A Dynamic K-means Based Clustering Algorithm Using Fuzzy Logic for CH Selection and Machine Learning Based Data Transmission

Anupam Choudhary (choudharyanupam.knpc@gmail.com)
Kalaniketan Polytechnic college Jabalpur,MP,India https://orcid.org/0000-0002-8578-7687

Abhishek Badholia
MATS University

Anurag Sharma
MATS University

Brijesh Patel
MATS University

Sapna Jain
MATS University

Research Article

Keywords: Wireless Sensor Network, Dynamic K-means, Fuzzy logic, Machine learning, Clustering

Posted Date: November 1st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-172424/v1

A Dynamic K-means based clustering algorithm using fuzzy logic for CH selection and machine learning based data transmission

Anupam Choudhary*
Department of Computer Science and Engineering, Kalaniketan Polytechnic College, Jabalpur, India
Email: choudharyanupam.knpc@gmail.com

Dr. Abhishek Badholia
Department of Computer Science and Engineering
MATS School of Engineering and Information Technology, Aarang, Raipur, India
Email: drabhishekb@matsuniversity.ac.in

Dr. Anurag Sharma
Department of Computer Science and Engineering
MATS School of Engineering and Information Technology, Aarang, Raipur, India
Email: anusiraag@gmail.com

Dr. Brijesh Patel
MATS School of Engineering and Information Technology, Aarang, Raipur, India
Email: brijeshp@matsuniversity.ac.in

Sapna Jain
Department of Computer Science and Engineering
MATS School of Engineering and Information Technology, Aarang, Raipur, India
Email: choudharysapnajain@gmail.com

*Corresponding author

Abstract

Clustering is effective method to increase network lifetime, energy efficiency, and connectivity of Sensor nodes in wireless sensor network. An energy efficient clustering algorithm has been proposed in this paper. Sensor nodes are clustered using K-means algorithm which dynamically forms number of clusters in accordance with number of alive nodes. Selection of suitable CH is done by fuzzy inference system by choosing three fuzzy input variable such as residual energy of Sensor node, its distance from cluster center and base station. Amount of data transmitted by member nodes to CH is reduced by machine learning that classify similar data at regular interval. The simulation results show that proposed algorithm outperforms other cluster based algorithms in terms of data received by base station, number of alive node per round, time of first node, middle node and last node to die for various density of sensor nodes and scalable conditions.

Keywords:

Wireless Sensor Network; Dynamic K-means; Fuzzy logic; Machine learning; Clustering

1. Introduction

Wireless Sensor Networks (WSNs) are self-organized network to collect various information of the surroundings. Sensor node(SN) has limited computational power, storage capacity and lifetime of batteries[1]. These restrictions instigate development of energy aware routing protocol. Applications of WSNs are traffic management, healthcare, military surveillance, fire detection in forest, flood warning, habitat monitoring, agriculture, industries, smart home, volcano monitoring, security, military surveillance maritime search and rescue etc. [2-13].

Hierarchical cluster based routing algorithm partitions SNs into number of clusters. Each cluster has a cluster head (CH) and number of member nodes. Member nodes transmit sensed data to their CH. After receiving data from all member nodes CH performs data aggregation and fusion to reduce amount of data. Then CH transmit data to base station (BS). Fig. 1 shows cluster based routing of data towards BS.

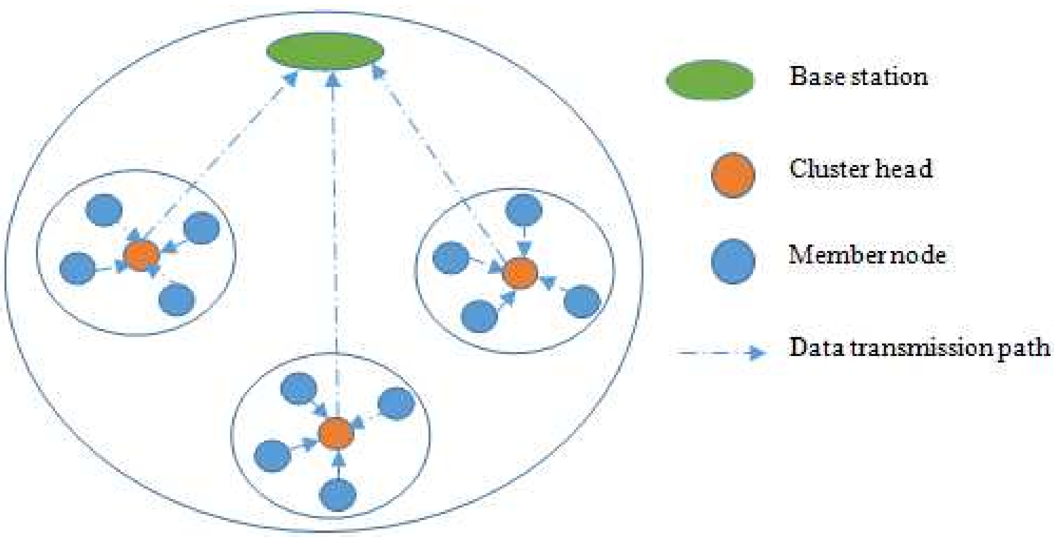

Fig. 1. Concept of clustering in WSN [7]

The proposed approach of developing the energy efficient clustering protocol is driven by following questions:

- What is a good clustering?
- How to find optimal number of clusters to be formed?
- What factors should be considered for selection of CH?

- How to reduce data transmission by member nodes to CH?

For good clustering intra-cluster similarities should be high and inter-cluster similarity must have to be low. Number of clusters to be formed should be in accordance with number of alive node. Factors that should be considered for determination of CH are location of the SN relative to BS and center of cluster, its residual energy, confidence and trust. To reduce overhead on message transmission from member nodes to CH similar pattern in sensed data must be identified and only one copy should be transmitted for every set of similar data.

In this paper, authors have proposed a new protocol for wireless sensor network that uses dynamic K-means algorithm for efficient clustering and optimal number of clusters formation. Selection of suitable CH for each cluster is determined by fuzzy inference system (FIS) that considers residual energy, distance from center of cluster and distance from BS of each SN. Machine learning is used by member nodes to find similar reading in sensed data. All distinct readings and one reading for every set of similar reading are forwarded to CH from member nodes. It results in reduction in data transmission from member nodes to CH. Result of simulations show that proposed protocol has significantly improved network lifetime of WSN.

Rest of the paper is organized as follows: Section 2 presents the related work in the field of cluster based protocols. Section 3 describes energy model adopted. Section 4 defines Methodology used for construction of proposed routing algorithm. Section 5 shows performed simulations in MATLAB and their results. Finally, section 6 provides conclusion and scope of future work.

2. Literature Review

In this section, most of well-known routing protocols have been discussed. Heinzelman et al. [14] have proposed a Low-energy adaptive clustering hierarchy (LEACH) protocol that introduced the concept of clusters in WSN. It is based on probabilistic model and each node has equal probability to become CH. Process of routing of data is simple and does not require much information. Major disadvantages of LEACH are: (i)Residual energy of SN is not considered in choosing CH. (ii) Clusters formed are not uniform.Heinzelman et al.[15] have described LEACH-C protocol that uses a centralized control technique that uses location information of the s. Base station forms clusters on the basis of SNs current location and energy level resulting in more balanced clusters formed by using the LEACH algorithm. Lindsey and Raghavendra [16] have proposed a Power-Efficient Gathering in Sensor Information Systems (PEGASIS). It uses greedy algorithm to organize SNs in form of a chain. Each node receive from and transfer data to its close neighbor. Fan and Song [17] have presented a Multi-hop LEACH (M-LEACH) protocol for multi-hop communication that takes scalability into consideration. Its negative aspect is that it can not be implemented in heterogeneous sensor network. Beiranvand et al. [18] have developed Improved LEACH (I-Leach) that select CH based on minimum distance from BS, larger remaining energy and more number of neighbors. For cluster formation and data transmission this algorithm considers the distance of SN from CH as well from BS. If BS is nearer to SN it send data directly to BS instead of CH. Yassein et al. [19] have described Vice-LEACH (VLEACH) algorithm that vice-CH in addition to CH, and member nodes. Vice-CH takes the responsibility of CH when it dies. The major flaw of this algorithm is that if vice CH Dies, it does not provide solution for this condition. Rabiaa et al. [20] have proposed an algorithm that uses K-means clustering using Davies-Bouldin index which is ratio of within-cluster and between-cluster distances. For optimal clustering value of Davies-Bouldin index must be as low as possible. Then Gaussian elimination algorithm is used to select the CH. Jerbi et al. [21] have developed Orphan-LEACH (O-LEACH) that aims to cover SNs which do not belong any cluster due to far away deployment. A cluster member perform the role of a gateway or CH that allows the joining of orphan nodes. Rajput and Kumaravelu [7] have used fuzzy c-means clustering for cluster formation. Selection of CH is based on level of centrality of a node in the cluster. Fuzzy c-means can not be used when number of clusters to be formed are not fixed. Kim et al. [22] have developed (CHEF) that uses fuzzy based approach to select CH. It is based on two parameters that are proximity distance and energy. Locally optimal node with high energy is elected as CH. In [23-25] some protocols have been described using fuzzy techniques. Machine learning (ML) techniques are very useful in WSN to reduce amount of data transmitted among SNs. It learn from their surroundings and based on their learning knowledge nodes transfer data to other nodes [26]. In Supervised learning, known input and their output are provided for learning purpose. This knowledge is used to predict result for unseen inputs. In unsupervised learning similarity in input data is used to classify them into different classes.

Review of literature summarized that different algorithms use discrete parameters like location of SNs, inter-cluster distance, residual energy, distance from BS and number of neighbours for CH selection but integrated approaches are not presented. The existing strategies suffers from significant overhead in data transmitted from member nodes to CH.

To overcome above issues an energy efficient dynamic K-means based protocol clustering approach for WSN has been proposed. The prime objective of this research is to increase network lifetime with selection of suitable CH by fuzzy inference system and reduction in data transmission from member nodes to CH by machine learning.

3. Radio energy model

Radio energy model is used for computation of energy dissipation during data transmission between transmitter and receiver. In WSN, data transmission consumes more energy than data processing [7]. SNs wirelessly transmit their data over a short range. Free space propagation model and multipath fading channel model [7,27] shown in Fig. 2 have been used in proposed work.

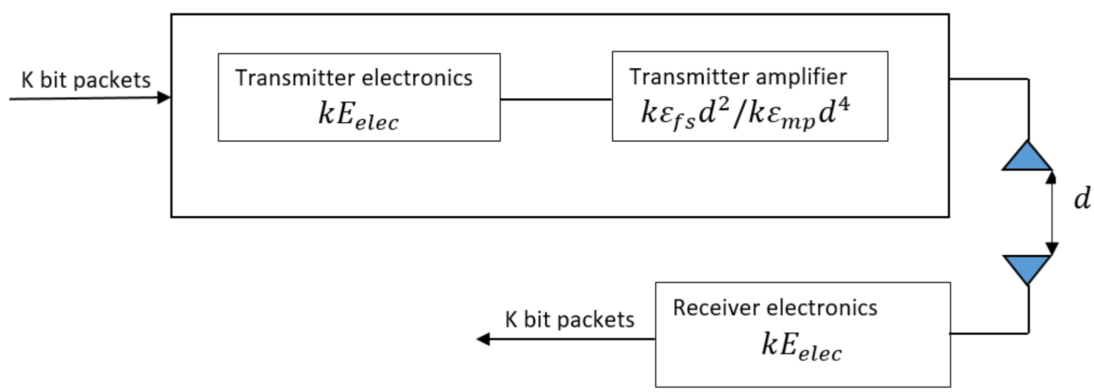

Fig. 2. Radio communication model

The dissipation of energy of the SN is calculated for the following operations:

(i) Transmission of data from cluster member to CH

Dissipation of energy for transmission of data from cluster member to CH is given by eq. (1):

$$E_{CM_{tx}CH} = \begin{cases} kE_{elec} + k\varepsilon_{fs}d^2 & if\ d \leq d_0 \\ kE_{elec} + k\varepsilon_{mp}d^4 & if\ d > d_0 \end{cases} \quad (1)$$

Here, $E_{CM_{tx}CH}$ is total energy required to transmit k bits of data from cluster member to CH.

E_{elec} is the energy consumption of electronic circuit of SNs.

ε_{fs} and ε_{mp} are energy required by amplifier at transmitter end for free space propagation and multipath fading channel model respectively .

d is the distance between cluster member and CH.

d_0 is reference distance calculated by eq. (2):

$$d_0 = \sqrt{\varepsilon_{fs}/\varepsilon_{mp}} \tag{2}$$

(ii) Reception of data at CH

The data transmitted by all cluster member is received at CH's receiver circuit. is the total energy required receive k bit of data from a cluster member node (E_{rx}) is computed by eq. (3):

$$E_{rx} = kE_{elec} \tag{3}$$

(iii) Transmission of data from CH to BS

CH aggregate the data received from all its cluster member and transmit it to BS. Amount of energy needed by a CH for aggregation and transmission is given by eq. (4):

$$E_{CH_{tx}BS} = \begin{cases} kE_{elec} + k\varepsilon_{fs}d^2 + E_{aggr} & if\ d \leq d_0 \\ kE_{elec} + k\varepsilon_{mp}d^4 + E_{aggr} & if\ d > d_0 \end{cases} \tag{4}$$

Here, $E_{CH_{tx}BS}$ is the total energy required to transmit K bits of data from CH to BS.
E_{aggr} is the energy required for data aggregation.
d is the distance between CH and BS.

Consider a WSN having q number of clusters and a cluster contains m number of SN. It results in (m-1) number of cluster member nodes and a CH in a cluster. Thus energy is consumed in (m-1) transmission from cluster member nodes to CH, (m-1) reception by CH and one transmission from CH to BS. Total energy consumed per round in a cluster ($E_{cluster}$) is determined by eq. (5):

$$E_{cluster} = (m-1)E_{CM_{tx}CH} + (m-1)E_{rx} + E_{CH_{tx}BS} \tag{5}$$

Amount of energy consumed per round in whole WSN (E_{round}) is given by eq. (6):

$$E_{round} = \sum_{i=1}^{q} E_{cluster_i} \tag{6}$$

4. Methodology

4.1. K-means Clustering Algorithm

K-means algorithm partitions a set of n object into k clusters based on similarities of objects [20]. It starts with randomly choosing k number of objects each of that initially represents a cluster mean or center. Then each of the remaining objects is assigned to the cluster having identical properties, based on the distance between the object and the cluster mean. It then computes the new mean for each cluster. This process repeats until the square-error criterion function given by eq. (7) converges to be optimal.

$$E = \sum_{i=1}^{k} \sum_{p \in C_i} \mathrm{dist}(p, m_i)^2 \qquad (7)$$

Here E is the sum of the square error for all objects in the data set, p represents an object and mi is the mean of characteristics of cluster Ci. Criterion function is used to make the resulting k clusters compact and distinct. The algorithm determines k partitions that minimize the criterion function resulting in compact and distinct clusters.

4.2. Fuzzy logic model

The fuzzy logic model shown in Fig. 4 consist of following modules

1. Fuzzifier: It convert crisp input values to fuzzy set.

2. Fuzzy rule base: It stores if-then rules.

3. Fuzzy inference engine: It takes input values and draws decision from fuzzy rules.

4. Defuzzifier: It converts the fuzzy variables to crisp values.

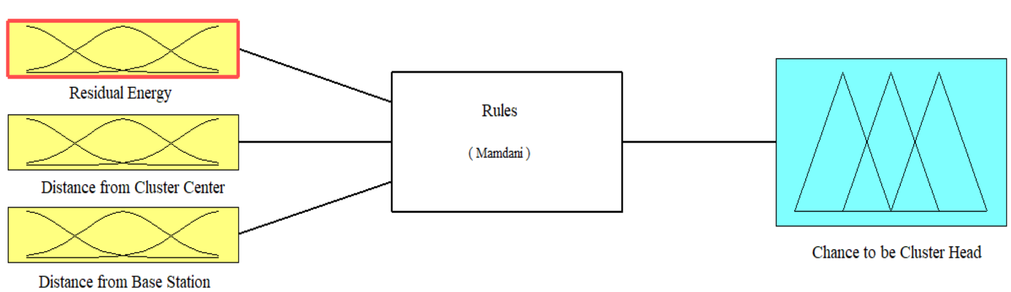

Fig. 3. Fuzzy system for proposed model

1. Fuzzication module

In proposed protocol, Mamdani's method fuzzy inference system (FIS) is used to select CH. Fig. 3 shows chosen input and output variables for the FIS. Three fuzzy input variables have been taken to elect CH.

Fig. 4(a) represent membership function plots for input variables residual energy (RE). The linguistic variable for this fuzzy set is very low, low, medium, high, and very high. Trapezoidal membership function has been considered by very low and very high variables. Triangular membership function has been taken by remaining variables.

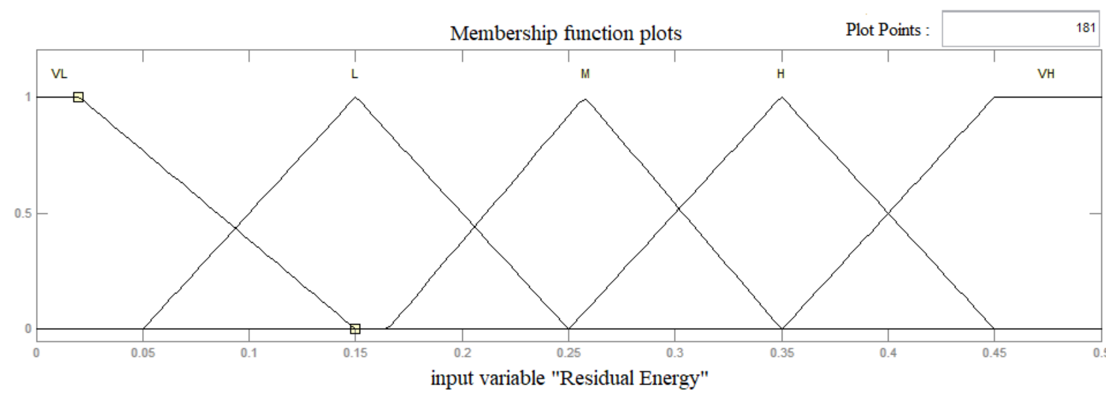

Fig. 4(a). Membership function for Residual Energy

Fig. 4(b) and shows membership for fuzzy set distance from cluster center (DCC). Near, medium and far are chosen as linguistic variable for this fuzzy set. The third fuzzy input variable is distance from BS (DBS). Its membership function has been shown by Fig. 4(c). Near, medium and far are considered as range of values for this fuzzy set.

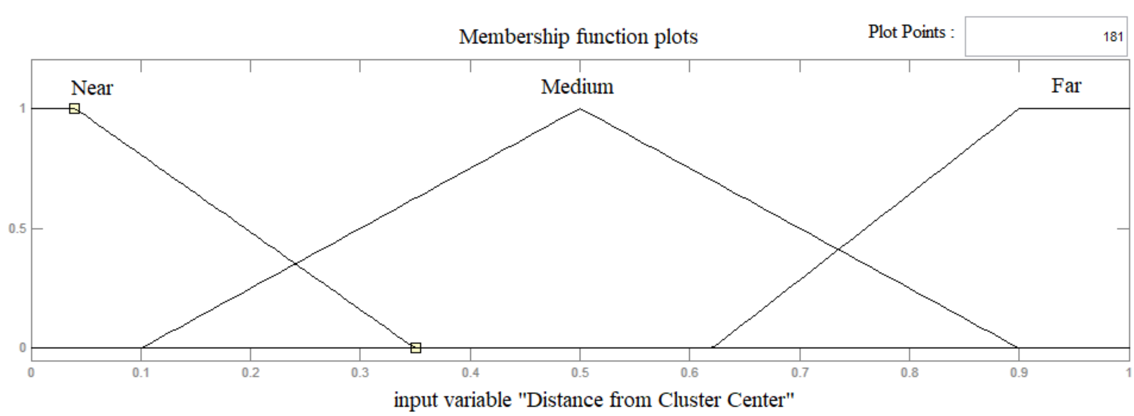

Fig. 4(b). Membership function plot for Distance from Cluster Center

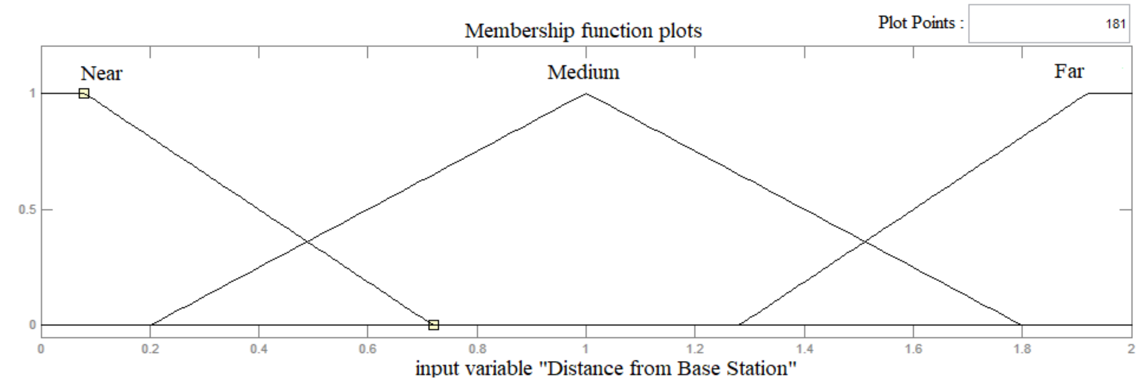

Fig. 4(c). Membership function plot for Distance from Base Station

2. Rule base and inference engine

Fuzzy inference system that has been considered uses 45 rules and Some of these are shown in Table 1. The form of rule is if A, B, C then O. 'A' represent fuzzy input variable residual energy. 'B' and 'C' represents input variables distance from cluster center and distance from BS. O represents chance to be CH (CCH) it is fuzzy inference output given by eq. (8)

$$CCH_i = FIS\ (RE_i, DCC_i, DBS_i) \quad (8)$$

Table 1
Sample Fuzzy rules

Residual Energy	Distance from Center of cluster	Distance from BS	Chance to be CH
Very Low	Far	Far	Very Weak
Very Low	Far	Medium	Very Weak
Very Low	Far	Near	Very Weak
Very Low	Medium	Far	Very Weak
Very Low	Medium	Medium	Very Weak
Very Low	Medium	Near	Very Weak
Very Low	Near	Far	Very Weak
Very Low	Near	Medium	Very Weak
Very Low	Near	Near	Weak
Very High	Far	Far	Weak
Very High	Far	Medium	Weak
Very High	Far	Near	Medium
Very High	Medium	Far	Weak
Very High	Medium	Medium	Strong
Very High	Medium	Near	Strong
Very High	Near	Far	Weak
Very High	Near	Medium	Strong
Very High	Near	Near	Very strong

Fig. 5 shows membership function for output variable chance to be CH (CCH). Very weak, weak, medium, strong and very strong are linguistic variables for this fuzzy set. The chance of a node to become CH is calculated by considering input parameters such as residual energy, distance from center of cluster and distance from BS by using fuzzy rules.

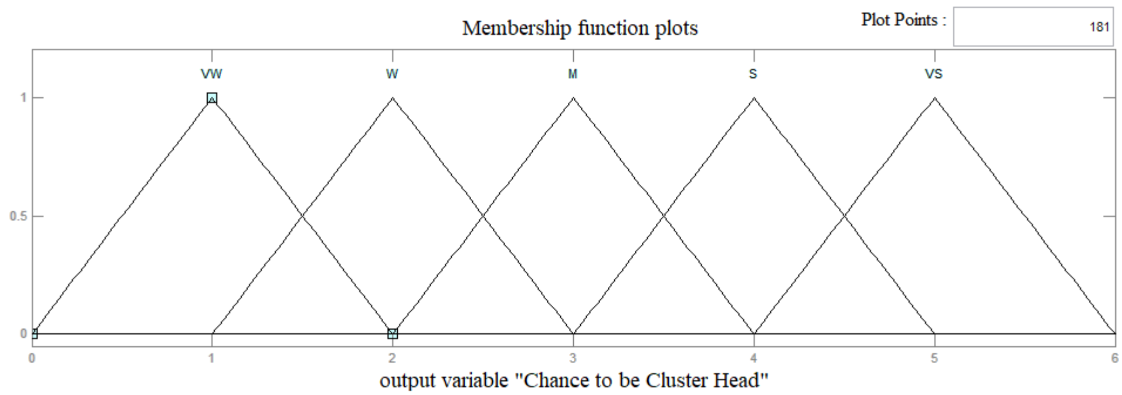

Fig. 5. Membership function plot for Chance to be Cluster Head

4.3. Machine learning model

Intel lab data set has been considered for classification with machine learning (ML) using python. A sample set of sensor reading for certain period has been taken for training and classification. Three attributes namely time, mote id and humidity has been selected from data set. A new variable "similarity" has been appended in the dataset. It contains value "similar" or "dissimilar". Classifier has been evaluated using following factors:

Precision factor: It is positive predictive value that indicates how good a model is at predicting the positive class. It is calculated as proportion of True Positive and Predicted Yes.

Recall factor: gives a measure of how correctly model is able to identify the relevant data. It is proportion of True Positive and Actual Yes.

F1-Score: It specifies what percent of the positive predictions are correct and is calculated by eq. (9)

F1-Score = 2*(Recall*Precision) / (Recall+Precision) (9)

Support: It is the number of actual occurrence of the class in the specified dataset.

An analysis of humidity data from considered data subset for certain time period is presented in Fig.6 of p4. Count shows number of occurrence of a humidity value. One reading is transmitted for every set of similar reading. This approach significantly reduce number of readings transmitted by member nodes to CH. It results in significant saving of energy of member nodes and increased network lifetime.

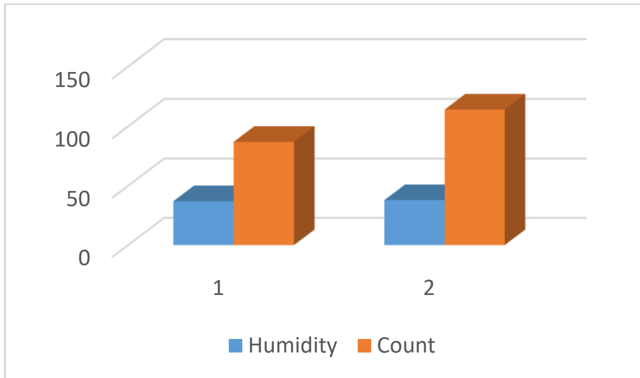

Fig. 6. Analysis of humidity data in the data subset

Table 2 shows the parameters of Random Forest classifier that has been evaluated using python.

Table 2
Parameters of Random Forest classifier

	Precision	**Recall**	**F1-Score**	**Support**
0	0.92	1	0.96	23
1	1	0.93	0.96	28
Avg/Total	0.96	0.96	0.96	51

4.4. Proposed algorithm (DKFM)

A Dynamic K-means based clustering algorithm using fuzzy logic for CH selection and machine learning based data transmission (DKFM) has been proposed on the basis of outcomes from literature review. This protocol uses dynamic K-means algorithm to form optimal number of clusters and reduction of intra cluster distance. A fuzzy inference system selects suitable CH by considering three fuzzy input variable (i) residual energy of SN (ii) distance of SN from cluster center (iii) distance of SN from base station. Amount of data transmitted by member nodes to CH is reduced by machine learning that classify similar data at regular interval. Following assumptions has been taken for proposed protocol:

- The SNs are randomly distributed in the target area.
- Network is homogeneous.
- All SNs and BS are stationary.
- Each node knows its residual energy and current position.
- All nodes are able to send the data to the BS.

The procedure of proposed routing protocol (DKFM) is as follows:

Step 1. Clustering using dynamic K-means

K-means algorithm is executed on target WSN having n nodes. It selects number of clusters to be formed (K) dynamically for each round by eq. (10)

$$K = \text{sqrt (initial nodes- dead nodes)}. \tag{10}$$

First it randomly selects k out of n nodes as the initial center. Each of the remaining nodes decides its cluster center nearest to it according to the Euclidean distance. After each of the nodes in the network is assigned to one of k clusters, the center of each cluster is calculated and all objects are reassigned using the updated cluster center calculated by eq. (11).

$$\text{Center (x,y)}=(1/n \sum_{i=1}^{n} Xi \,, 1/n \sum_{i=1}^{n} Yi) \tag{11}$$

This process is recursively executed until clusters formed in current round are identical as those formed in the previous round.

Step 2. Fuzzy based selection of CH

After the formation of clusters, FIS described in section 4.2 is used to select CH for each cluster. Then node selected as CH broadcast its status to all other nodes in the cluster.

Step 4. Schedule Creation

The selected CHs create TDMA schedule to define the time slot for each member in its cluster to forward data to it.

Step 5. Machine learning based data transmission

All cluster member send data to their CH using machine learning (described in section 4.3) in their allocated time slot. CH aggregates the received data from all member nodes and sends it to BS.

Step 6. Count dead nodes and alive nodes. If (alive node>0) start new round.

The above procedure has been represented by a

Fig. 7flow chart (Fig. 7.) and Algorithm 1.

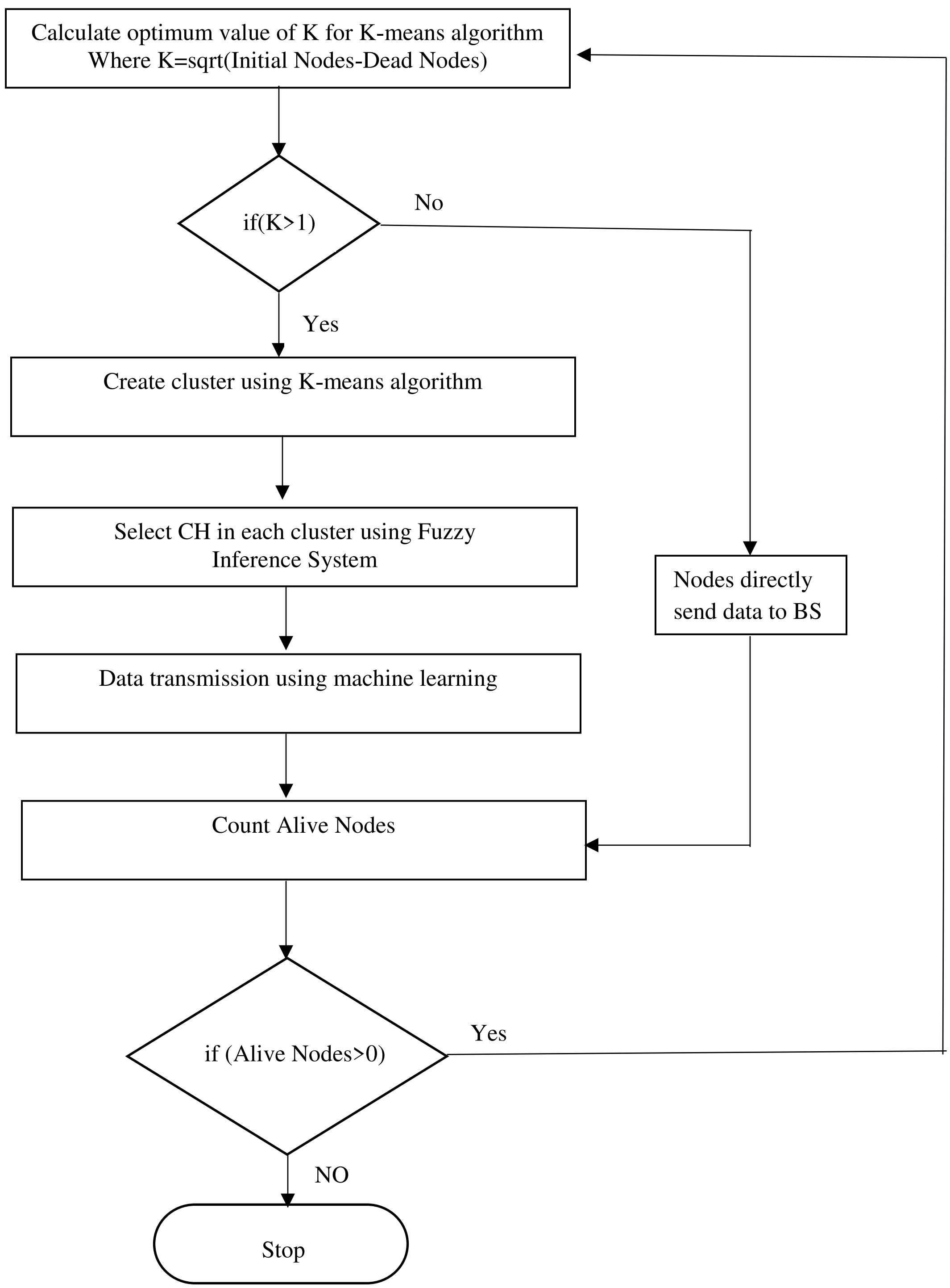

Fig. 7. Flow chart of proposed protocol

Algorithm 1. Proposed DKFM algorithm

```
Algorithm DKFM (CE).
{
// CE is matrix of n×3 dimension and denoted as: CE= {S1, S2, S3,………, Sn}
// n represent total no of SNs in WSN. Variable Si denotes X, Y
// coordinates and energy (E) of ith SN.
initial_node=n; dead_node=0;
do
{
  K= sqrt (initial_node-dead_node);
  if (k>1)
  {
    Create cluster using K-means algorithm;
     for i=1 to k do
      {
        for j=1 to size of cluster_i do
        {
             CH_i = Select S_j using Fuzzy inference system;
             CH_i creates schedule for each member of cluster_i;
             Cluster member send data to CH_i using machine learning
             CH_i aggregate data and send it to BS;
```

```
            // CHi  is cluster head of ith cluster and BS is base station

        }
      }

}

 else

   Each Sj directly send data to BS;

 for all Sj € CE whose s_j^E=0

  {

dead_node=dead_node+1;

alive_node=initial_node - dead_node;

  }

} while (alive_node > 0);
```

5. Simulation results and discussion

The proposed protocol(DKFM) is compared with LEACH[14] and I-LEACH[18] in terms of network lifetime, number of alive node per round, data received by base station, time of first node, middle node and last node to die. MATLAB R2016a tool is used to implement LEACH, I-LEACH and proposed protocol. Table denotes the various network simulation parameters and their values that have been considered [12,28].

Table 3
Network parameters used for simulation

Parameters	Values
Sensing area size	100 m×100 m, 100 m×150 m, 200 m×200 m, 200 m×250 m
Location of BS	(50,50)
Number of SNs	100,150,200,300,400
Initial energy of SNs	0.5J
Energy consumption of electronic circuit (E_{elec})	50 nJ/bit
Energy consumption for data aggregation (E_{aggr})	5nJ/bit/message
Free space communication energy(ε_{fs})	10pJ/bit/m^2
Multipath communication energy(ε_{mp})	0.0013pJ/bit/m^4
Data packet size(k)	2000 bits

Simulation has been carried out for increase in node density and size of sensing area for evaluating the performance of the DKFM against I-LEACH and LEACH. Table 4 to Table 7 denotes the statistics for increase in node density of the WSN. Ten iterations for 100,200,300 and 400 randomly deployed nodes in sensing area of 100 m×100 m and location of sink at (50,50) have been performed. Performance of algorithms is represented in terms of Node Dead First (NDF), Node Half Dead (NHD) and Node Dead Last (NDL). The stability of network is described by number of SN remain alive for long period of time. NHD values are used to represent network stability.

Table 4

Network lifetime on the basis of NDF, NHD, NDL for sensing area=100 m×100 m, no of SNs=100

Iteration No.	DKFM			I-LEACH			LEACH		
	NDF	NHD	NDL	NDF	NHD	NDL	NDF	NHD	NDL
1	3093	4310	4593	1318	1617	3368	1248	1599	2218
2	2627	4321	4590	1262	1595	3143	1204	1606	2329
3	3208	4297	4649	1206	1614	3351	1209	1648	2157
4	3356	4301	4606	1290	1601	3626	1294	1623	2366
5	3243	4300	4588	1290	1577	4361	1295	1674	2520
6	3222	4322	4609	1300	1614	4669	1219	1596	2582
7	3093	4310	4593	1318	1617	3368	1248	1599	2218
8	3214	4306	4659	1186	1584	2207	1284	1620	2213
9	2960	4269	4621	1267	1613	3790	1252	1592	2299
10	2857	4358	4557	1307	1607	2426	1236	1602	2280
Average	3087	4309	4607	1274	1604	3431	1249	1616	2319

Table 5

Network lifetime on the basis of NDF, NHD, NDL for sensing area=100 m×100 m, no of SNs=200

Iteration No.	DKFM			I-LEACH			LEACH		
	NDF	NHD	NDL	NDF	NHD	NDL	NDF	NHD	NDL
1	2767	4552	4797	1145	1391	2173	1138	1373	1806
2	2777	4556	4769	1133	1398	2636	1141	1380	1880
3	2503	4561	4765	1141	1397	3029	1111	1384	1971
4	2439	4547	4813	1136	1388	3213	1120	1377	1905
5	3011	4549	4740	1147	1391	2032	1126	1393	1801
6	3048	4567	4766	1176	1393	2301	1145	1378	1848
7	3032	4571	4738	1176	1391	3143	1128	1390	1887
8	3203	4540	4770	1086	1390	2251	1117	1391	2001
9	2828	4551	4794	1121	1400	3863	1128	1390	1834
10	3144	4571	4784	1150	1383	2345	1087	1372	1913
Average	2875	4557	4774	1141	1392	2464	1124	1383	1885

Table 6

Network lifetime on the basis of NDF, NHD, NDL for sensing area=100 m×100 m, no of SNs=300

Iteration No.	DKFM			I-LEACH			LEACH		
	NDF	NHD	NDL	NDF	NHD	NDL	NDF	NHD	NDL
1	3069	4638	4828	1084	1395	2789	1128	1388	2027
2	3168	4636	4814	1145	1408	2347	1123	1391	1916
3	3273	4629	4833	1122	1403	3446	1151	1390	1956
4	3026	4627	4815	1145	1401	2817	1128	1397	1870
5	3069	4638	4828	1084	1395	2789	1128	1388	2027
6	3117	4625	4822	1071	1403	2009	1107	1395	1840
7	2886	4653	4834	1127	1407	3050	1170	1399	1818
8	3130	4650	4816	1110	1394	3290	1121	1386	1856
9	3302	4641	4819	1080	1402	2337	1080	1392	1860
10	3379	4625	4842	1061	1407	2522	1096	1392	1998
Average	3142	4636	4825	1103	1402	2740	1123	1399	1917

Table 7
Network lifetime on the basis of NDF, NHD, NDL for sensing area=100 m×100 m, no of SNs=400

Iteration No.	DKFM			I-LEACH			LEACH		
	NDF	NHD	NDL	NDF	NHD	NDL	NDF	NHD	NDL
1	3287	4667	4849	1286	1613	3358	1276	1632	2612
2	3206	4675	4857	1339	1684	3109	1305	1640	2320
3	3396	4683	4855	1197	1605	3544	1267	1614	2338
4	2821	4701	4890	1252	1639	2750	1245	1635	2559
5	2444	4692	4852	1163	1643	2665	1169	1639	2310
6	3289	4664	4851	1238	1621	3781	1272	1628	2494
7	3018	4671	4862	1316	1685	2575	1278	1620	2563
8	3368	4677	4869	1251	1635	3219	1259	1637	2354
9	3476	4693	4861	1324	1634	4077	1299	1641	2199
10	3296	4669	4845	1252	1610	2923	1257	1632	2462
Average	3160	4679	4859	1262	1637	3200	1263	1632	2421

Results shows that DKFM has achieved average time of NDF 142% or 1813 number of rounds, 147% or 1838 number of rounds for 100 nodes; 152% or 1734 number of rounds, 156% or 1751 number of round for 200 nodes; 185% or 2039 number of round, 180 % or 2019 number of round for 300 nodes; 150% or 1898 number of round, 150% or 1897 number of round for 400 nodes better than I-LEACH and LEACH respectively.

Stability period of DKFM is 169 % or 2705 number of rounds, 167% or 2693 number of rounds for 100 nodes; 227% or 3165 number of rounds, 230% or 3174 number of round for 200 nodes; 231% or 3234 number of round, 231 % or 3237 number of round for 300 nodes; 186% or 3042 number of round, 187% or 3047 number of round for 400 nodes better than I-LEACH and LEACH respectively.

Average time of NDL of DKFM is 34% or 1176 number of rounds, 99% or 2288 number of rounds for 100 nodes; 94% or 2310 number of rounds, 153% or 2889 number of round for 200 nodes; 76% or 2085 number of round, 152% or 2908 number of round for 300 nodes; 52% or 1659 number of round, 101% or 2438 number of round for 400 nodes better than I-LEACH and LEACH respectively.

Fig.8 (a-d) and Fig.9 (a-d) show the graphs representing number of alive nodes per round and data received by base station for 100,200,300,400 nodes respectively for sensing area size 100 m×100 m and sink is located at (50,50). In all scenarios proposed algorithm is more efficient than I-LEACH and LEACH.

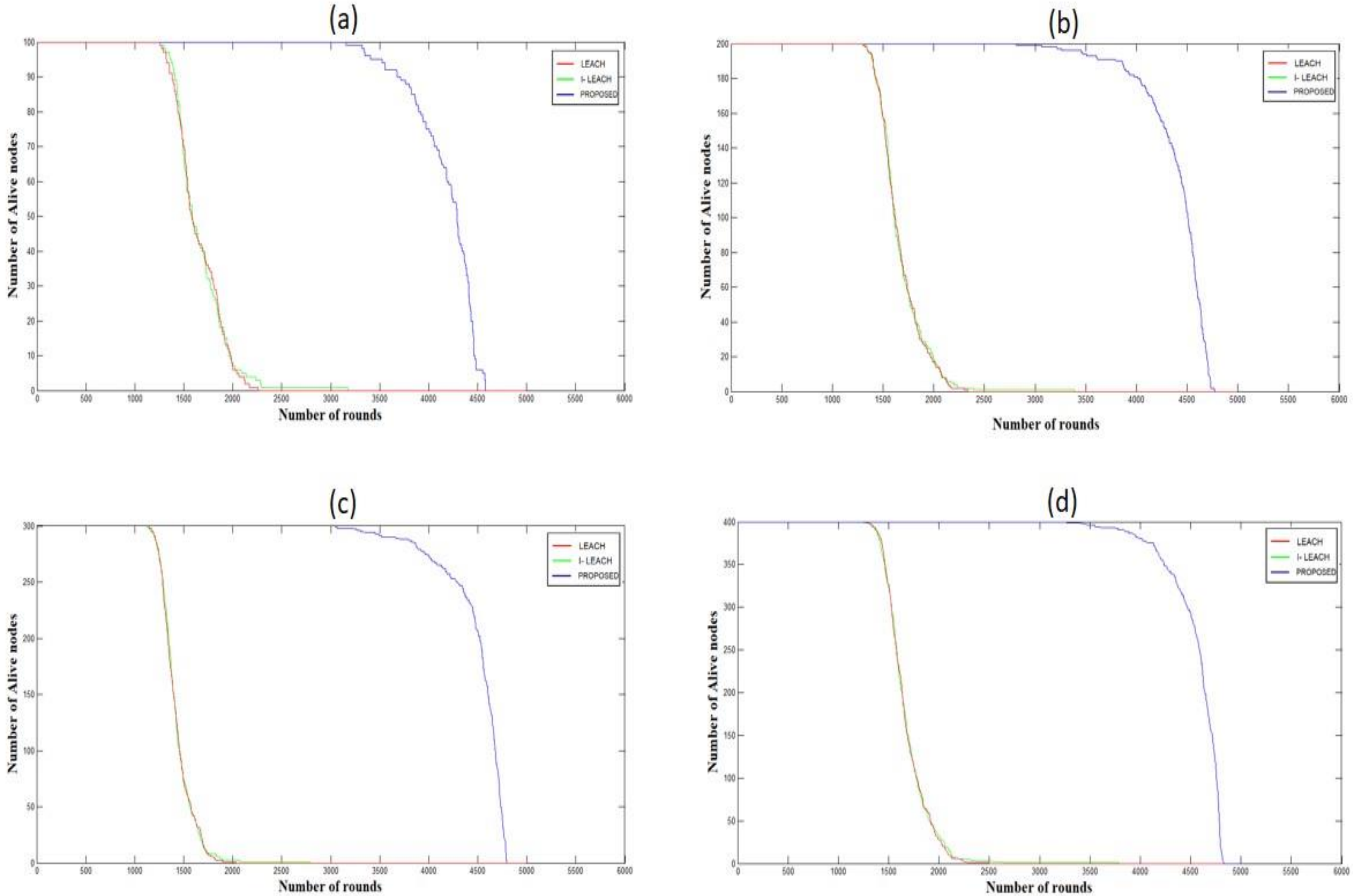

Fig. 8 (a-d). Number of Alive nodes vs Rounds for sensing area of 100 m x 100 m with 100, 200, 300 and 400 SNs respectively

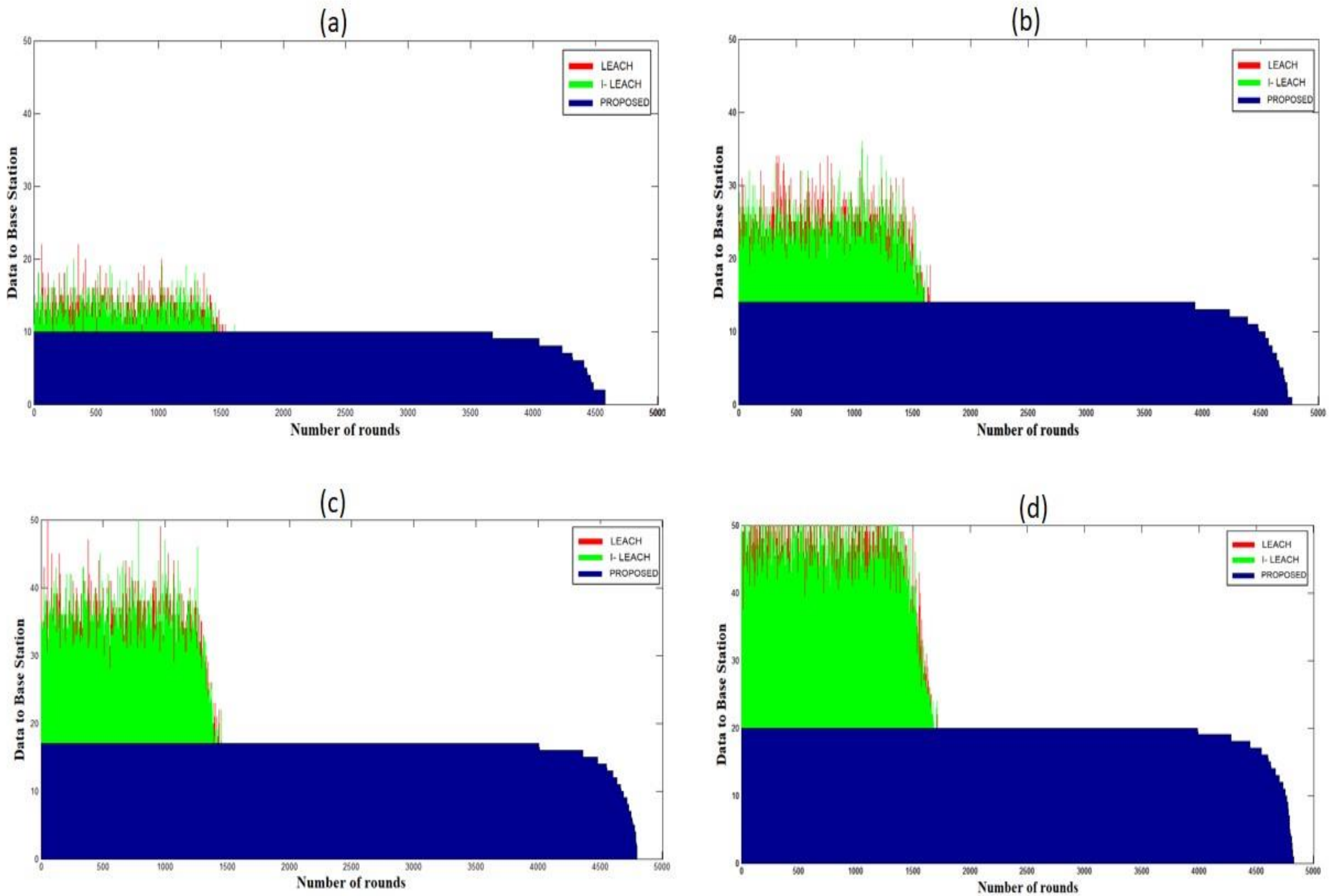

Fig. 9 (a-d). Data received by BS for sensing area of 100 m x 100 m with 100, 200, 300 and 400 SNs respectively

The algorithms considered are also evaluated for increase in size of sensing area. We have considered 150 number of SN in sensing area having sizes of 100 m×100 m, 100 m×150 m, 200 m×200 m, 200 m×250 m. BS is located at (50,50) . Table **8**8 to Table 11 represent the statistics for increase in size of sensing area in terms of NDF, NHD and NDL for five random deployment of SN. Increase in size of sensing area increases transmission distances resulting in more energy consumption as per the explanation shown in section 4.

Table 8
Network lifetime on the basis of NDF, NHD, NDL for sensing area=100 m×100 m, no of SNs=150

Iteration No.	DKFM			I-LEACH			LEACH		
	NDF	NHD	NDL	NDF	NHD	NDL	NDF	NHD	NDL
1	3206	4465	4663	1243	1602	4041	1277	1611	2258
2	3307	4434	4699	1319	1611	2758	1322	1618	2369
3	3017	4492	4709	1305	1587	3442	1275	1605	2414
4	3122	4491	4791	1223	1619	2754	1264	1609	2576
5	2666	4462	4740	1274	1607	2603	1321	1614	2344
Average	3064	4469	4720	1273	1605	3120	1292	1611	2392

Table 9
Network lifetime on the basis of NDF, NHD, NDL for sensing area=100 m×150 m, no of SNs=150

Iteration No.	DKFM			I-LEACH			LEACH		
	NDF	NHD	NDL	NDF	NHD	NDL	NDF	NHD	NDL
1	3050	4346	4605	1241	1602	2414	1317	1316	2125
2	2639	4364	4597	1240	1609	2649	1285	1284	2238
3	3019	4268	4556	1291	1597	2666	1207	1206	2131
4	3150	4262	4633	1349	1593	2448	1322	1321	2388
5	3142	4299	4627	1189	1619	3259	1216	1215	2337
Average	3000	4308	4604	1262	1604	2687	1269	1268	2244

Table 10
Network lifetime on the basis of NDF, NHD, NDL for sensing area=200 m×200 m, no of SNs=150

Iteration No.	DKFM			I-LEACH			LEACH		
	NDF	NHD	NDL	NDF	NHD	NDL	NDF	NHD	NDL
1	1295	3456	4396	796	1388	3041	767	1377	2187
2	1506	3522	4371	740	1437	2240	706	1423	2002
3	1166	3444	4297	723	1401	3281	725	1394	1942
4	1496	3467	4345	827	1406	2744	836	1388	2224
5	1234	3252	4383	774	1407	2201	781	1397	2419
Average	1339	3428	4358	772	1408	2701	763	1396	2155

Table 11
Network lifetime on the basis of NDF, NHD, NDL for sensing area=200 m×250 m, no of SNs=150

Iteration No.	DKFM			I-LEACH			LEACH		
	NDF	NHD	NDL	NDF	NHD	NDL	NDF	NHD	NDL
1	707	2658	4133	322	1206	2251	315	1201	2052
2	628	2441	4153	338	1212	3605	324	1186	2087
3	638	2713	4386	264	1197	2774	248	1200	2284
4	720	2415	3948	315	1097	2368	309	1081	1773
5	553	2344	4217	390	1199	2795	378	1195	2074
Average	649	2514	4167	326	1182	2759	315	1173	2054

It has been indicated that DKFM has attained average time of NDF 141% or 1791 number of rounds, 137% or 1772 number of rounds for 100×100m area; 138% or 1738 number of rounds, 136% or 1731 number of round for 100×150 m; 73% or 567 number of round, 75 % or 576 number of round for 200× 200 m ; 99% or 323 number of round, 106% or 334 number of round for 200×250 m better than I-LEACH and LEACH respectively.

Stability period of DKFM is 178 % or 2864 number of rounds, 177% or 2858 number of rounds for 100×100 m area; 169% or 2704 number of rounds, 240% or 3040 number of round for 100×150 m; 143% or 2020 number of round, 146% or 2032 number of round for 200×200 m; 113% or 1332 number of round, 114% or 1341 number of round for 200×250 m better than I-LEACH and LEACH respectively.

Average time of NDL of DKFM is 51% or 1600 number of rounds, 97% or 2328 number of rounds for 100 m×100m area; 71% or 1917 number of rounds, 105% or 2360 number of round for 100 m×150 m; 61% or 1657 number of round, 102% or 2203 number of round for 200 m×200 m; 51% or 1408 number of round, 103% or 2113 number of round for 200 m×250 m better than I-LEACH and LEACH respectively.

Fig. 10 (a-d) and Fig. 11 (a-d) show the effect of increase in size of sensing area on number of alive nodes per round and data received by BS. Four different sensing area having size 100 m×100 m, 100 m×150m, 200 m×200 m, 200m×250m, 150 number of nodes and location of BS at (50,50) have been taken. In all situations proposed algorithm is more energy efficient, stable and scalable than I-LEACH and LEACH.

The effect of increase in node density and size of sensing area on average time of NDF, NHD and NDL of simulated protocols have been summarized and shown in Fig. 12 (a-c) and Fig. 13 (a-c) respectively. The

Proposed protocol shows considerable improvement in network lifetime than two conventional in all scenarios

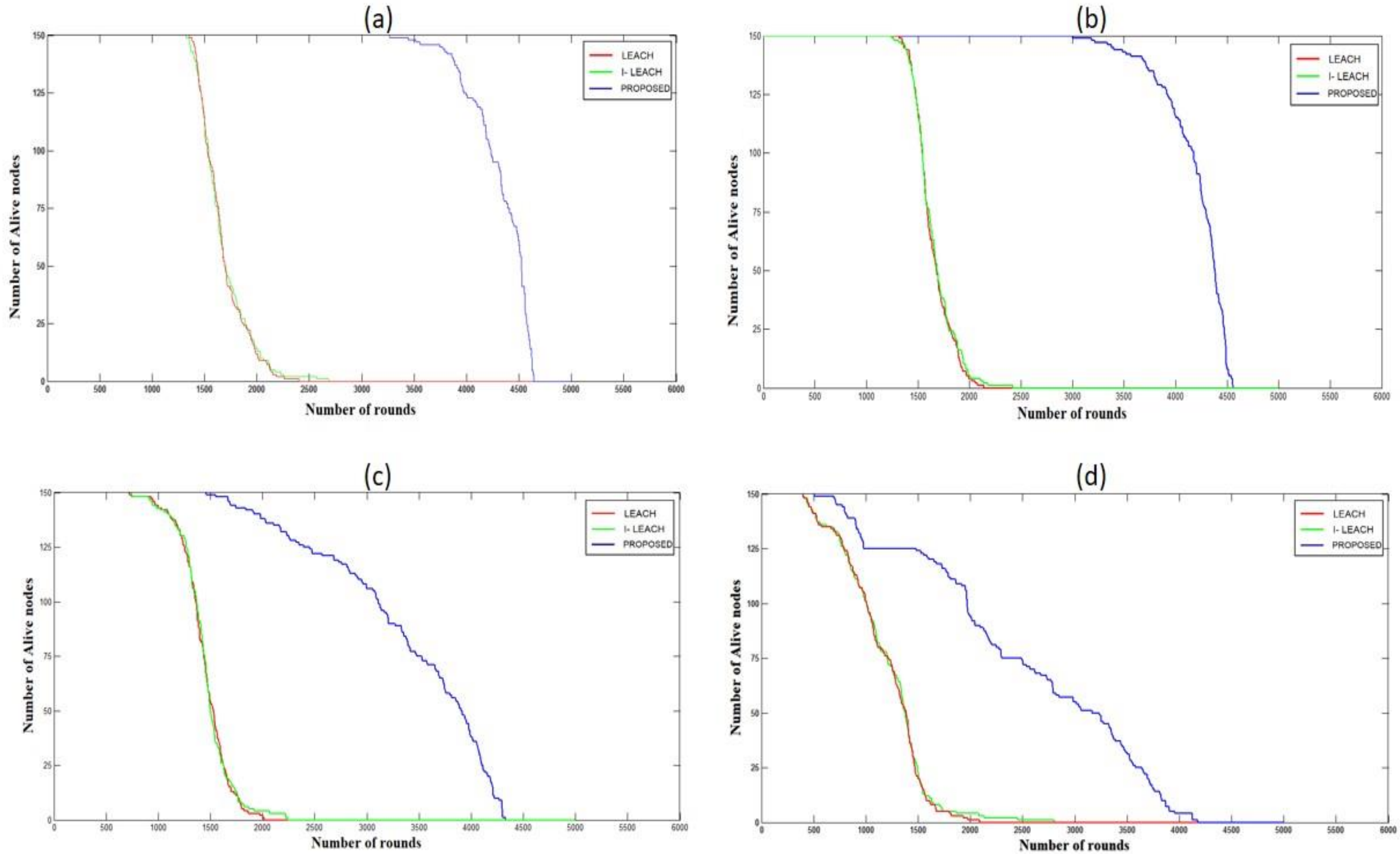

Fig. 10 (a-d). Number of Alive nodes vs Rounds for 150 SNs in sensing area of 100 m x 100 m, 100 m x 150 m, 200 m x 200 m and 200 m x 250 m respectively

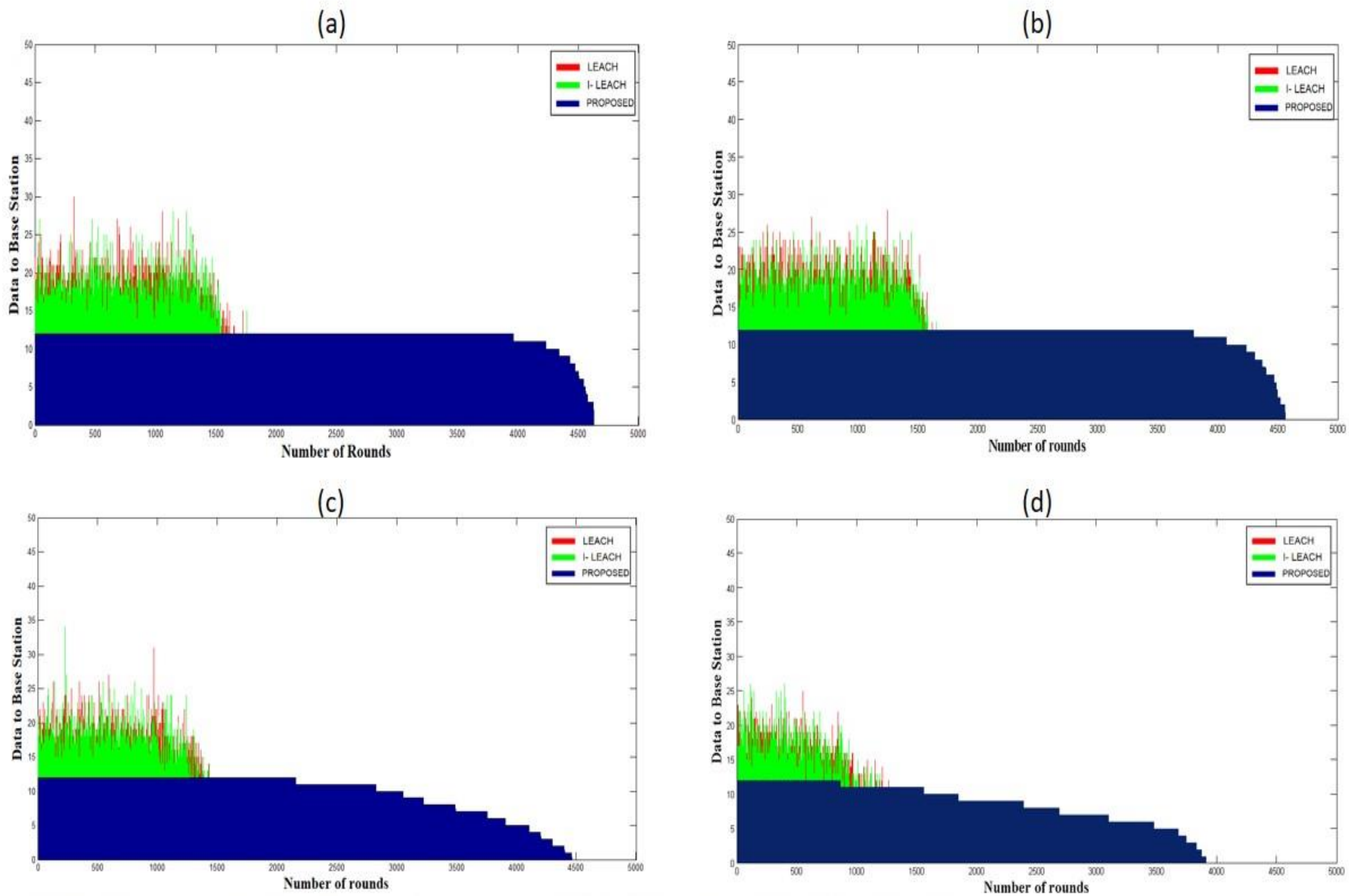

Fig. 11 (a-d). Data received by BS for 150 SNs in sensing area of 100 m x 100 m, 100 m x 150 m, 200 m x 200 m and 200 m x 250 m respectively

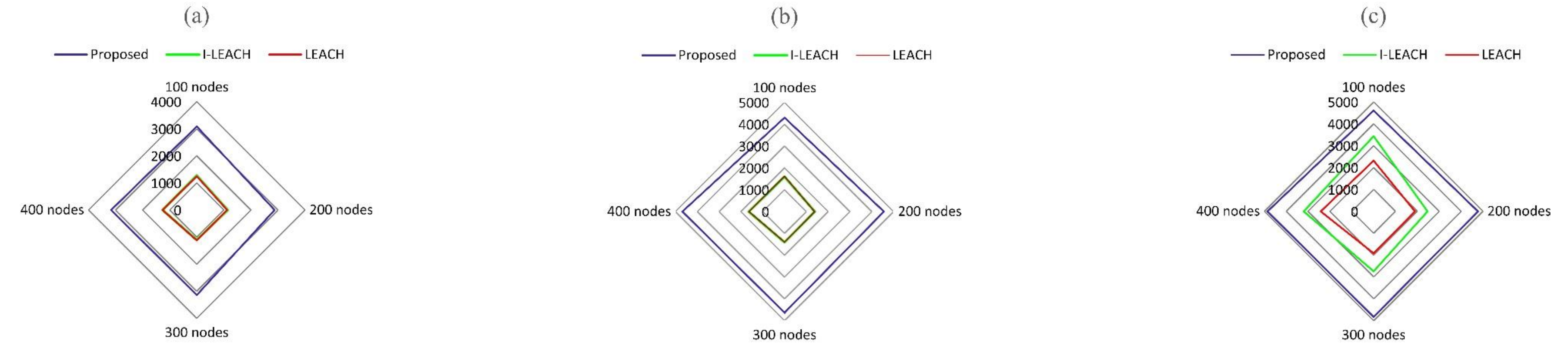

Fig. 12 (a-c). Effect of increase in node density on average time of NDF, NHD and NDL for simulated protocol

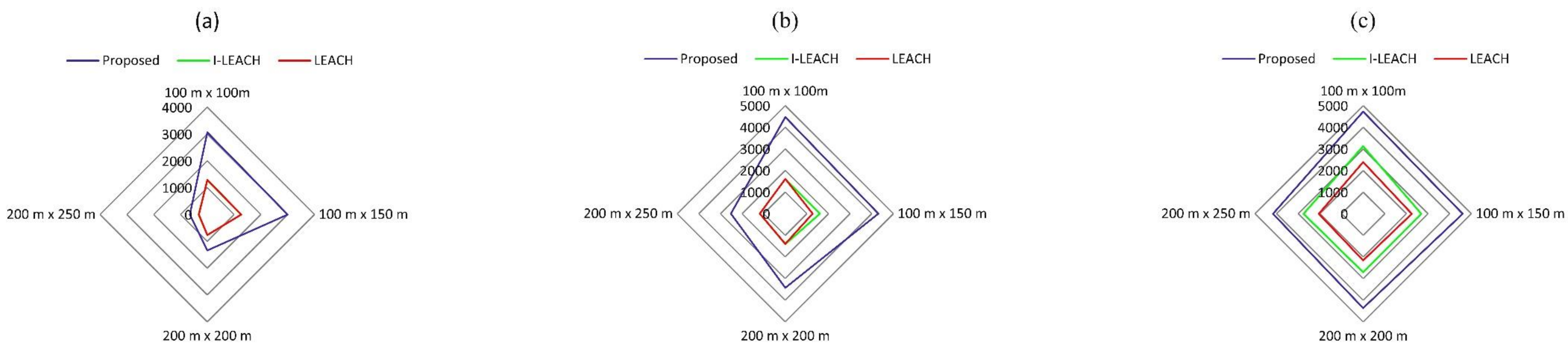

Fig. 13 (a-c). Effect of increase in size of sensing area on average time of NDF, NHD and NDL for simulated protocol

6. Conclusion

In this work A Dynamic K-means based clustering algorithm using fuzzy logic for CH selection and machine learning based data transmission (DKFM) for wireless sensor network has been proposed. It forms the optimum number of clusters using a dynamic K-means clustering such that intra cluster data transmission distance of SNs are reduced. A fuzzy inference system has been used to select suitable CH considering three fuzzy input variable such as residual energy of SN, its distance from cluster center and base station. Amount of data transmitted by member nodes to CH has been reduced by machine learning that classify similar data at regular interval. In future performance of proposed algorithm will be compared using other network simulator. Further it can be extended for heterogeneous network having mobile SNs and BS to gain more flexibility in real time applications.

Ethical approval-

This study does not contain any studies with human participants or animals performed by any of the authors

Funding details-

This study is not funded by any person or organisation

Conflict of interest-

Authors declare that they have no conflict of interest

Informed Consent

The study does not contain any identifying information or personal data of any of the individual

Author's contribution

Anupam Choudhary: Conceptualization, Methodology, Writing-reviewing and editing. **Dr.Abhishek Badholia**:Writing-original draft and formal analysis. **Dr.Anurag Sharma**: Writing-original draft. **Dr.Brijesh patel**: Supervision. **Sapna jain**:Validation

References

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: A survey," Comput. Networks, vol. 38, no. 4, pp. 393–422, 2002.

[2] N. B. Melnikova, D. Jordan, and V. V. Krzhizhanovskaya, "Experience of using FEM for real-time flood early warning systems: Monitoring and modeling Boston levee instability," J. Comput. Sci., vol. 10, pp. 13–25, 2015.

[3] F. Kong, J. Li, and Z. Lv, "Construction of intelligent traffic information recommendation system based on long short-term memory," J. Comput. Sci., vol. 26, pp. 78–86, 2018.

[4] B. Alghamdi and H. Fouchal, "A mobile wireless body area network platform," J. Comput. Sci., vol. 5, no. 4, pp. 664–674, 2014.

[5] N. Dessart and P. Hunel, "Data collection using WSN for counting individuals and habitat characterization," J. Comput. Sci., vol. 5, no. 4, pp. 624–632, 2014.

[6] J. Aponte-Luis, J. A. Gómez-Galán, F. Gómez-Bravo, M. Sánchez-Raya, J. Alcina-Espigado, and P. M. Teixido-Rovira, "An efficient wireless sensor network for industrial monitoring and control," Sensors (Switzerland), vol. 18, no. 1, 2018.

[7] A. Rajput and V. B. Kumaravelu, "Scalable and sustainable wireless sensor networks for agricultural application of Internet of things using fuzzy c-means algorithm," Sustain. Comput. Informatics Syst., vol. 22, pp. 62–74, 2019.

[8] G. R. Asha and A. Gowrishankar, "Energy efficient clustering and routing in a wireless sensor networks," Procedia Comput. Sci., vol. 134, pp. 178–185, 2018.

[9] S. Yahiaoui, M. Omar, A. Bouabdallah, E. Natalizio, and Y. Challal, "An energy efficient and QoS aware routing protocol for wireless sensor and actuator networks," AEU - Int. J. Electron. Commun., vol. 83, pp. 193–203, 2018.

[10] P. Singh, A. Khosla, A. Kumar, and M. Khosla, "3D localization of moving target nodes using single anchor node in anisotropic wireless sensor networks," AEU - Int. J. Electron. Commun., vol. 82, pp. 543–552, 2017.

[11] Huafeng Wu , Jiangfeng Xian , Xiaojun Mei , Yuanyuan Zhang , Jun Wang , Junkuo Cao , Prasant Mohapatra, Efficient target detection in maritime search and rescue wireless sensor network using data fusion. Computer Communications,no.136, pp. 53–62, 2019.

[12] Amit Sharma ,Pradeep Kumar Singh,Ashutosh Sharma, Rajiv Kumar, "An efficient architecture for the accurate detection and monitoring of an event through the sky", Computer Communications,Vol.148, pp. 115-128, 2019.

[13] H. W. Ferng, R. Tendean, and A. Kurniawan, "Energy-efficient routing protocol for wireless sensor networks with static clustering and dynamic structure," Wirel. Pers. Commun., vol. 65, no. 2, pp. 347–367, 2012.

[14] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, ", 'Energy-Efficient Communication Protocols for Wireless Microsensor Networks', Proceedings of the 33rd Hawaaian International Conference on Systems Science (HICSS), January 2000," vol. 00, no. c, pp. 1–10, 2000.

[15] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," IEEE Trans. Wirel. Commun., vol. 1, no. 4, pp. 660–670, 2002.

[16] S. Lindsey and C. S. Raghavendra, "PEGASIS: Power-efficient gathering in sensor information systems," IEEE Aerosp. Conf. Proc., vol. 3, pp. 1125–1130, 2002.

[17] X. Fan and Y. Song, "Improvement on LEACH protocol of wireless sensor network," in 2007 International Conference on Sensor Technologies and Applications, SENSORCOMM 2007, Proceedings, 2007.

[18] Z. Beiranvand, A. Patooghy, and M. Fazeli, "I-LEACH: An efficient routing algorithm to improve performance & to reduce energy consumption in wireless sensor networks," IKT 2013 - 2013 5th Conf. Inf. Knowl. Technol., pp. 13–18, 2013.

[19] M. Bani Yassein, A. Al-zou'bi, Y. Khamayseh, W. Mardini, "Improvement on LEACH Protocol of Wireless Sensor Network (VLEACH)," Int. J. Digit. Content Technol. its Appl., vol. 3, no. 2, pp. 132–136, 2009.

[20] E. Rabiaa, B. Noura, and C. Adnene, “Improvements in LEACH based on K-means and Gauss algorithms,” Procedia Comput. Sci., vol. 73, no. Awict, pp. 460–467, 2015.

[21] W. Jerbi, A. Guermazi, and H. Trabelsi, “O-LEACH of Routing Protocol for Wireless Sensor Networks,” Proc. - Comput. Graph. Imaging Vis. New Tech. Trends, CGiV 2016, pp. 399–404, 2016.

[22] J. M. Kim, S. H. Park, Y.J. Han, T. M. Chung, “CHEF: Cluster Head Election mechanism using Fuzzy logic in Wireless Sensor Networks,” ICACT, pp.654-659,2008.

[23] Z. Arabi, “HERF: A Hybrid Energy Efficient Routing using a fuzzy method in Wireless Sensor Network,” International Conference on Intelligent and Advanced Systems(ICIAS),pp. 1-6, 2010.

[24] G. Ran, H. Zhang, S. Gong, “Inproving on LEACH Protocol of Wireless Sensor Networks Using Fuzzy Logic,” Journal of Information and computational Science 7,(3), pp.767-775,2010.

[25] H. Ando, L. Barolli, A. Durresi, F. Xhafa and A.Koyama, “An Intelligent Fuzzy based Cluster Head Selection System for WSNs and its performance evaluation for D3N Parameter,” International Conference on Broadband, Wireless Computing, Communication and Applications, pp.648-653, 2010.

[26] V. V. Ghate ,Dr.V. Vijaykumar, “Machine learning for data aggregation in WSN: A survey,” Int. J. Pure Appl. Math.118(24),2018.

[27] J. Y. Chang, “A Distributed Cluster Computing Energy-Efficient Routing Scheme for Internet of Things Systems,” Wirel. Pers. Commun., vol. 82, no. 2, pp. 757–776, 2015.

[28] A. S. Toor and A. K. Jain, “Energy Aware Cluster Based Multi-hop Energy Efficient Routing Protocol using Multiple Mobile Nodes (MEACBM) in Wireless Sensor Networks,” AEU - Int. J. Electron. Commun., vol. 102, pp. 41–53, 2019.