valid.

The graph representing the above average energy consumption values are depicted in figure 7. On the x- axis the respective clustering algorithm is shown and the average energy consumption value is marked on the y-axis. Here as the average energy consumption is less for Gaussian Mixture Modeling Clustering algorithm, its energy efficiency is said to be high.

Conclusion:

[051] Wireless Sensor Networks are inevitable in the current scenario. Almost in every domain the importance of WSNs are ever increasing. Such technology must be elegantly designed for the intended applications. In designing the appropriate sensor network, the vital element is node – base station communication where in clustering plays a crucial role. The design of clustering algorithms must be very effective such that the energy efficiency is of top priority. In this paper Machine Learning based Clustering techniques have been part of research subject. Three clustering algorithms K Means, Hierarchical and Gaussian Mixture are used for clustering and the network is divided into clusters. Multiple data transfers are done and the energy consumption for the data transfer is calculated. Finally the average energy consumption is calculated for which Gaussian Mixture modelling technique resulted in low energy consumption comparatively than the other two techniques. It could be concluded that GMM clustering technique better performed in terms of energy efficiency.

[052] It is to be understood that the above description is intended to be illustrative, and not restrictive. For example, the above-discussed embodiments may be used in combination with each other. Many other embodiments will be apparent to those of skill in the art upon reviewing the above description.

[053] The benefits and advantages that the present invention may offer have been discussed above with reference to particular embodiments. These benefits and advantages are not to be interpreted as critical, necessary, or essential features of any or all of the embodiments, nor are they to be read as any elements or constraints that might contribute to their occurring or becoming more evident.

[054] Although specific embodiments have been used to describe the current invention, it should be recognized that these embodiments are merely illustrative and that the invention is not limited to them. The aforementioned embodiments are open to numerous alterations, additions, and improvements. These adaptations, changes, additions, and enhancements are considered to be within the purview of the invention.

We Claim:

1. A computer-implemented method for simulating a network comprising 100 nodes and a sink node, wherein the nodes are randomly spread over an area, and the 100 nodes have equal capability, the method comprising:
    a. Dividing the 100 nodes into clusters using a machine learning-based clustering algorithm selected from the group consisting of k-means (KMC), Agglomerative hierarchical (HC), and Gaussian Mixture Model (GMMC);
    b. Applying a cluster formation optimization method to evaluate the reasonableness of the number of clusters formed, wherein the cluster formation optimization method comprises employing the Silhouette method;
    c. Obtaining four clusters after performing clustering using the selected machine learning-based clustering algorithm;
    d. Computing energy efficiency by transmitting messages from various nodes to a base station; and e. Comparing the performance of message transmission among the three machine learning-based clustering algorithms.
2. A system for simulating a network comprising 100 nodes and a sink node, the system comprising:
    a. A processor configured to execute a computer program for dividing the 100 nodes into clusters using a machine learning-based clustering algorithm selected from the group consisting of k-means (KMC), Agglomerative hierarchical (HC), and Gaussian Mixture Model (GMMC);

b. The processor further configured to apply a cluster formation optimization method to evaluate the reasonableness of the number of clusters formed, wherein the cluster formation optimization method comprises employing the Silhouette method;

c. The processor further configured to obtain four clusters after performing clustering using the selected machine learning-based clustering algorithm;

d. The processor further configured to compute energy efficiency by transmitting messages from various nodes to a base station; and

e. The processor further configured to compare the performance of message transmission among the three machine learning-based clustering algorithms.

Dated this 26th day of June 2023

Applicant

Andhra University

ABSTRACT

A MACHINE LEARNING BASED SYSTEM TO OPTIMIZED CLUSTER HEAD SELECTION IN WIRELESS SENSOR NETWORKS AND METHOD THEREOF

[055] The present invention discloses a machine learning based system to optimized cluster head selection in wireless sensor networks and method thereof. In the present invention, the system comprises an Energy focussed protocol designed in which initially the nodes are divided into clusters and for each cluster, a cluster head shall get selected. Thereafter data has been transmitted between various nodes to base station and the energy consumption is computed. In this paper, Clustering is done using various machine learning algorithms and the energy consumption results are compared.

Accompanied Drawing **[FIGS. 1-7]**

Dated this 26th day of June 2023

Applicant

Andhra University

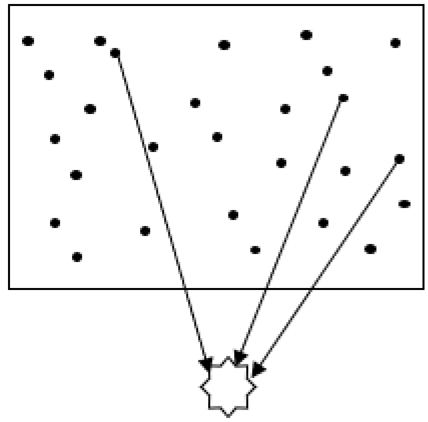
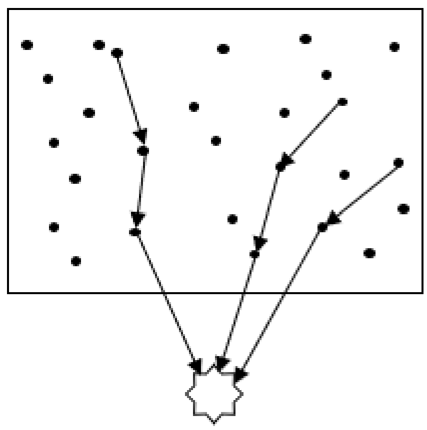


Fig. 1(a): Direct Transmission from source node to base station



1(b): Relay mode Transmission from source node to base station

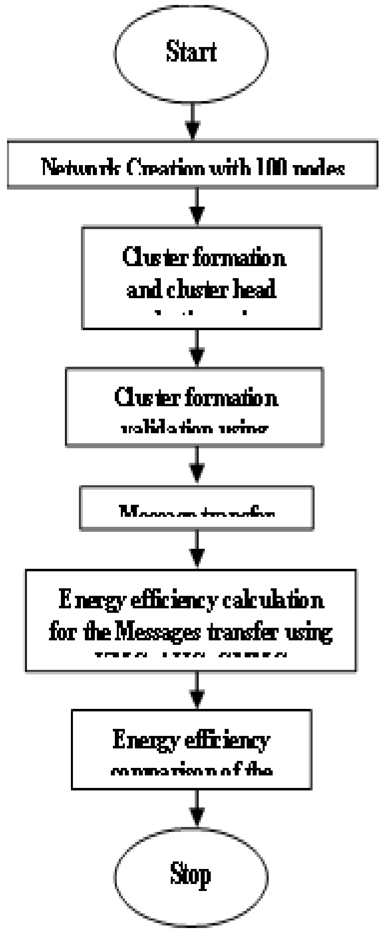


Fig 2: Stepwise procedure of the proposed research work.

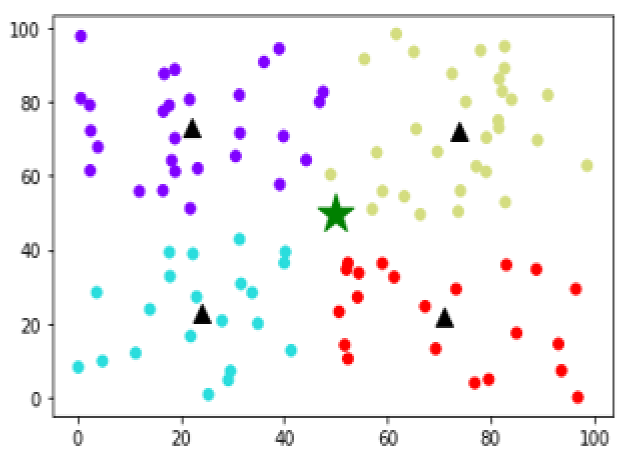


Fig. 3. Cluster formation using K-Means

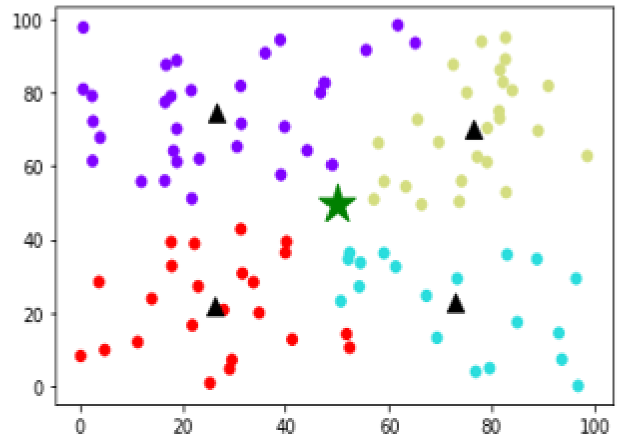


Fig. 4. Cluster formation using Hierarchical

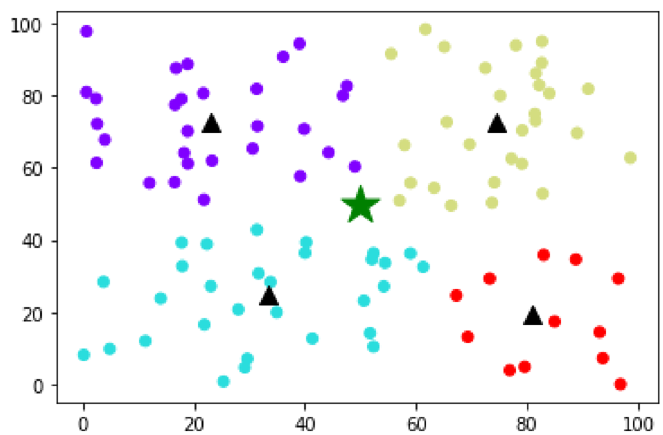


Fig. 5. Cluster formation using Gaussian Mixture

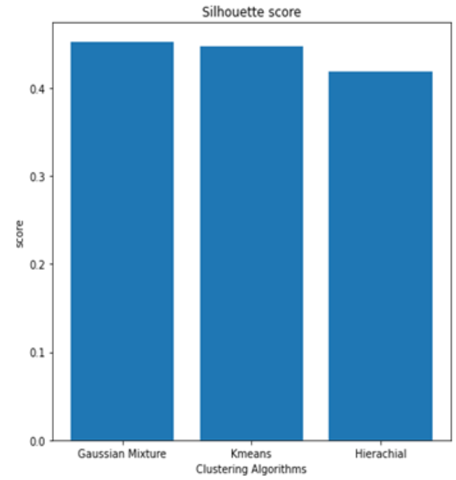


Fig 6: Silhouette scores

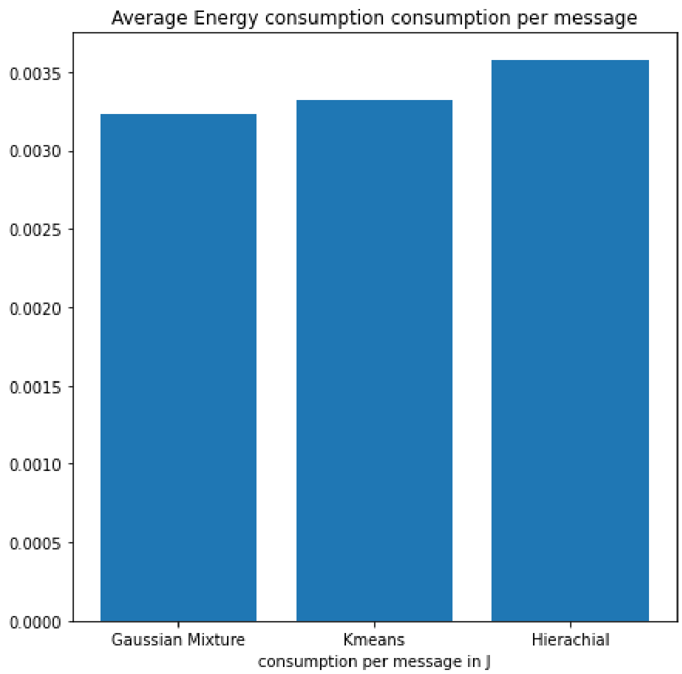


Fig 7: Energy efficiency comparison of the three algorithms

Scheme	Silhouette score	No. of Clusters formed
Hierarchical Clustering	4.085584741574557	4
K-means Clustering	4.415059119028434	4
Gaussian Mixture Clustering	4.4440583494004243	4

Table 1: Silhouette scores of various clustering schemes

Parameter	Values considered
Network size	100 X 100
Number of Nodes	100
Initial Energy of nodes	5 J
No. of messages	10
Average no. of bits per message	200
Number of Cluster Heads	4
Base station position	(50,50)
Energy consumption for receiving and transmitting a bit	96 µJ / bit

Table 2. Parameters for the simulation

Clustering Algorithm used	Average Energy consumption
K Means Clustering	0.0033188589500907015 J
Hierarchical Clustering	0.0035731528935163846 J
Gaussian Mixture Modeling Clustering	0.0032297323159530184 J

Table 3. Average Energy consumption for the three algorithms

Dated this 26th day of June 2023