Energy Efficient Heterogeneous WNS Clustering Using Machine Learning

Kamini Maheshwar
PhD Scholar
Computer Science & Engineering,
Rabindranath Tagore University
Bhopal, India
veenadhari1@gmail.com

Dr. S. Veenadhari
Associate Professor
Computer Science & Engineering,
Rabindranath Tagore University
Bhopal, India

Almelu
PhD Scholar
Computer Science & Engineering,
Rabindranath Tagore University
Bhopal, India

Abstract— **Heterogeneous wireless sensor network (HWSN) fulfills the requirements of researchers in the design of real-life application to resolve the issues of unattended problem. Wireless sensor networks are used in diverse areas such as battlefields, security, hospitals, universities, etc. It has been used in our everyday lives. Its development is rising day by day. Wireless sensor network includes hundreds to thousands of sensor nodes which aid in gathering various information like temperature, sound, location, etc. Recharging or modifying sensor nodes which might have limited battery power is usually difficult. Therefore, energy conservation is a crucial concern in sustaining the network. Clustering the networks is definitely one of the most common solutions for rendering WSNs energy. In this paper, review and compare different energy-efficient clustering protocols for WSNs.**

Keywords—HWSN, Clustering, Energy Efficient.

I. INTRODUCTION

Heterogeneous wireless sensor network (HWSN) consists of sensor nodes with different ability, such as different computing power and sensing range is based on the irregular sensor model used to approximate the behavior of sensor nodes. In Heterogeneous network have advanced nodes with more hardware requirement. Heterogeneous nodes have diverse hardware as well as different initial energy. The heterogeneous wireless sensor network is becoming more and more communal recently. It shows that heterogeneous nodes can prolong network lifetime without any significant increase in the cost. So, heterogeneous sensor network supports to save energy that will result in prolonging the lifetime of the WSNs. Many clustering protocols have been precisely designed for WSNs to reduce energy consumption that will extend the lifetime of the network.
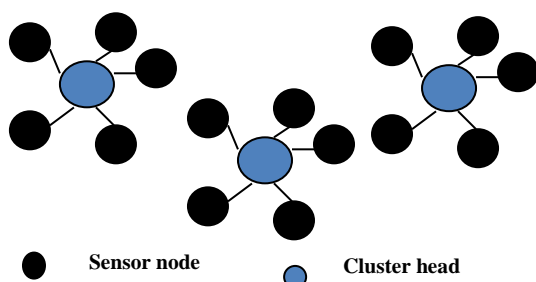

Sensor node   **Cluster head**

Fig.1. Cluster based WSN

II. CLUSTERING TECHNIQUES FOR WIRELESS SENSOR NETWORK

- Low Energy Adaptive Clustering: In this technique, it gives chances to all the nodes to become cluster head. Cluster head consumes more energy so it will make balanced energy consumption. Amount of time a single node will not run out of energy simply because it has always been selected as the cluster leader. It requires two operational phases: Setup phase and Steady phase.

- K-means algorithm: it relies mainly on two factors while picking the cluster head such as Residual energies of each node and Euclidian distances between the nodes. In this, every node sends the data to the main node, which will store in the separate file. Upon collecting knowledge from all of its nodes, it executes the clustering algorithm k-middle.
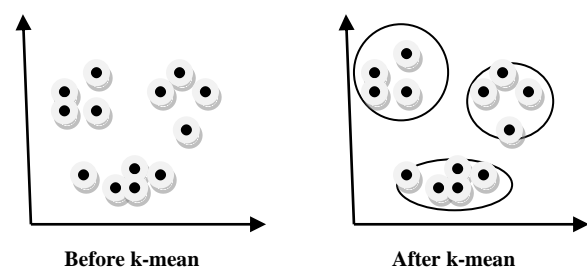

Before k-mean   **After k-mean**

Fig.2. K-mean Clustering

- Load Balanced Clustering: In the load controlled clustering method, an associate node is chosen primarily to support the head of the cluster execute its data collection and data processing function. Node must move data to the base point. The cluster head handles the data collected and moves it on to the node linked. Associate node provides the base station with this information. Multi-hop transfer of data is used to discourage associate node from an early death.

- Hierarchical Clustering Energy Efficient Algorithm: It is an extended version of LEACH's different hope

architecture. It usages the probabilistic approach for the clustering. If the node is selected as CH than all the non-CH nodes will get notification of their CH with the join invitation. Once CH is formed with their non-CH nodes, then nodes will send data to its CH and CH will transfer to Base Station. With the help of hierarchical techniques, network lifetime is improved.

- Weight-Based Clustering: It is used in heterogeneous environments. In weight dependent clustering, cluster head selection is focused on network sensor node weight. Weight is calculated by means of different parameters such as strength, node size. Each clustering node period its weight. Clustering is done with the help of weight, so the usages of resources in WSN become less. This selects the best heads of the network and thereby increases lifetime with the performance of WSN.

- Fuzzy based Clustering: In this technique cluster head selection is done with the support of different parameters including capacity, size and node density. All of these parameters are set as the Fuzzy Inference System, which provides the input for collection of cluster head. It also helps to achieve the load balancing in the terms to maintain the stable connectivity for a particular duration. It expanded the WSN's life span and transmitted more and more critical data to sink [1].

III. CLUSTERING AND ROUTING OPTIMIZATION ALGORITHM FOR HETEROGENEOUS WIRELESS SENSOR NETWORKS

The clustering routing algorithm of WSN is divided into initialization stage and stabilization stage. Initialization stage is an important stage of cluster head stage and cluster formation. Once the cluster head and cluster are formed, the entire system will enter the stable working phase, which is the phase where the cluster head nodes transmit data. The actual working principle is similar to LEACH .The data has a clustering algorithm that can cluster common things into one category through analysis and calculation. It can be better applied in WSN cluster head calculation, cluster calculation and data transmission [2]

IV. A NOVEL CLUSTERING PROTOCOL IN WIRELESS SENSOR NETWORK

The protocols used in the WSN maximize the life of the system. The CH election procedure is totally energy dependent. This technique uses the node energy for the assortment of CH, which makes the network energy efficient. The CH selection procedure starts with the generation of random number. The random value is then matched with the threshold function. If that number comes smaller then threshold value then that node become CH for that round. The system performance is dependent on the threshold function used for CH selection. An efficient threshold can enhance the network life by selecting the CH as per the energy level of the system. In this technique, the threshold function uses the initial energy and aggregate energy of the network for the CH selection. It is based on the heterogeneous environment in which the nodes in the network have the different energies. The nodes are categorized on the basis of their energy levels. The nodes are categorized as simple nodes and super nodes. The super nodes are provided with some extra energy than the simple nodes. The threshold for both the nodes is different as per their energy levels [3].

V. MULTI-OBJECTIVE HETEROGENEOUS CLUSTERING APPROACH FOR EFFICIENT-ENERGY OPTIMIZATION IN WSN

Multi-Objective Heterogeneous Clustering (MOHC): Technique is used to form cluster and selection of cluster head. This technique is divided into three stages: cluster head identification phase, cluster formation phase, and re-cluster phase. The selection of cluster head using four parameters.

Delay: Delay is time used in sensing the environment and send this information from nodes to sink. Delay is Inversely related to residual energy more residual energy of the sensor node triggers more numbers of data packets.

$$D = ((E_{initial} - E_{residual})/E_{initial} + r)) * RTD$$

Where,
D=delay,
RTD=round trip time,
r=random value between 0 &1.
Energy Consumption of CH: Energy consumption shows the total energy required by the node as a CH to receive and transfer the data to the BS. Energy consumption of CH (Econ) in each round.

$$Power\ CH = E_{rec} + ET\ (CH_i - BS)$$

Where,

E_{rec} = Energy required to receive the data from other nodes in the cluster.

ET $(CH_i - BS)$=Transmission energy to transfer data from CH_i to BS.

Coverage: Rate of energy dissipation of node is directly proportional to d^2 when $d < d0$. If distance between two node is even slightly greater than d0, drastically energy reduction occurs.

Connectivity: Connectivity shows that the distance between CH and BS is less than d0. It confirms less energy consumption during data transmission. Its value will be 1 if distance is less than d0 otherwise value will be 0.

Connectivity = 1, if $distance_{node} - BS < d0$ or 0, if $distance_{node} - BS \geq d0$

Re-cluster: clusters are reorganized after clusters head identification process,. All CHs broadcast the message of being a CH and all normal nodes select the nearest CH based on the signal strength using join message. This phase again reduces the energy consumption and guarantees the lower distance between nodes and respective CH [4].

VI. POWER OPTIMIZATION ALGORITHM FOR HETEROGENEOUS WSN USING MULTIPLE ATTRIBUTES

It enhance the lifetime of WSN with lowering the energy consumption. The selection of CH as we select the CHs

on the basis of five conflicting factors: delay, CH coverage, CH lifetime, average distance to CH and maximum power of nodes. After selection process we have obtained the clusters again after every round of CH selection using minimum distance to selected CHs. The distance is measured on the basis of received signal strength. It has three phases.

Cluster Formation: All Nodes having different energy and characteristics due to heterogeneous network. clusters are made on probabilistic approach. In each cluster CH selection approach is applied to select the best CH in each cluster.

CH Identification: The attributes and the process to calculate the values are as follows:

Delay: Delay is inversely proportional to residual energy. Its value is calculated as:

$$D = \left(\frac{(Einitial - Eresidual)}{Einitial + r}\right) * RTD$$

Where D=delay, RTD=round trip time ,r=random value between 0 &1.

Cluster Head Coverage of Nodes: It shows the number of normal nodes in cluster which will have a distance less than d0 to particular node or optional CH. Its value is calculated as:

$$ClusHdCov = \frac{(Tota\_Nodes(Minimum(Distance(Nodei - CHs)) \leq d0))}{node\_cluster}$$

Cluster Head Lifetime: It shows the lifetime of the node if it will be a CH. It is calculated as:

$$ClustHd\_Life = \frac{Eres\_CHi}{Average\_Transmission\_Energy\_Required}$$

Where, Eres=Residual energy

Average Distance to CH: Average distance from member nodes to the node if it will be a CH.

$$Average\_CHs = \frac{\sum_{i=1}^{node\_cluster} Distance(nodei-CH)}{node\_cluster}$$

Maximum Power of Nodes: the maximum power required to transmit data to CH among the nodes in cluster.

Power Max = Max(Transmission Power(Node$_i$ -CH))

Re-cluster:: When CHs are selected then all clusters are again formed on the basis of distance from member nodes to the selected CHs. This distance is majored on the basis ofintensity of signal received from the CHs. After completion of this round, all member nodes attached with the nearest CH[5].

VII. ENERGY BASED CLUSTERING HETEROGENEOUS WIRELESS SENSOR NETWORK: A STUDY

Base station serves as a gateway to send data to another network Receiving data from cluster head involves high cost, thus when the distance between sensor networks is very large, and then only we go for cluster heads.(Stable Election Protocol) SEP have two types of nodes: initial node and advance node, advance nodes have double initial energy than normal nodes. Figure shows the communication between the cluster and their

corresponding CH and base station. Here, the arrangement of the nodes is assumed to be random in nature and distance between each subsequent area circle is denoted by d0 [6].

VIII. ENERGY EFFICIENT CLUSTERING IN HETEROGENEOUS ENVIRONMENT

Stable Election Protocol (SEP) works in the heterogeneous environment. It applies weighted election probabilities for every node to elect as CH on the basis of their energy level. The advanced nodes in the system are provided with additional energy in comparison to normal nodes. The sensor nodes are divided into the three-level hierarchy based on their energy levels. Sensor nodes are organized as simple nodes, middle nodes, and super nodes. The sensor nodes use their energy levels as criteria to elect the cluster head among them. The main focus is on reducing the cost of communication and lessening the amount of data transfer to the sink. Each node transfers the data to its nearest cluster head and then after aggregating the data cluster head forward that data to the sink. Each node gets the chance of becoming cluster head in every 1/p round exactly once with a probability of p.

The energy of normal nodes Enrm = Eo

The energy of middle nodes Emid = Eo (1+β)

The energy super nodes Esup =Eo (1+γ),Where β=γ/2

With the change in the energy levels of the system the probabilities of the nodes of becoming cluster head for different nodes also gets changed by:

$$Pnrm = \frac{P}{(1+a*\gamma+i*\beta)}$$

$$Pmid = \frac{P*(1+\beta)}{(1+a*\gamma+i*\beta)}$$

$$Psup = \frac{P*(1+\gamma)}{(1+a*\gamma+i*\beta)}$$

This technique multiplies the ratio of node's present energy to maximum energy of all nodes with the threshold function T(n).

$$T(n) = T(n) * \frac{present\ energy\ of\ node}{initial\ energy}$$

If the value of a random number is smaller as compared to the threshold, the node gets selected as a CH. The threshold function for normal, middle and super nodes :

$$T(n_{nrm}) = \begin{cases} \frac{Pnrm}{1-Pnrm[(r\ mod\frac{1}{Pnrm})]} * \frac{Eremaining}{Einitial}, & \text{if } n\ \varepsilon\ G" \end{cases}$$

0 otherwise. The total count of normal nodes is n*(1-a-i). G' represents the set of normal nodes who do not get chance of becoming CH from the past 1/Pnrm rounds, and r represents current round.

$$T(n_{mid}) = \begin{cases} \frac{Pmid}{1-Pmid[(r\ mod\frac{1}{Pmid})]} * \frac{Eremaining}{Einitial}, & \text{if } n\ \varepsilon\ G" \end{cases}$$

The total number of middle nodes are n* b. G'' represents the group of middle nodes who do not get chance of becoming CH from the past 1/Pmid rounds, and r represents current round.

$$T(n_{sup}) = \begin{cases} \dfrac{Psup}{1-Psup[(r \bmod \frac{1}{Psup})]} * \dfrac{E remaining}{E initial}, & \text{if } n \, \varepsilon \, G" \end{cases}$$

The total number of super nodes are n*m. G'' represents the set of super nodes. [7]

IX. Machine Learning in Wireless Sensor Networks:Challenges and Opportunities

Machine learning algorithms for learning models development is majorly classified into three types supervised learning algorithms, unsupervised learning algorithm and reinforcement learning.

Supervised Learning:These algorithms require some external assistance for performing their task. The input is made up of training data set and test data set. It is based on the patterns learned from the training data set and when applied to the test data set.

i. Support vector machines(SVMs)
ii. Bayesian statistics
iii. Decision tree (DT)
iv. Neural networks (NNs)
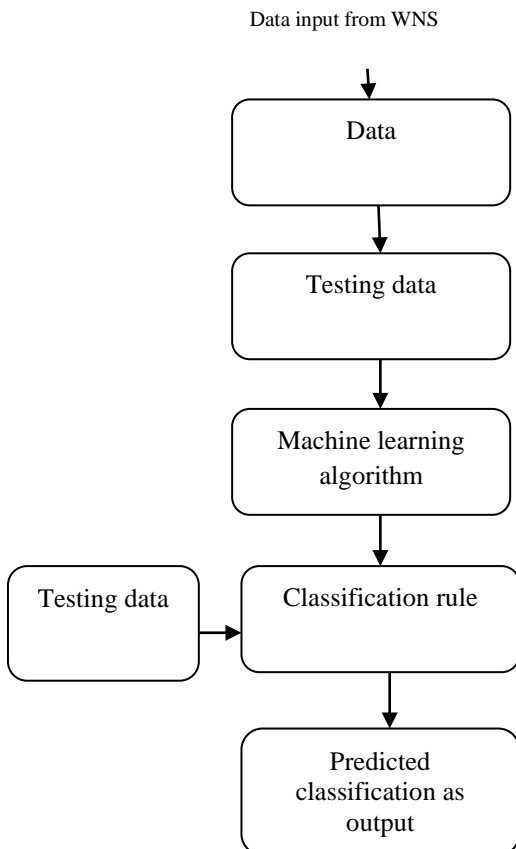v. K-nearest neighbor (k-NN)

Data input from WNS

Fig 4: Use of machine learning in WNS

Support vector machines: Support vector machine (SVMs) is a technique under supervised machine learning in which models are developed for analysis of data required for classification and regression analysis. This technique is also known as support vector networks.

Bayesian statistics: This learning technique uses mathematical procedures while applying probability methods for dealing with problems which are statistical in nature.

Decision tree: This modeling approach based on prediction and is widely used in data mining, statistics and machine learning.

Neural networks: It depends upon data and the network is built with the help of cascading chains of decision units which is often termed as perception and radial basis function.

K-nearest neighbor: it is used for query processing jobs in wireless sensor networks. In this algorithm, the data sample for testing is classified based on the labels of the nearest data samples.

Unsupervised Learning: It is associated with the idea of using a collection of observations collected or grouped from a distribution. there is no output vector and the learning that takes place under this technique does not require any labels for the input data.. When new This learning technique is largely used for data clustering and truncation of features. There are two types algorithms under this classification of learning are:-

K-means clustering: This clustering algorithm is very popular in wireless sensor networks because of its simplicity in execution .This machine learning technique categorizes data into different classes or clusters and works sequentially by randomly selecting k nodes. These selected k nodes acts as initial centroids. This procedure is repeated again and again to re-compute the centroids using a threshold value which is predefined for the node andthis iterative procedure will end only when convergence condition is met. The algorithm when used help in creating groups automatically based on the similarities in the characteristics.

Principal component analysis: This machine learning algorithm is used for data compression and for reduction in dimensions.

Reinforcement Learning: Reinforcement learning actuates the sensor node in a wireless sensor network to perceive process and communicate data with its environment. This machine learning strategy purely depends upon two measures: trial and error search and delayed outcome

Q-learning technique: This learning algorithm is most agreed upon and broadly used machine learning algorithm for addressing issues related to routing in wireless sensor networks [8].

X. Related Work

A. Enhanced EECP: Enhanced Energy Efficient Coverage Preserving Protocol for Heterogeneous Wireless Sensor Networks

The Enhanced EECP protocol have three different types of heterogeneous sensor nodes i.e. Class-A, Class-B and Class-C type nodes. It have different battery capacity, sensing and transmitting range. In the network, initial energy of Class-A node < Class-B node < Class-C node. To transmit sensed data to sink, multi-hop hybrid routing

is used where Class-A type nodes (deployed in inside region) transmit their data packets by following multi-hop direct communication towards the sink while Class-B (deployed in outside regions) and Class-C type nodes (deployed in middle regions) transmit their data packets by following multi-hop clustering architecture. In Enhanced EECP protocol, coverage efficient CHs are initially selected by sink using fuzzy logic technique [9].

B. Real-Time Implementation of Energy-Efficient Clustering Protocol in WSN

This technique works into two phases: Network setup phase and Data transmission phase.

Cluster Setup phase: In the cluster setup phase, all the geographical positions of the deployed sensor nodes are properly mapped on Cartesian plane. Clustering algorithm is applied to generate control packet. Control packet holds the ids of opted CHs and then broadcasted by base station in its communication range. On the receipt of the control packet, nodes will immediately switch into network setup mode and decide their mode as CM or CH by looking their ids into control packet. If the node is elected as CH then it will wait a while and rebroadcast the same control packet in its communication range to ensure the receipt of the control packet by every node. Then nodes will adjust their sleeping time according to modes.

Data transmission phase: In the data transmission phase, members start collecting the environmental parameters such as temperature and humidity along with the available battery power. This information then forwarded to the CH on every random time delay of 10 to 15 minutes. Nodes are programmed to send energy information only on demand by base station. Whereas, CHs will forward their energy information to the base station on after every 3 minutes to ensure that they are alive. The cluster formation procedure will not repeat until either the power level any of the opted CH is below to the threshold or any one of the CH is dead [10].

C. Energy Efficient Clustering Protocol for Heterogeneous Wireless Sensor Network: A Hybrid Approach using GA and K-means

GA is an optimization technique based on the principles of evolution and natural genetics and randomized search. It performs search operation in very complex, large and multimodal landscape in order to provide a near optimal solution to a problem. Each solution is a chromosome. This process (of selection, crossover and mutation) is continued for fixed number of iterations or till a termination condition is fulfilled. GA and K means to solve clustering problem of WSN. Most of the routing protocols based on GA choose random population but these may not be efficient for sparsely deployed sensor networks because it may lead to the selection of CHs with low density. This may affect the final clustering result. A solution to this limitation is selection of good quality CHs in the population. A combination of meta-heuristic approach for supporting high exploration and deterministic approach for selection of good quality CHs can provide better clustering solution for HWSNs. In the proposed method, the population of GA is seeded with K-means in order to get good clustering result [11].

D. Distributed Energy Efficient Clustering Protocol

In DEEC protocol, cluster heads are selected by a probability based on the ratio between residual energy of each node and the average energy of the network. The approach of being cluster heads for nodes are different according to their initial and residual energy. DEEC protocol estimates the ideal value of network life-time, which is used for calculating the reference energy that each node should expend during the round.

E. Developed Distributed Energy-Efficient Clustering for Heterogeneous Wireless Sensor Networks (DDEEC)

In DEEC, where all nodes use the initial and residual energy level to define the cluster heads. Nodes average probability Pi to be a cluster head will have changed as in this protocol, each node needs to have the global knowledge of the networks, DDEEC like DEEC estimate the ideal value of network lifetime, which is used to compute the reference energy that each node should expend during each round. In this scheme, the network is organized into a clustering hierarchy, and the cluster heads collect measurements information from cluster nodes and transmit the aggregated data to the base station directly.

F. Enhanced Distributed Energy Efficient Clustering Protocol for Heterogeneous Wireless Sensor Networks (EDEEC)

Three types of nodes in prolonging the lifetime and stability of the network. Nodes average probability Pi to be a cluster head will have changed as Hence, it increases the heterogeneity and energy level of the network.

G. Balanced Energy Efficient Network Integrated Super Heterogeneous Protocol (BEENISH)

Cluster Heads (CHs) are selected on the basis of the residual energy level of nodes. Nodes average probability Pi to be a cluster head will have changed as Simulation results show that it performs better than existing clustering protocols in heterogeneous WSNs. This protocol achieves longer stability, lifetime and more effective messages than Distributed Energy Efficient Clustering (DEEC), Developed DEEC (DDEEC) and Enhanced DEEC (EDEEC).

H. Modified Balanced Energy Efficient Network Integrated Super Heterogeneous Protocol (m-BEENISH)

BEENISH protocol provides four classifications to the nodes based on energy levels. The more the classification of energy levels and probability constraints the better are the outcomes and efficiency parameter. m-BEENISH is vitally an improvement in heterogeneous WSNs, in which new mechanism is adapted to provide feasibility to the existing BEENISH protocol[12].

Table I: Gives the Study of hetergeneous wireless sensor network clustering

Author	Description	Results
Ling Chen et al.(2020)	Clustering and routing optimization in WNS	Reduce the consumption of network energy castoff
Jayant Kumar Rout(2019)	Clustering Protocol in Wireless Sensor Network	energy efficient and enhanced the system performance
Abhishek Prajapati et al(2019)	Multi-Objective Heterogeneous Clustering (MOHC) Technique	Optimization of the energy consumption
Kumkum Dubey et al(2019)	Power Optimization Algorithm for Heterogeneous WSN	Reduction of communication cost between member node and respected cluster head
Sushant Kumar et al(2018)	Energy based Clustering Heterogeneous Wireless Sensor Network	Increases optimization for cluster head network
Piyush Rawat et al(2018)	Energy Efficient Clustering in Heterogeneous WNS	The network lifetime improved
Deep Kumar Bangotra et al(2018)	Machine Learning in Wireless Sensor Networks	Machine learning provides solution that have problems related to networking and data processing in WSNs
Sonam Maurya et al(2018)	Enhanced EECP for Heterogeneous WNS	Enhanced EECP protocol prolongs the overall network lifetime and also achieves approximately 2:5 times better improvement in coverage lifetime
Shashi Bhushan et al(2018)	A Hybrid Approach using GA and K-means	network life time, throughput 10% and 20% increased

XI. CONCLUSION

In this paper, we studied various protocols that have been used in wireless sensor network. How to use different technique for reducing energy consumption in creating network and cluster head. Hybrid approach combining GA and K-means is used to do clustering of WSN. When the distance between the sensors is less, the direct communication between sensor networks and bus station

is in a stable form. As it increases, one should opt for CH network. The fuzzy logic approach is implemented, which is capable of taking real time decisions.

REFERENCES

[1] Santosh Anand, Manoj K.C, " A Survey on Clustering Approaches to Strengthen the Performance of Wireless Sensor Network," in IEEE xplore, September 2020.

[2] Ling Chen, Wenwen Liu, Daofu Gong, Yan Chen, "Clustering and Routing Optimization Algorithm for Heterogeneous Wireless Sensor Networks," in IEEE, 2020.

[3] Jayant Kumar Rout," A Novel Clustering Protocol in Wireless Sensor Network, "in IEEE, 2019.

[4] Abhishek Prajapati, Vinay Kumar, Prince Rajpoot, Kumkum Dubey, Shubham Harsh Singh, Neetu Verma," Multi-Objective Heterogeneous Clustering Approach for Efficient-Energy Optimization in WSN,"in IEEE,2019.

[5] Kumkum Dubey, Adarsh Kumar Yadav, Pappu Kumar, Prashant Shekhar, Prince Rajpoot, Shobhit Kumar," Power Optimization Algorithm for Heterogeneous WSN using Multiple Attributes,"IEEE,2019.

[6] Sushant Kumar, Aman Gupta, Kumar Shaurya, Shashank Kumar, Amrit Mukherjee," Energy based Clustering Heterogeneous Wireless Sensor Network: A Study,"IEEE, 2018.

[7] Piyush Rawat, Siddhartha Chauhan," Energy Efficient Clustering in Heterogeneous Environment,"IEEE Xplore, 2018.

[8] Deep Kumar Bangotra , Yashwant Singh ,Arvind Selwal," Machine Learning in Wireless Sensor Networks:Challenges and Opportunities,"IEEE, December 2018.

[9] Sonam Maurya, Vinod Kumar Jain," Enhanced EECP: Enhanced Energy Efficient Coverage Preserving Protocol for Heterogeneous Wireless Sensor Networks,"IEEE,2018.

[10] Himanshu Ranjan Seth,, Chiranjeev Kumar,, Haider Banka ," Real-Time Implementation of Energy-Efficient Clustering Protocol in WSN"IEEE,2018.

[11] Shashi Bhushan, Raju Pal, Svetlana G Antoshchuk," Energy Efficient Clustering Protocol for Heterogeneous Wireless Sensor Network: A Hybrid Approach using GA and Kmeans,"IEEE,august 2018.

[12] Anurag Chatap, Sumedha Sirsikar," Review On Various Routing Protocols For Heterogeneous Wireless SensorNetwork,"IEEE, 2017.