INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2023, VOL. 69, NO. 4, PP. 793–801
Manuscript received May 30, 2023; revised October, 2023. DOI: 10.24425/ijet.2023.147703

K-Means and Fuzzy based Hybrid Clustering Algorithm for WSN

Basavaraj M. Angadi and Mahabaleshwar S. Kakkasageri

***Abstract*—Wireless Sensor Networks (WSN) acquired a lot of attention due to their widespread use in monitoring hostile environments, critical surveillance and security applications. In these applications, usage of wireless terminals also has grown significantly. Grouping of Sensor Nodes (SN) is called clustering and these sensor nodes are burdened by the exchange of messages caused due to successive and recurring re-clustering, which results in power loss. Since most of the SNs are fitted with non-rechargeable batteries, currently researchers have been concentrating their efforts on enhancing the longevity of these nodes. For battery constrained WSN concerns, the clustering mechanism has emerged as a desirable subject since it is predominantly good at conserving the resources especially energy for network activities. This proposed work addresses the problem of load balancing and Cluster Head (CH) selection in cluster with minimum energy expenditure. So here, we propose hybrid method in which cluster formation is done using unsupervised machine learning based k-means algorithm and Fuzzy-logic approach for CH selection.**

***Keywords*—Wireless Sensor Networks; Cluster; K-Means algorithm; Fuzzy Logic**

I. INTRODUCTION

IN most applications, Wireless Sensor Networks (WSN) are utilized to simplify and manage complicated problems. In WSN, energy conservation is a top priority. It is significant since the network's lifetime is mostly determined by the WSN's energy usage [1]. As a result, balancing and conserving energy use is given top priority. So it is necessary to design algorithms that consume least amount of energy [2] which can be achieved with the usage of evolving computational techniques [3] - [5].

The fundamental purpose of clustering schemes for WSN is to organize sensors into clusters and select a Cluster Head (CH) for each cluster. As we all know, direct communication between the SNs and base station especially in large scale network consumes more energy, causing the WSN to expire sooner. In WSN, data from cluster members is aggregated by CH and delivered to the BS/gateway/sink for efficient energy utilization as shown in the Fig. 1. In a two-layered clustering protocol, the first layer is responsible for determining the best CH set, while the second layer is responsible for transferring data to the BS [6].

The difficulty of the clustering technique increases linearly with the size of the network. The selection of the cluster head is another significant difficulty that has a direct impact on network performance. In large-scale WSNs, using more than one sink nodes can improve the network's scalability and lifespan [7]. Clustering mechanisms should not only make data transfer easier in WSNs, but also take into account the sensor nodes' restrictions to meet reliable data transfer, energy efficiency and scalability. During the cluster formation process, clusters are constituted in such a way that load balance is maintained and a minimum number of messages are exchanged.

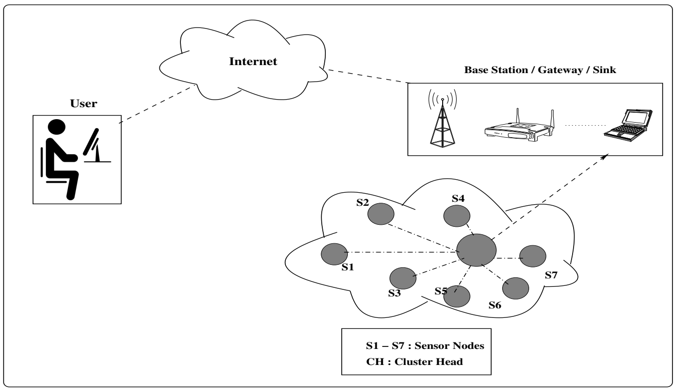

Fig. 1. Network Scenario

It is necessary to develop effective clustering protocols for controlling network topology and overcoming challenges such as CH selection, load balancing in the cluster, lowering cluster control messages. As a result, by removing redundant energy use, energy consumption is reduced. In this work, we are going to propose hybrid clustering mechanism where unsupervised machine learning based K-Means algorithm is used for cluster formation. After cluster formation, CH is selected using fuzzy logic based Multi-Attribute Decision-Making (MADM) technique.

The organization of the remainder of the paper is as follows: Brief review of clustering algorithms based on computational intelligence techniques such as machine learning and fuzzy logic is presented in section 2. Section 3 presents proposed clustering model. Simulation parameters and result analysis are briefed in 4 and 5 sections respectively. Finally section 6 concludes the work.

Basavaraj M. Angadi and Mahabaleshwar S. Kakkasageri are with Faculty of Electronics and Communication Engineering Department, Basaveshwar Engineering College, Bagalkote, Karnataka, INDIA (e-mail: bmaec@becbgk.edu, mskec@becbgk.edu).

II. Related Works

In unsupervised learning the developed model extracts the relationship and there is no output associated to the inputs. Similar patterns are grouped together in this learning method. Routing, data aggregation, connectivity challenges, anomaly detection and clustering are the some of the issues that unsupervised learning [8] solves with WSNs. Fuzzy-C-Means (FCM), K-Means (KM), Hierarchical-based, and K-Medoids algorithms are examples of unsupervised learning based clustering algorithms [9]. The performance of KM and FCM are investigated in [10] to comprehend which one has a superior capacity to build balanced clusters, so that researchers can choose the best strategy for increasing lifespan of the network.

The K-Means algorithm is a well-known clustering algorithm that looks for approximate solutions. To group Sensor Nodes (SN) into cluster, the k-means employs the Euclidean distance as a similarity metric and has received a lot of attention with frequent usage[11] [12].

Traditional K-Means has various drawbacks which appear in real-time applications data sets, the most significant of which is that it exclusively considers the distance criteria for grouping. The paper in [13] presents a unique modified K-Means which it makes use of the power of bargaining game modeling. In this technique cluster centroids operate as bargaining players in which each cluster centroid position varies depending on the behavior of the other cluster centroid and the location of the data objects. To tackle the issues of processing huge datasets sequentially and scalability for processing datasets of any size in an elegant manner, linear clustering techniques are often more efficient than nonlinear clustering algorithms. So to solve the nonlinearly separable clustering problem, a multiple K-Means clustering algorithm [14] is used in which an aggregate clusterer of many clusterers is formed using k-means as a foundation clusterer.

Lifetime of Wireless Sensor Network can be extended using a combination of K-Means algorithm and an Optimal Path Selection Method [15]. Firstly, centre of each cluster is chosen as a data-gathering location using the K-Means method. A spanning tree algorithm is used to determine the optimal/best path to avoid direct transmission between CHs and BS which reduces data transmission time between CHs. The K-Means clustering method is assessed using three different mathematical metrics in terms of execution time by considering dissimilar datasets and clusters [16].

Load balancing was achieved by picking the optimal settings for fuzzy-based clustering in order to determine an acceptable cluster head for data gathering and routing [17] [18]. An energy-efficient clustering for hierarchical routing protocols based on fuzzy logic is proposed in [19]. The Fuzzy Inference System (FIS) is a useful tool for integrating influencing elements (parameters) to improve CH selection [20] based on remnant energy, average communication distance, and communication quality. Several fuzzy logic based energy-aware uneven clustering methods were presented in [21]-[23] especially to handle the hot spot problem and manage differences in the estimation of CH radius.

III. Proposed Work

Clustering is a key technique used in WSNs to reduce energy consumption, increase network scalability, and prolong network lifetime. K-Means and fuzzy-based clustering algorithms are popular approaches used to cluster sensor nodes in WSNs. To overcome the limitations of these individual clustering algorithms, a hybrid clustering algorithm based on K-Means and fuzzy logic is proposed to improve the accuracy and efficiency of clustering in WSNs. Network environment, computational and proposed models are discussed in the following subsections.

A. Computational Models

1) Network Model: The goal of this work is to present a hybrid clustering strategy for WSN comprising of stationary Sensor Nodes (SN). A two-dimensional network with N sensors divided into K-clusters and a sink node/Base Station (BS) are considered with an assumption that the BS has a fixed physical location and unlimited resource supply. Each node has a unique node identifier that corresponds to a certain communication range. Every cluster has Cluster Head (CH) selected based on fuzzy based Multi-Attribute Decision Making (MADM) technique. The CH in each cluster receives the sensed data from all cluster member SNs. The data aggregated by CHs is transmitted to user via BS through internet as shown in the Fig. 2.

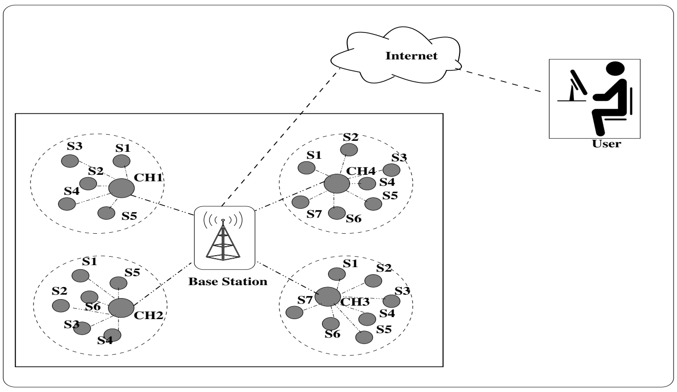

Fig. 2. Network Environment

2) Energy Model: In a SN, energy is used primarily during transmission, reception and aggregation of the data. Usually energy consumption for aggregation is fixed for certain period of time, but the Transmission Energy (TX_{ERNG}) and Receiving Energy (RX_{ERNG}) vary as per the free space path and multipath models that are dependent on distance as shown in equation (1) (2) and (3).

$$TX_{ERNG} = l_p \times E_{elec} + l_p \times e_{fs} \times d^2 \quad (1)$$

For $d < d_0$

$$TX_{ERNG} = l_p \times E_{elec} + l_p \times e_{mp} \times d^4 \quad (2)$$

For $Others$

Where l_p is the length of data (packets), e_{fs} and e_{mp} are the factors used by free space path and multipath models. $E_{elec}$is the energy required to convert 1 bit of data to a signal and d is the distance between transmitter and receiver, $d_0 = \sqrt{\frac{e_{fs}}{e_{mp}}}$ is the reference distance.

$$RX_{ERNG} = l_p \times E_{elec} \tag{3}$$

B. Proposed Model

In this proposed work, the deployed sensor nodes are divided into 'K' clusters using K-Means algorithm. Once clusters are formed, the CH is selected using Fuzzy Inference System (FIS) based on MADM technique as shown in Fig. 3. In the FIS, parameters such as residual energy (R_Energy), Nearest Neighbors (NN) and distance from Base Station (d_BS) are considered for selection of the CH.

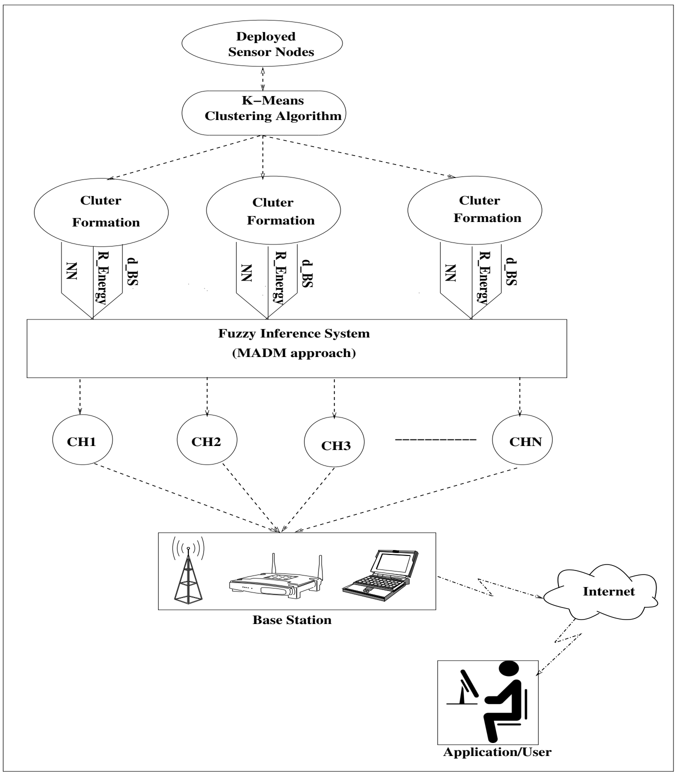

Fig. 3. Proposed Clustering Model

1) Cluster Formation: In the K-Means algorithm, initially 'K' numbers of points are selected as centroids randomly. The sensor nodes closest to the centroid are assigned to it for formation of 'K' number of clusters. After assignment of all the SNs to its nearest centroid, once again the centroids are recomputed and sensor nodes are re-assigned to the new closest centroid of each cluster. If any re-assignment is required, recomputation of the centroid continues until there is no change in the location of the newly selected centroid as shown in Algorithm 1.

Let us consider set of sensor nodes denoted by S_{Node} as in equation (4)

$$S_{Node} = [S_1, S_2, S_3, ...S_n] \tag{4}$$

Algorithm 1 Cluster Formation

1: *Nomenclature: Sensor nodes - S_{Node}, Centroids - C_{ntroid} $d(x_i, x_c)^2$ - Euclidean distance between centroid C_j and node x_{ij}*
2: *Input : $S_{Node} = [S_1, S_2, S_3, ...S_n]$ number of clusters - k*
3: *Output : Clusters $C = [C_1, C_2, C_3, ...C_k]$.*
4: **Begin**
5: *Initialize k centroids*
6: ***for** each iteration **do***
7: *Compute $d(x_i, x_c)^2$*
8: *Assign S_{Node} to closer C_{ntroid}*
9: ***for** S_{Node} = 1 $\rightarrow n$ **do***
10: $C_{ntroid} = [C_1, C_2, C_3, ...C_k]$
11: ***end for***
12: ***end for***
13: ***for** each cluster k **do***
14: *Update $C_{ntroid}(X, Y)$ using equation (8)*
15: ***end for***
16: ***if** change in $C_{ntroid}(X, Y)$ **then** Go to step: 6*
17: ***else***
18: *end clustering*
19: ***end if***
20: **End**

Where 'n' is the number of sensor nodes. Initially the set of 'K' number of clusters with randomly selected centroids C_{ntroid} is given by equation (5) as

$$C_{ntroid} = [C_1, C_2, C_3, ...C_k] \tag{5}$$

The objective function of the K-Means algorithm is given by equation (6) as

$$F_{obj} = \sum_{i=1}^{n} \sum_{j=1}^{k} d(x_i, x_c)^2 \tag{6}$$

where $i = 1, 2, 3, ...n$ sensor nodes and $j = 1, 2, 3, ...k$ clusters with centroids.

In equation (6), $d(x_i, x_c)^2$ is the Euclidean distance between centroid C_j and node x_{ij}. Here i refers to the sensor node and j is the cluster. After assignment of all the sensor nodes to its closest centroid, there will be re-computation of the centroid for newly formed clusters and is the average location point of all the sensor nodes in the cluster given by equation (7) as:

$$C_{ntroid}(X, Y) = \left(\frac{1}{S_{Node}} \sum_{i=1}^{S_{Node}} x_i, \frac{1}{S_{Node}} \sum_{i=1}^{S_{Node}} y_i \right) \tag{7}$$

If there no change in the location of centroids in the newly formed clusters, then clustering process is stopped and clusters are formed. Since centroid for the cluster is nominated considering only the distance, it will not play the CH role effectively. So it is necessary to elect CH for the cluster which can effectively balance the load in the clusters.

2) Cluster Head Selection: In a wide range of scientific and engineering domains, Multi-Criteria Decision-Making (MCDM) techniques have been used to resolve the quantitative decision-making difficulties. Multi-Objective Decision-Making (MODM) and Multi-Attribute Decision-Making (MADM) techniques are the two basic types of MCDM. In spite of all the parameters under consideration, MODM chooses nondominant alternatives. Where as, MADM approaches quantitatively compare and rank alternatives based on the degree of appropriateness of the attributes being taken into consideration for the study. Here in this work, we are using MADM based technique for selecting the CH.

Algorithm 2 illustrates the selection of CH using fuzzy logic. Fuzzy-based CH selection is accomplished using three parameters, includes Residual Energy (R_Energy), Nearest Neighbors (NN) and distance from Base Station (d_BS) to determine the reward for cluster head selection. Fuzzifier accepts three input parameters and carryout the Fuzzification with the involvement of the knowledge rule that uses membership function and fuzzy-based rule construction. Defuzzifier uses the predefined rules for computation of the output as shown in Fig 4.

Algorithm 2 Cluster Head Selection Process

Nomenclature:Residual Energy - R_Energy, Nearest Neighbors - NN, distance from Base Station - d_BS, Cluster Head - CH, number of clusters-k
Input : R_Energy, NN, d_BS
Output : Efficient CH
Begin
Read fuzzy inputs
for** each k **do
    Fuzzify (Membership Function, Fuzzy Based Rules)
end for
for** each k **do
    Defuzzify (Predefined Rules)
end for
Efficient CH selected
End

The membership function for the considered input sets are shown in table I. The linguistic variables low and high are considered for residual energy. For nearest neighbors the linguistic variables are medium and dense. Similarly short and far are considered for distance from base station. In all three cases the triangular membership functions are taken for consideration.

For the selection of CH, five output member functions as shown in Fig. 5 are considered. Table II. illustrates the nine rules considered for the selection of best CH. Based on the values assigned to the linguistic variables of the logic, membership function and predefined fuzzy rules, best cluster head is elected.

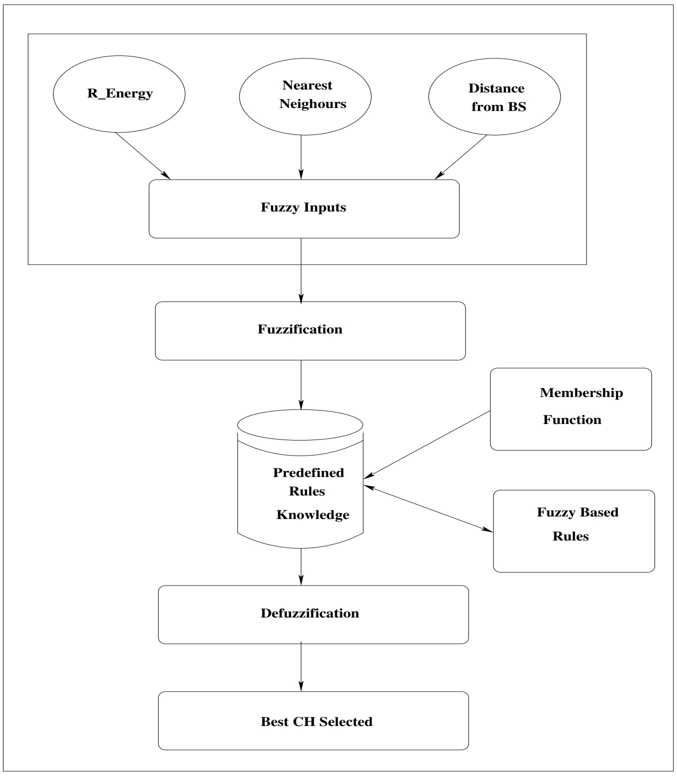

Fig. 4. Fuzzy Inference System for Custer Head Selection

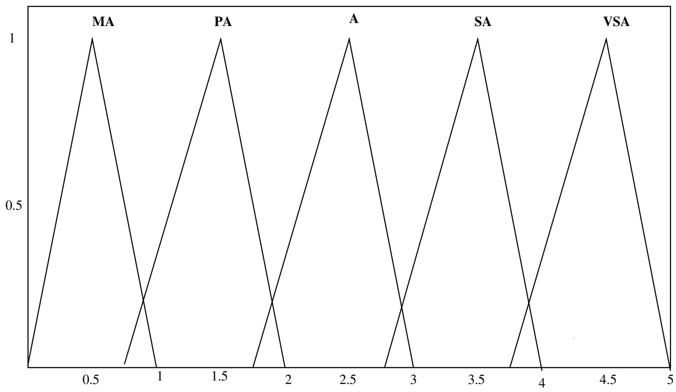

Fig. 5. Membership function plot for CH selection

IV. Simulation

We have used “C++” programming language as discrete event simulator to simulate the proposed scheme. Initially appropriate network environment is created with random deployment of the sensor nodes. The ’N’ numbers of deployed Sensor Nodes (SN) in the environment region of $100m$x$100m$ are grouped into ’k’ clusters using K-Means clustering algorithm. For the deployed SNs energy is allocated randomly. In this section, simulation inputs and performance parameters are presented.

A. Simulation Inputs

The parameters considered for simulation are illustrated in table III.

B. Performance Parameters

The following performance metrics are taken into account for evaluating the efficiency of the proposed work.

TABLE I
LINGUISTIC INPUT VARIABLES MEMBERSHIP FUNCTION

Sl. No.	Membership Functions		
	Residual Energy (R_Energy)	**Nearest Neighbors (NN)**	**distance from BS (d_BS)**
1	Low (0)	Dense (0)	Far (0)
2	High (1)	Medium (1)	Short (1)

TABLE II
FUZZY RULES FOR CH SELECTION

Residual Energy	Nearest Neighbors	distance from BS	Confidence Factor	Weights
Low	Medium	Short	Partially Agree (PA)	2
Low	Medium	Far	Marginally Agree (MA)	1
Low	Dense	Short	Partially Agree (PA)	2
Low	Dense	Far	Marginally Agree (MA)	1
High	Medium	Short	Very Strongly Agree (VSA)	5
High	Medium	Far	Agree (A)	3
High	Dense	Short	Strongly Agree (SA)	4
High	Dense	Far	Agree (A)	3

TABLE III
SIMULATION PARAMETERS

Sl. No.	Parameters	Specifications
1.	Simulation area in meters [A]	100*100
2.	Number of nodes [N]	50-300
3.	Transmission Range	100m
4.	Channel Type	Wireless
5.	Initial energy of a node	1.5J
6.	Energy of transmitting each bit	50nJ/b
7.	Energy of receiving each bit	50nJ/b
8.	Length of the packet	128KB
9.	Node distribution	Random

- **Computational Delay:** It is defined as the amount of time required to form the cluster and select the cluster head in each cluster. It is expressed in milli second (ms).
- **Cluster Head Identification Delay:** Is defined as the amount of time required to elect the cluster head in each cluster and is expressed in milli second (ms).
- **End to End Delay:** Is defined as the time required for a packet to be generated at the sensor node till it is received by the sink and is expressed in milli second (ms).
- **Total Energy Consumption:** Is the total energy consumed by all sensor nodes and cluster heads present in the network. It is measured in milli Joules(mJ).
- **Network Overheads:** It is defined as fraction of bits in every sensor data packet that are not part of actual data. In general, it is considered as the total number of bits transmitted per successfully delivered data packets and is expressed in bits (b).
- **Packet Delivery Ratio:** Is defined as the ratio of total number of packets successfully received to the total packets that are actually sent. It is measured in percentage (%).
- **Difference Factor:** Is defined as the difference between the number of maximal and minimal cluster members in each cluster.
- **Cluster Validity Index (CVI):** It is defined as the ratio of minimum distance between the cluster centroids to the maximum distance between any pair of sensor nodes within that cluster. The CVI measures clustering quality in forming clusters that are well separated and firmly packed.
- **Clustering Accuracy:** Is defined as the the degree to which a clustering algorithm accurately assigns sensor nodes to the correct clusters based on some similarity criteria or distance measure. It is the ratio of average similarity between each cluster and its most similar cluster to the similarity between the clusters being compared and expressed in percentage (%).
- **Standard Deviation of Residual Energy (SDRE):** Is a metric that quantifies the degree of variation of sensor node residual energy from the mean value. SDRE is defined as the mean variance between residual energy levels at sensor nodes and is determined using equation equation (8) as:

$$SDRE = \sqrt{\sum_{i=1}^{N} \left(\frac{(E_i - \mu)^2}{N} \right)} \tag{8}$$

Where E_i is the residual energy of i^{th} sensor node, μ is the mean residual energy and N is the number of nodes. SDRE provides the information about the even distribution of the energy in the network.

V. RESULT ANALYSIS

The simulation results of proposed work are compared with the Probabilistic Based Optimized Adaptive Clustering Scheme for Energy-Efficiency in Sensor Networks [24] referred as L-DDRI in the graphs. The k-means algorithm especially for clustering in WSN offers more advantages over K-Nearest Neighbors (KNN) and k-medoids with respect to scalability, communication cost, performance and reduced communication overhead. Performance of the k-means algorithm with different k values shown in Fig. 6 is better than k-medoids and KNN, which indicates that the SNs deployed are clearly distinguished and well placed in the cluster.

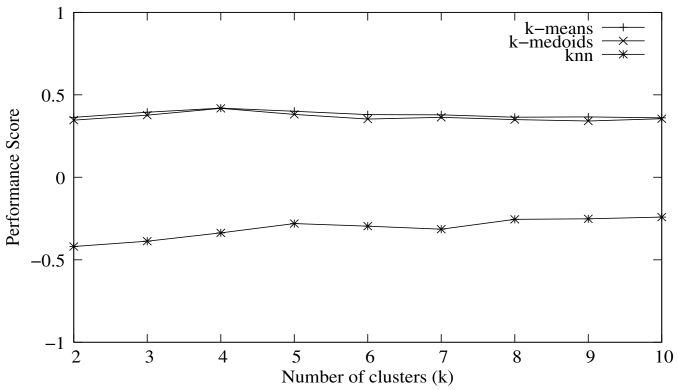

Fig. 6. Number of Clusters (k) Vs. Performance Score

The k-means algorithm uses mainly distance metric, preferably Euclidean distance for clustering Sensor Nodes (SNs). Communication cost for each clustering approach can be determined by summing the Euclidean distances between the sensor nodes and their respective cluster centers. The graph in Fig. 7 compares the communication cost of k-means, k-medoids and KNN algorithm for different k values, keeping number of SNs constant. Communication cost decreases as the number of clusters (k) increases due to decreased distance from centroid to SNs in which k-means algorithm performs better than k-medoids and KNN.

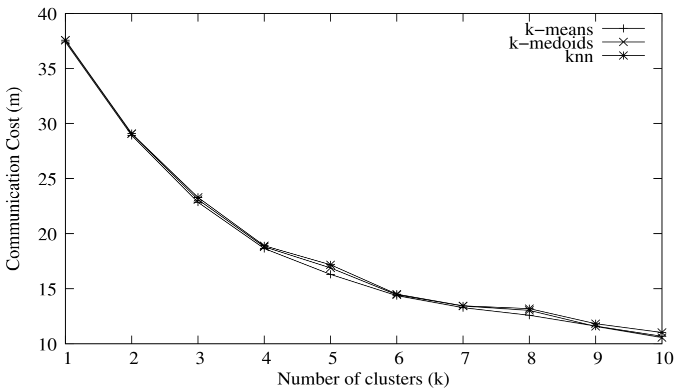

Fig. 7. Number of Clusters (k) Vs. Communication Cost

Communication cost is also observed by varying number of SNs with fixed k value. As there is increase in the number of SNs, commuinication cost also increases due to increased distance between SNs and their respective centroid. Comparision of commuinication cost with different number of nodes for k=1, k=2, k=3 and k=4 is shown in Fig. 8(a), Fig. 8(b), Fig. 8(c) and Fig. 8(d) respectively. It is observed that, k-means algorithm gives better clusters than k-medoids and KNN.

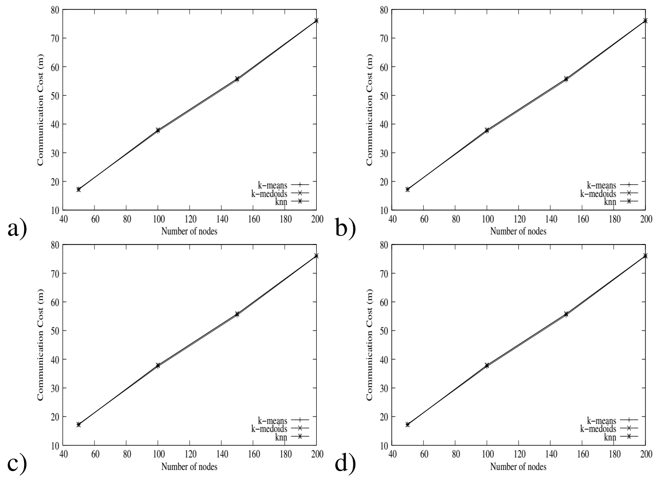

Fig. 8. Number of Nodes Vs. Communication Cost a) For k=1, b) For k=2, c) For k=3, d) For k=4

K-means algorithm is preferred for clustering because the computation complexity is less than k-medoids and KNN. The computation complexity of the algorithm will increase if more number of iterations are required for clustering. Since KNN requires labelled data for clustering, number of iterarions required by k-means and k-medoids are only compared and is depicted in Fig. 9. It is observed that, k-means algorithm uses less number of iterations than k-medoids and hence has reduced computation complexity.

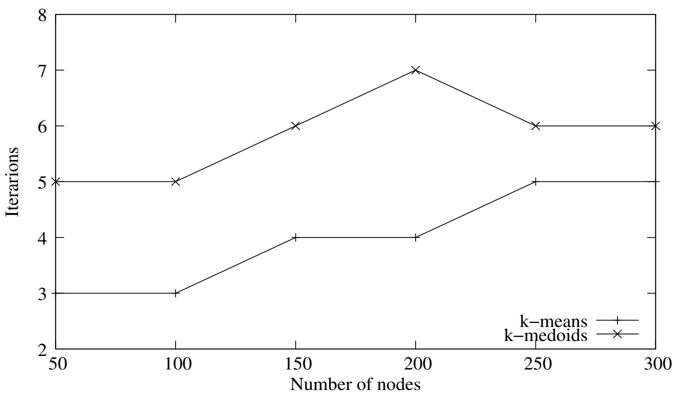

Fig. 9. Number of Nodes Vs. Iterations

The average clustering overhead for the k-means and k-medoids is compared by varying number of SNs as shown in Fig. 10. It is observed that, exchange of data for clustering in k-means is less than k-medoids. The decrease in clustering overhead increases the performance of the algorithm which influences selection of k-means for clustering.

Computational delay with varying number of nodes and sensor nodes' communication range is depicted in Fig. 11. It is observed that there is reduction in the computational delay for the proposed scheme in comparison with the existing L-DDRI because k-means require simple computations for clustering SNs.

Cluster Head (CH) identification delay for various number of nodes and communication range is shown in Fig. 12. Since

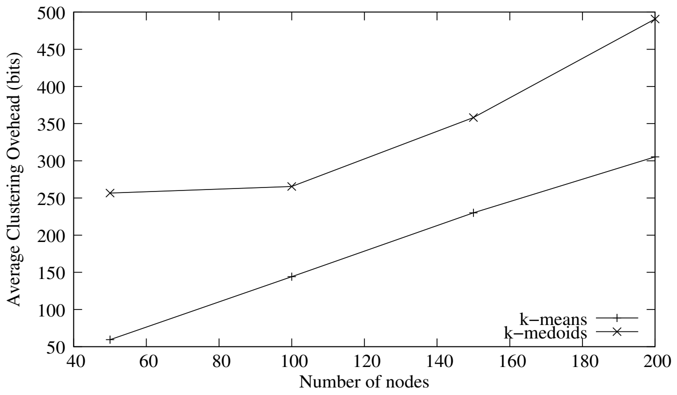

Fig. 10. Number of Nodes Vs. Average Clustering Overhead

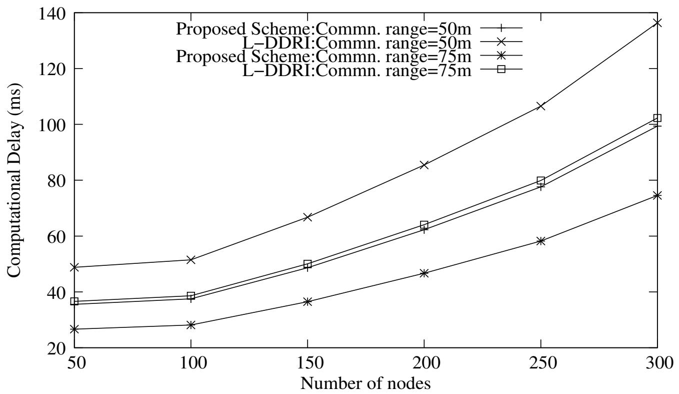

Fig. 11. Number of Nodes Vs. Computational Delay

fuzzy logic is used to elect only CH for the formed clusters, the CH identification delay is less for the proposed algorithm compared to L-DDRI.

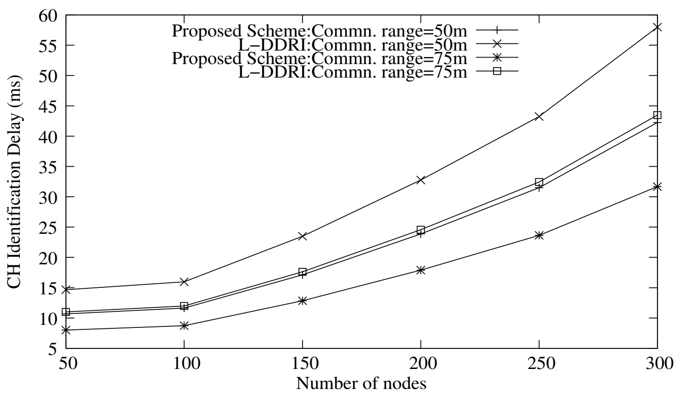

Fig. 12. Number of Nodes Vs. Cluster Head Identification Delay

End to end delay with varying number of nodes and communication range is depicted in Fig. 13. It is observed that, end to end delay for the proposed scheme is less than the L-DDRI. There is reduction in the delay as the communication range increases due to enhanced SNs transmission capability.

As network grows (increment in number of nodes), conventionally the amount of energy consumed will also increase. The total energy consumption for both proposed and L-DDRI scheme is shown in Fig. 14. It is seen that, performance of

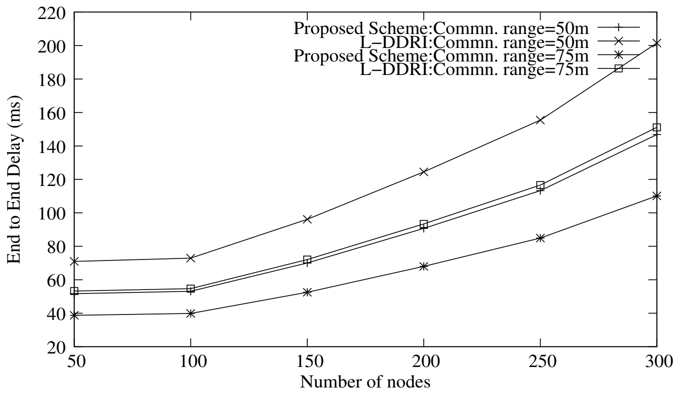

Fig. 13. Number of Nodes Vs. End to End Delay

the proposed scheme is better than the L-DDRI in terms of total energy consumption. As a result of the lower energy consumption, higher the residual energy hence extension of the network lifetime.

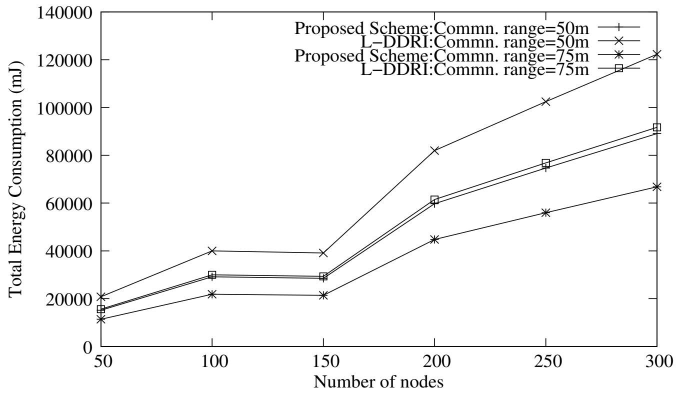

Fig. 14. Number of Nodes Vs. Total Energy Consumption

Figure 15 outlines the network overheads for varying number of nodes along with different communication range for proposed and L-DDRI schemes. It is observed that the overheads increases as number of node increases but comparing to existing scheme there is reduction in overheads for the proposed scheme.

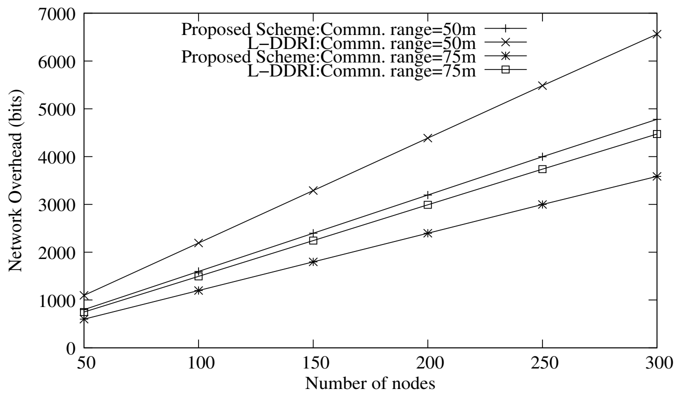

Fig. 15. Number of Nodes Vs. Network Overhead

The total number of packets received (PDR) by the base station from the source node for both proposed and L-DDRI scheme is shown in Fig. 16. It is observed that, there is slight decrement in the PDR for larger number SNs due to increase in the network overhead. As we increase the communication range, number of packet drop will be less which results in increase of PDR. When we compare the results, it is seen that the total number of packets received for the proposed scheme is higher than L-DDRI.

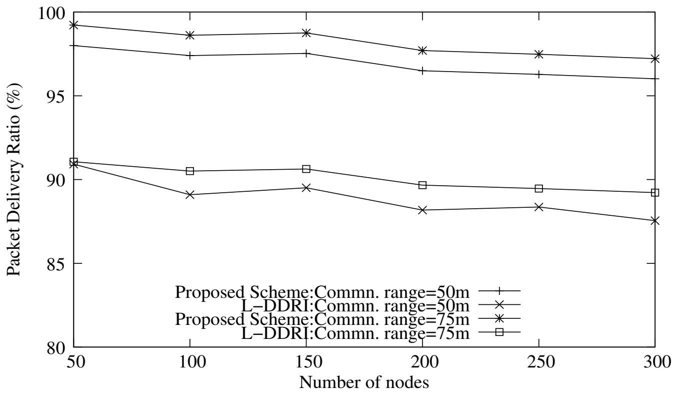

Fig. 16. Number of Nodes Vs. Packet Delivery Ratio

An unbalanced energy consumption between clusters due to non-uniform distribution of CHs will expedite the rate of node death. When the CHs are uniformly distributed, then members of every cluster will be closer together. By comparing the difference between the number of maximal and minimal cluster members (Difference Factor), it is possible to determine the evenly distribution of CHs. As illustrated in Fig. 17, it is observed that the proposed algorithm yields lesser difference value comparing with L-DDRI.

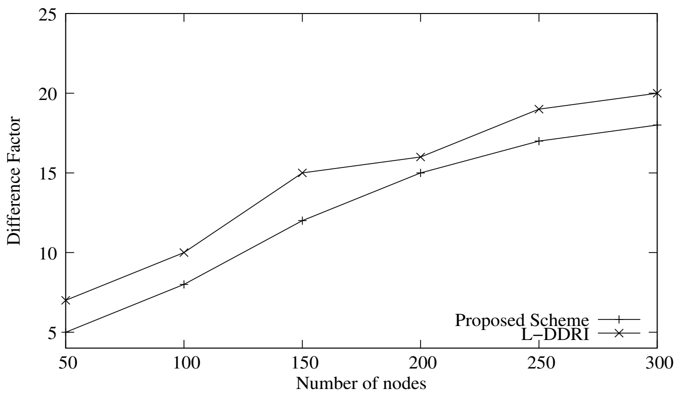

Fig. 17. Number of Nodes Vs. Difference Factor

Cluster Validity Index (CVI) is one of the parameter used to measure the quality and validity of clustering results. This also provides a quantifiable evaluation of the effectiveness of clustering algorithms. The Fig. 18 shows variation of the CVI with respect to number of SNs and observed that, proposed algorithm gives more compact and well-separated clusters.

Accuracy in clustering signifies extent to which a clustering algorithm distributes sensor nodes to the correct groupings considering some similarity criterion or distance metric. A high-accuracy clustering algorithm will be able to separate SNs effectively into meaningful and distinct categories while

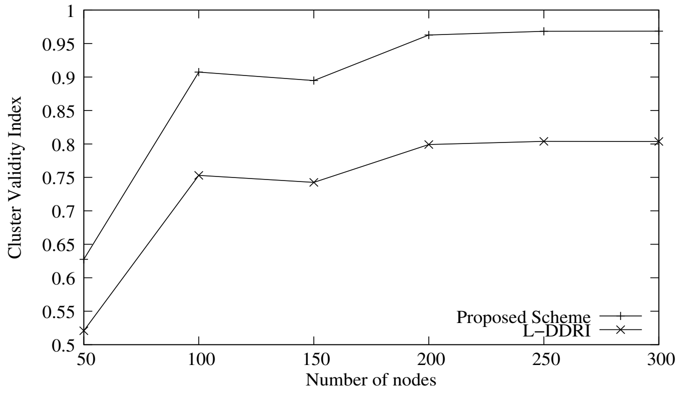

Fig. 18. Number of Nodes Vs. Cluster Validity Index

minimizing cluster overlap. Clustering accuracy for both proposed and L-DDRI scheme is shown in Fig. 19. It is observed that, increase in number of sensor nodes enhance the quality of clustering results of proposed scheme. A greater number of nodes will capture broader range of information, resulting in better clusters and improved cluster separation.

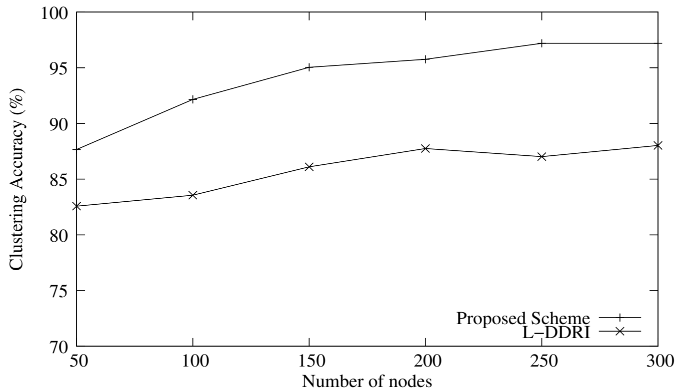

Fig. 19. Number of Nodes Vs. Clustering Accuracy

The residual energy in WSN refers to the amount of energy left in the nodes after they have performed their tasks. The standard deviation of residual energy provides useful information regarding the distribution of energy in the network. High value of standard deviation indicates the significant variation in the energy levels of the nodes, which may cause some nodes to run out of energy earlier than others. The graph in Fig. 20 shows the variation of SDRE with respect to number of SNs for both proposed and L-DDRI scheme. Low value of SDRE provides that the energy is distributed uniformly across the network, which can improve the network's lifetime and hence its reliability.

VI. Conclusion

The main issues with the WSN are energy consumption, node life cycle, and performance. Clustering is widely used to reduce energy consumption and improve network stability. The proposed K-Means and Fuzzy based Hybrid Clustering Algorithm for WSN approach deals with the issue of increasing network lifetime while balancing load throughout the network's sensor nodes. We have also analyzed the performance

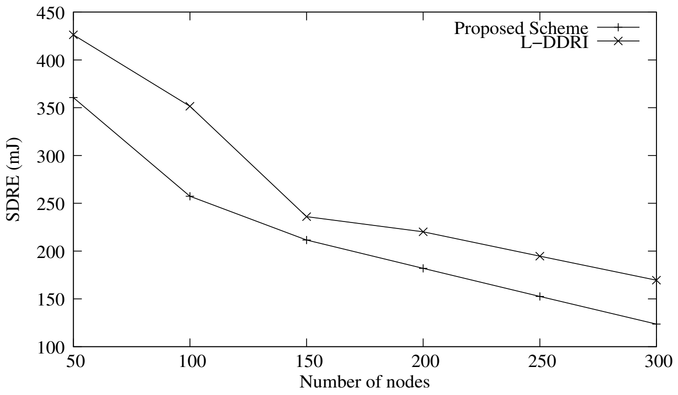

Fig. 20. Number of Nodes Vs. Standard Deviation of Residual Energy

metrics such as energy consumption, residual energy, end to end delay, overhead, computational and CH identification delay and clustering accuracy by varying different number of nodes.

The simulation results show that, the proposed scheme performs better as compared to the L-DDRI algorithm. Communication range is varied at different levels and it is seen that, at higher communication range level performance metrics are improved compared with existing L-DDRI scheme. In our work, only three parameters and nine linguistic variables are considered for the selection of cluster head using fuzzy logic. In future more number of parameters and linguistic variables can be considered to improve the network performance.

References

[1] Merabtine Nassima, Djamel Djenouri and Djamel-Eddine Zegour, "Towards energy efficient clustering in wireless sensor networks: A comprehensive review", *IEEE Access*, vol. 9, pp. 92688-92705, 2021. https://doi.org/10.1109/ACCESS.2021.3092509

[2] Verma, Sandeep, Neetu Sood, and Ajay Kumar Sharma, "Genetic algorithm-based optimized cluster head selection for single and multiple data sinks in heterogeneous wireless sensor network", *Applied Soft Computing*, vol. 85, 2019. https://doi.org/10.1016/j.asoc.2019.105788

[3] Primeau, Nicolas, Rafael Falcon, Rami Abielmona, and Emil M. Petriu, "A review of computational intelligence techniques in wireless sensor and actuator networks", *IEEE Communications Surveys and Tutorials*, vol. 20, no. 4, pp. 2822-2854, 2018. https://doi.org/10.1109/COMST.2018.2850220

[4] Amutha, J., Sandeep Sharma and Sanjay Kumar Sharma, "Strategies based on various aspects of clustering in wireless sensor networks using classical, optimization and machine learning techniques: Review, taxonomy, research findings, challenges and future directions", *Computer Science Review*, vol. 40, 2021. https://doi.org/10.1016/j.cosrev.2021.100376

[5] Raj, Jennifer S., "Machine learning based resourceful clustering with load optimization for wireless sensor networks", *Journal of Ubiquitous Computing and Communication Technologies (UCCT)*, vol. 2, no. 01, pp. 29-38, 2020. https://doi.org/10.36548/jucct.2020.1.004

[6] Panchal, Akhilesh, and Rajat Kumar Singh, "EHCR-FCM: Energy efficient hierarchical clustering and routing using fuzzy C-means for wireless sensor networks", *Telecommunication Systems*, vol. 76, no. 2, pp. 251-263, 2021. https://doi.org/10.1007/s11235-020-00712-7

[7] Shahidinejad Ali and Saeid Barshandeh, "Sink selection and clustering using fuzzy-based controller for wireless sensor networks", *International Journal of Communication Systems*, vol. 33, no. 15, 2020. https://doi.org/10.1002/dac.4557

[8] Sinaga Kristina P., and Miin-Shen Yang, "Unsupervised K-means clustering algorithm", *IEEE access*, vol. 8, pp. 80716-80727, 2020. https://doi.org/10.1109/ACCESS.2020.2988796

[9] Mouton Jacques P., Melvin Ferreiraand Albertus SJ Helberg, "A comparison of clustering algorithms for automatic modulation classification", *Expert Systems with Applications*, vol. 151, 2020. https://doi.org/10.1016/j.eswa.2020.113317

[10] Hassan Ali Abdul-hussian, Wahidah Md Shah, Mohd Fairuz Iskandar Othman and Hayder Abdul Hussien Hassan, "Evaluate the performance of K-Means and the fuzzy C-Means algorithms to formation balanced clusters in wireless sensor networks", *International Journal of Electrical and Computer Engineering*, vol. 10, no. 2, 2020. (2088-8708) 10, no. 2 (2020). http://doi.org/10.11591/ijece.v10i2.pp1515-1523

[11] Angadi Basavaraj M., Mahabaleshwar S. Kakkasageri, and Sunilkumar S. Manvi, "Computational intelligence techniques for localization and clustering in wireless sensor networks", *In Recent Trends in Computational Intelligence Enabled Research*, Academic Press, pp. 23-40, 2021. https://doi.org/10.1016/B978-0-12-822844-9.00011-6

[12] Ahmed Mohiuddin, Raihan Seraj and Syed Mohammed Shamsul Islam, "The k-means algorithm: A comprehensive survey and performance evaluation", *Electronics*, vol. 9, no. 8, 2020. https://doi.org/10.3390/electronics9081295

[13] Rezaee, Mustafa Jahangoshai, Milad Eshkevari, Morteza Saberi and Omar Hussain, "GBK-means clustering algorithm: An improvement to the K-means algorithm based on the bargaining game", *Knowledge-Based Systems*, Vol. 213, 2021. https://doi.org/10.1016/j.knosys.2020.106672

[14] Bai Liang, Jiye Liang and Fuyuan Cao, "A multiple k-means clustering ensemble algorithm to find nonlinearly separable clusters", *Information Fusion*, vol. 61, pp. 36-47, 2020. https://doi.org/10.1016/j.inffus.2020.03.009

[15] Jlassi Wadii, Rim Haddad, Ridha Bouallegue and Raed Shubair, "A combination of K-means Algorithm and Optimal Path Selection Method for Lifetime Extension in Wireless Sensor Networks", *International Conference on Advanced Information Networking and Applications*, Springer, pp. 416-425, 2021. https://doi.org/10.1007/978-3-030-75078-7$_4$2

[16] Ghazal, T.M., Hussain, M.Z., Said, R.A., Nadeem, A., Hasan, M.K., Ahmad, M., Khan, M.A. and Naseem, M.T., "Performances of K-means clustering algorithm with different distance metrics", *Intelligent Automation and Soft Computing*, vol. 30, no.2, pp. 735-742, 2021. https://doi.org/10.32604/iasc.2021.019067

[17] Rajaram V. and N. Kumaratharan, "Multi-hop optimized routing algorithm and load balanced fuzzy clustering in wireless sensor networks", *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 3, pp. 4281-4289, 2021. https://doi.org/10.1007/s12652-022-04273-2

[18] Lata Sonam, Shabana Mehfuz, Shabana Urooj and Fadwa Alrowais, "Fuzzy clustering algorithm for enhancing reliability and network lifetime of wireless sensor networks", *IEEE Access*, vol. 8, pp. 66013-66024, 2020. https://doi.org/10.1109/ACCESS.2020.2985495

[19] Hamzah Abdulmughni, Mohammad Shurman, Omar Al-Jarrah and Eyad Taqieddin, "Energy-efficient fuzzy-logic-based clustering technique for hierarchical routing protocols in wireless sensor networks", *Sensors*, vol. 19, no. 3, 2019. https://doi.org/10.3390/s19030561

[20] Rajput Anagha and Vinoth Babu Kumaravelu, "Fuzzy-based clustering scheme with sink selection algorithm for monitoring applications of wireless sensor networks", *Arabian Journal for Science and Engineering*, vol. 45, no. 8, pp. 6601-6623, 2020. https://doi.org/10.1007/s13369-020-04564-w

[21] Chauhan Vinith and Surender Soni, "Energy aware unequal clustering algorithm with multi-hop routing via low degree relay nodes for wireless sensor networks", *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 2, pp. 2469-2482, 2021.https://doi.org/10.1007/s12652-020-02385-1

[22] Mehra Pawan Singh, "E-FUCA: enhancement in fuzzy unequal clustering and routing for sustainable wireless sensor network", *Complex and Intelligent Systems*, vol. 8, no. 1, pp. 393-412, 2022. https://doi.org/10.1007/s40747-021-00392-z

[23] Dwivedi Anshu Kumar and Awadhesh Kumar Sharma, "EE-LEACH: Energy Enhancement in LEACH using Fuzzy Logic for Homogeneous WSN", *Wireless Personal Communications*, vol. 120, no. 4 pp. 3035-3055, 2021. https://doi.org/10.1007/s11277-021-08598-7

[24] Vasudha and Anoop Kumar, "Probabilistic Based Optimized Adaptive Clustering Scheme for Energy-Efficiency in Sensor Networks", *International Journal of Computer Networks and Applications*, vol. 8, no. 3, 2021. https://doi.org/10.22247/ijcna/2021/209187