Journal of Intelligent Systems and Internet of Things
Vol. 13, No. 02, PP. 52-59, 2024
DOI: https://doi.org/10.54216/JISIoT.130204

Machine Learning and Internet of Things Driven Energy Optimization in Wireless Sensor Networks through Crossbreed Clustering

Ahmed Saeed Alabed[1], Rajesh Kumar Samala[2], Asha KS[3], Sorabh Sharma[4], Amit barve[5], Deepak Minhas[6]

[1] Assistant Professor, Information Systems and Computer Science Department, Ahmed Bin Mohammed Military College, Qatar
[2] Professor, Department of uGDX, ATLAS SkillTech University, Mumbai, Maharastra, India
[3] Associate Professor, Department of Electronics and Communication Engineering, Faculty of Engineering and Technology, JAIN (Deemed-to-be University), Bangalore, Karnataka, India
[4] Centre of Research Impact and Outcome, Chitkara University, Rajpura- 140417, Punjab, India
[5] Associate Professor, Department of Computer Science and Engineering, Parul University, Vadodara, Gujarat, India
[6] Chitkara Centre for Research and Development, Chitkara University, Himachal Pradesh-174103 India
Emails: ahmed.alabed@abmmc.edu.qa; rajesh.samala@atlasuniversity.edu.in; ks.asha@jainuniversity.ac.in; sorabh.sharma.orp@chitkara.edu.in; damit.barve17535@paruluniversity.ac.in; deepak.minhas.orp@chitkara.edu.in

Abstract

Key challenges in Wireless Sensor Networks (WSNs) include reduced dormancy, energy efficacy, reportage worries, and network lifetime. To solve the issues of energy efficiency and network longevity, more study of cluster-based WSNs is required. In order to address the challenges and constraints of WSNs, creative approaches are needed. WSNs use machine-learning techniques because of their unique characteristics. These characteristics include high communication costs, low energy reserves, high mobility, and frequent topological shifts. The current method picks cluster heads at random at the beginning of each cycle, not considering the remaining energy of these nodes. It is possible that the newly chosen CH nodes will have the lowest energy level in the network and will die off fast as a result. Energy is wasted while communicating over long distances between cluster heads and the BS, which occurs frequently in a big network due to Internet of things. This would mean that WSNs have a finite lifespan. Therefore, to increase the network's longevity and efficiency, we propose a machine-learning-based strategy called energy proficient crossbreed clustering methodology (ECCM). The experimental results reveal that the ECCM is superior to the LEACH approach, increasing residual energy by 35%, extending network lifetime by 37%, and increasing throughput by 15%.

Received: September 04, 2023 Revised: February 03, 2024 Accepted: June 20, 2024

Keywords: ECCM; WSN; BS; LEACH; IoT.

1. Introduction

The original uses for WSNs in the military and heavy industry were considerably different from the current uses in the consumer and light industrial sectors. These minuscule sensor nodes comprise sensing, Components for data processing and communication that make use of sensor networks' distributed computing paradigm [1]. When compared to traditional sensors, sensor networks represent a significant advancement. To form a sensor network, a vast number of Sensor Nodes (SNs) are scattered within or around the phenomena. The placement of sensors is not something that needs to be engineered or planned [2]. This enables the haphazard placement of sensor nodes. The cooperative nature of the sensor nodes is another characteristic that distinguishes sensor networks. As opposed to sending the fusion nodes the raw data, Sensor nodes can send just the necessary and partially processed data by using their computing capability to perform simple computations locally. Sensor network applications require wireless ad hoc networking techniques [3]. For traditional wireless ad hoc networks, many protocols and algorithms have been suggested yet, conventional methods are inadequate for the special needs and functional specifications of sensor networks [4]. To support this assertion, we make the following comparisons and contrasts between sensor networks and ad hoc networks:

- Compared to an ad hoc network, a sensor network may have several times as many nodes.
- The sensor nodes are scattered over different areas.
- It is not uncommon for sensor nodes to fail.
- A sensor network's topology is characterized by constant development.

As opposed to the point-to-point connectivity that the great majority of ad hoc networks employ, Often, broadcast communication is used by sensor nodes [5].

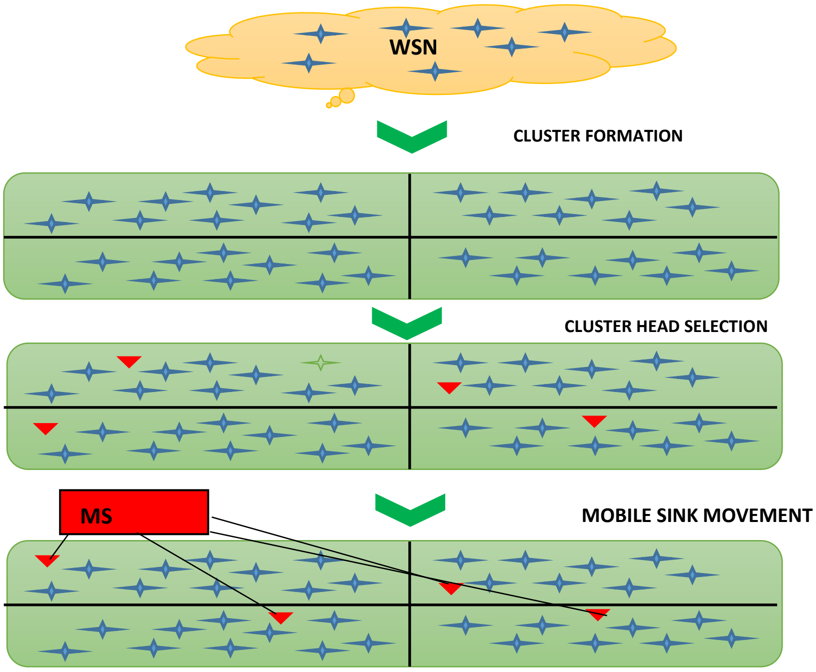

Figure 1: Clustering in WSN.

The node's most important hardware resides in the sensing unit, which features application-specific sensors like temperature, pressure, humidity, and sound [6]. Adjacent nodes may be used to relay the sensing data to the sink node. In a temperature-monitoring setup, a thermistor will serve as the sensor. The sensing apparatus required will be unique to each use case. A microcontroller, or CPU, is a computer's brain. The processing unit coordinates the activities of the entire node by issuing commands to the appropriate devices [7].

A transceiver is an electronic device that transmits and receives signals utilising the same circuit or container.

The sensor node receives electricity from the power unit. A battery, often one that can be recharged, powers this gadget [8]. Improvements in power units make solar charging feasible. Long-term power backup and savvy power awareness can automate charging from both solar panels and standard outlets. As a result, sensor networks become more viable in several contexts.

This classification establishes whether the WSN employs similar sensor nodes or sensor nodes with distinct functions [9]. Computing is easy if all of the sensor nodes in the network are of the same kind, if the sensor nodes are of different types, computing is quite hard to get all of the sensor data for processing. In a homogenous sensor network, the base station and the sensor nodes have the same features, including processing speed and data storage capacity. In this case, data gathering is determined by the structure of data sharing [10]. Data gathering and distribution in homogenous networks have been studied extensively using both flat and hierarchical designs.

By using machine learning in the process of analysing sensor data, inter-network communication can be significantly enhanced. In sensor networks, machine learning techniques are useful instruments for information extraction and correlation detection [11], especially in circumstances when numerical models are either inaccessible or extremely costly. ML approaches make it possible to choose the cluster head node, the identification of the best course of action the extraction of useful information from the data set, decrease in packet delivery time and the increase in WSN lifespan.

Since optimizing the use of available energy at each sensor node maximizes the lifetime of the network, designing wireless sensor networks (WSNs) with one of its main goals, an increasing number of scholars are becoming interested in the topic of energy-aware communication in WSNs. Several algorithms based on machine learning have been used to track effective energy harvesting methods for WSNs [12]. There are several benefits of using machine learning in the energy-gathering process. A machine-learning technique that calculates the energy that can be collected in a specific amount of time will significantly improve WSN display. It can be challenging to achieve high energy efficiency in WSNs. Traditional WSN procedures pose a challenge to the ability of systems to progressively adapt because of the tremendous uniqueness of these approaches. The utilisation of methods of machine learning makes it conceivable to provide an acceptable response in circumstances like these [13].

2. The Existing Work Done

Due to limited resources, such as high energy consumption and early SN expiry, WSN cannot allow sensor nodes to interact directly with BS or to communicate with BS via numerous hops. A direct connection has various downsides, such as excessive energy consumption, data duplication (near sensing nodes delivering data with very tiny differences), and rapid demise of the farthest nodes [14]. Some have suggested a two-tiered system as a way to handle the problems at hand. A hierarchical approach is used in conversations. Groups of nodes begin to coalesce.

Most hierarchical network structures have two levels, with the cluster leaders at the top and the member nodes below them. Regular transmissions of data from lower-level nodes to the appropriate CH. Subsequently, the data is collected by the cluster head and transmitted to the base station [15]. The CH node consumes more power than the rest of the network since it is always sending data over very long distances. When a set number of rounds have been played, the selected CH may become incapacitated or even perish if its energy reserves have been depleted. To maintain energy efficiency and fair distribution of processing among sensor nodes, CH's duties are constantly adjusted [16]. Multi-hop (inter-cluster) communication takes place between clusters, whereas single-hop (intra-cluster) communication takes place within a cluster. Cluster-based and grid-based systems are the most common hierarchical techniques used today.

It has been demonstrated in the literature that hierarchical clustering procedures can be developed using many different strategies [17]. The protocols are developed with fundamental goals in mind, such as minimising network downtime and maximising energy efficiency. Clustering techniques are used to improve load balancing, robustness, and data aggregation, as well as to streamline node administration and reduce energy consumption. Formation of clusters [18]. A CH node's job is to gather data from other nodes, compile it, and send it either directly or indirectly to the base station. Before being made available to end users, the input from a cluster head node undergoes additional evaluation at the base station. BS could be situated anywhere inside or beyond the range of the network [19]. Away from the sensor nodes, BS is typically deployed in the great outdoors. Challenges like efficient cluster generation, equitable load allocation, cluster head choice, and cluster reformation have been addressed in many various ways in the literature to improve energy efficiency via clustering. Quite a few of them are discussed below [20].

Energy-efficient routing protocols are still being developed in WSN, but the authors have introduced a new one: The hierarchy of low-energy adaptive clustering. By rotating a big number of nodes, Low Energy Adaptive Clustering Hierarchy (LEACH) sought to randomly pick a cluster hierarchy (CH), allowing energy expended through contact to be spread among all of the network's nodes [21-23]. There are two stages to each given operation: the initial setup and the final steady state.

Increased network presentation and energy efficiency based on weighted criteria. The leader is selected using these criteria. Through a direct link, the data is sent from the member nodes to the relevant CHs [24]. Unless the message is routed through the BS, messages between clusters are sent along a chain of selected cluster heads (CHs). The re-election of CH, however, causes an increase in network operating expenses because of the incorrect CH election. Furthermore, direct intra-cluster communication leads to unequal energy use [25].

3. The Objective of the Research Work

The WSN necessitates novel approaches to the existing network challenges and constraints. Machine learning algorithms give a toolkit for increasing a network's flexibility in the face of novel circumstances. Further research is necessary to address the problems of network longevity and energy efficiency in cluster-based WSNs.

4. The Proposed Work

Environments that undergo rapid change are tracked in real time via WSNs. As a result, sensor networks frequently use machine learning to adjust to new environments. Machine learning in WSNs applies statistical and mathematical methods to provide technological answers to the challenge of improving WSNs' performance. Extending the network's operational life is a major difficulty in the field of WSNs. The sensor nodes' energy consumption will determine how long they last. The transmission of data consumes the vast majority of the power available at the sensor nodes. The most common methods for managing energy consumption effectively include clustering, cluster head, and multi-hop communication.

Figure 2 depicts the conceptual framework of the proposed ECCM. Each node in a WSN belongs to a cluster, and the cluster head is in charge of collecting data from the nodes and sending it to the base station. To provide the necessary level of coverage, sensors are often deployed in dense clusters, allowing certain nodes to enter a sleep state and conserve significant amounts of energy. The distance, energy level, and sensor node degree are used to determine which nodes will serve as cluster leaders. How long a WSN lasts depends greatly on its cluster head. The best cluster leader is the one that has a high available energy reserve, a large number of nearby nodes, and a short path to the hub.

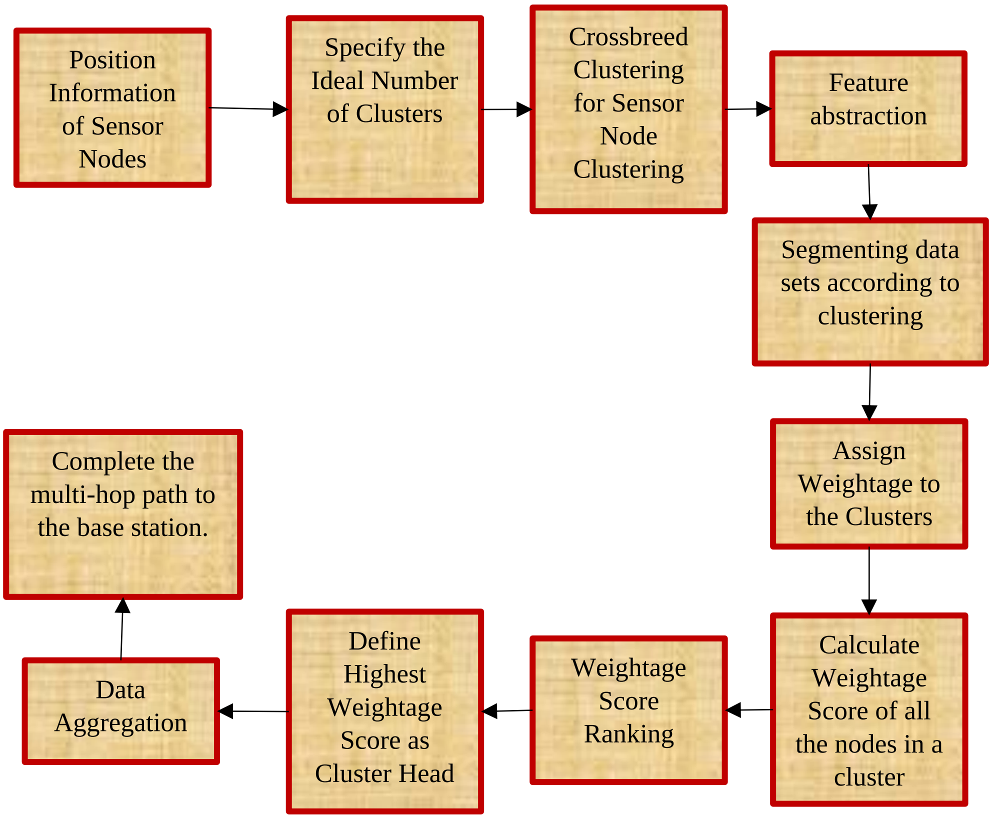

Figure 2: The Proposed Algorithm Block Diagram.

K-means clustering and the Nave Bayes classification algorithm are two clustering approaches that are combined in the proposed method to classify the sensor nodes into the four groups shown in Figure 2.

The Proposed Algorithm:
Input: X-coordinate value (x) and Y-coordinate value (y)
Output: k Clusters.
Method: K-means clustering for the training dataset and Gaussian Naïve Bayes classifier for the test dataset.

1. Using min-max normalisation, X or Y features can be normalised.
2. The data is split into a training set (70%) and an evaluation set (30%).
3. Select K completely at random input places as the training set.
4. To the spots where j is an integer between 1 and k, assign the cluster centre's C_j.

(i). To prevent further movement of the cluster centres, repeat the steps below for each mi-th node. Using the aforementioned algorithm, the ith node is allocated to the cluster node that is closest to it in terms of distance.

$$d_i = \min d(m_i, C_j) \quad (1)$$

(ii). Each cluster's centre should be relocated so that it lies near the cluster's geometric mean.

$$C_j = \frac{1}{N}\sum_{i=1}^{N} m_i \quad (2)$$

5. Thus, the output of the K means technique is used to supplement the training dataset that serves as input to the Naive Bayes classifier.
6. Mean and standard deviation are used to summarise the characteristics of the training dataset.
7. The training dataset is partitioned into k subsets, one for each cluster.
8. The average and standard deviation are used to summarise a subset of data that has been clustered for readability.
9. The equation for the Gaussian probability density function is provided.
10. The likelihood of each cluster is calculated.
11. Estimating and comparing the likelihood of each cluster is how the data in the test dataset is cluster categorised.

5. Result and Analysis

This research simulated the ECCM algorithm in Python and MATLAB to guarantee its performance. The simulation setup settings (Table 1) are listed below. The LEACH algorithm is compared to the suggested technique for clustering and selecting cluster heads.

Table 1: Simulation Set-up details.

Constraint	**Value**
Initial Energy of Nodes	2 Joules
Nodes	500
Communication Range	20m
Base station Location	(1000,1000)
Packet Size	4000 Bits
Statistics accumulation Vitality	5 nj/bit/signal
Transmission Energy	50 nj/bit
Network Size	1000m*1000m
Communicate Amplifier Vitality	100 pj/bit/m^2

5.1. Residual Energy

When creating a WSN, energy consumption is a crucial factor that must be taken into account. The amount of energy wasted during packet transmission is shown by the average energy required to deliver a packet from a source node to its destination. Figure 3 displays the results of a LEACH study of the recommended system's residual energy. Generally speaking, as the suggested method gives the cluster's most stable CH priority, it uses less energy than LEACH, as indicated by the diagram.

5.2. Network Lifetime

The lifespan of the network is determined by measuring the interval of time between the first and the last node to fail Moreover, this is the point at which a node's reserve energy is exhausted. Figure 4 displays the fraction of still-active nodes in each iteration.

5.3. Throughput:

The average pace at which data may be securely carried via a link is known as throughput. Typically, B/S is the unit of expression for throughput.

Table 2: Enactment Evaluation of the Proposed Algorithm.

Number of Rounds	**Network Lifetime**		**Throughput**		**Residual Energy (J)**	
	LEACH	**ECCM**	**LEACH**	**ECCM**	**LEACH**	**ECCM**
500	500	500	4	6	100	100
800	412	480	9	19	78	92
1100	349	456	12	20	65	89
1400	109	432	15	23	53	71
1700	115	405	15	26	29	56

2000	50	225	15	26	10	45

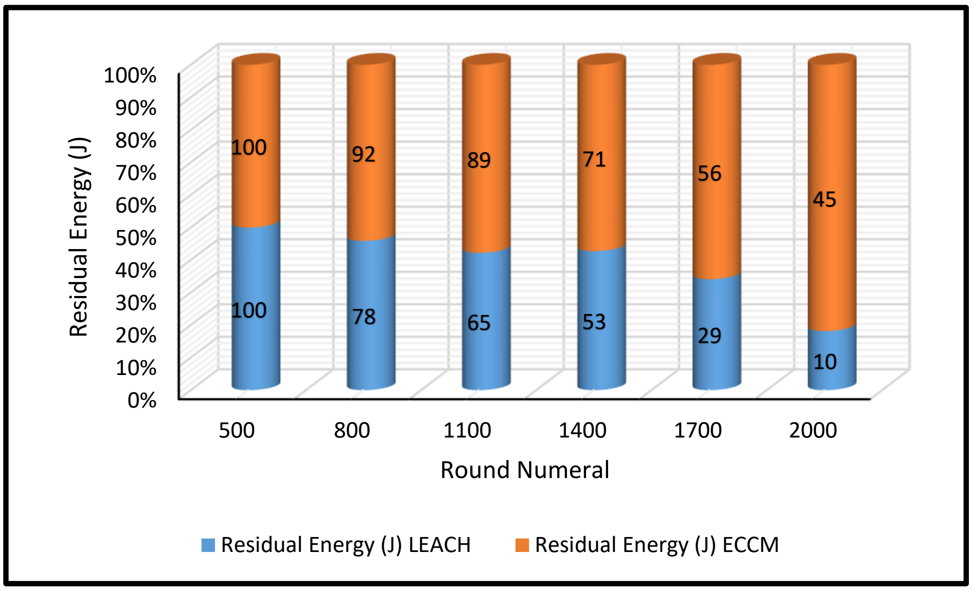

Figure 3: Evaluation of the Proposed Algorithm's Residual Energy (J).

Table 2 and Figure 3 illustrate this, showing that LEACH's 12% residual energy will run out after 2000.cycles, the 45% residual energy of ECCM does not. The energy efficiency of the suggested system in comparison to LEACH, thus grows linearly with time.

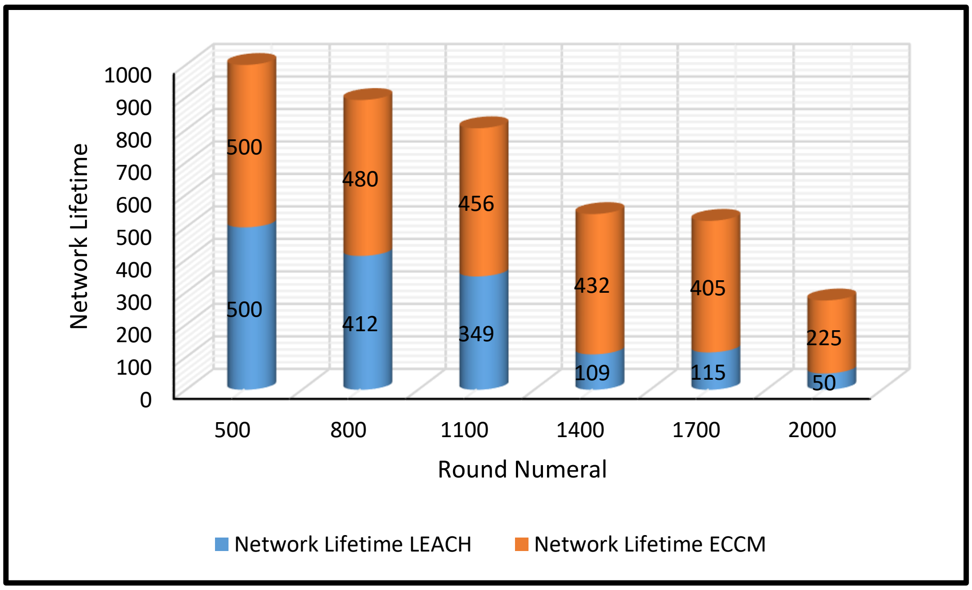

Figure 4: Evaluation of the Proposed Algorithm's Network Lifetime.

Table 2 demonstrates that LEACH has just 50 active nodes after 2000 cycles, compared to ECCM's 225 nodes. Figure 5 presents the throughput and number of cycles of an evaluation of the LEACH and recommended techniques. Table 2 shows that throughput increases linearly with the number of cycles and that the suggested method performs better than the LEACH algorithm.

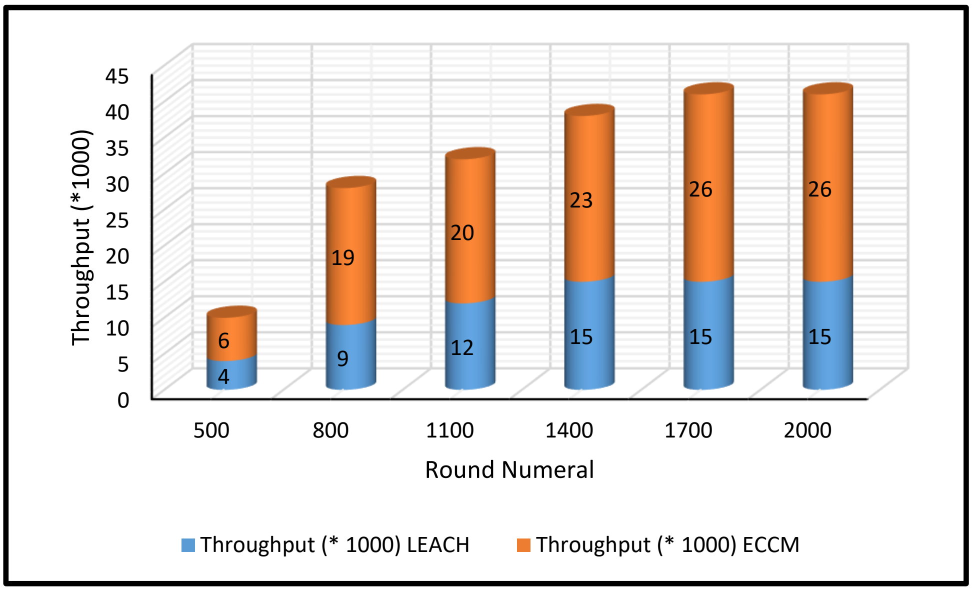

Figure 5: Evaluation of the Proposed Algorithm in Throughput.

When network resources are scarce, clustering becomes a major issue. Therefore, it's important to conserve as much power as possible. The proposed ECCM's performance was compared to that of the standard LEACH technique and found to be superior. A survey of variants of the LEACH algorithm was also undertaken.

6. Conclusion and Future Scope

The reduction of network latency, issues over network coverage, energy efficiency, and network lifetime are all key difficulties in WSNs. It is necessary to conduct additional research on cluster-based WSNs to find solutions to the problems of network durability and energy efficiency. Because of their singular properties, WSNs employ methods of machine learning. Rapid communication costs, low energy reserves, rapid mobility, and frequent topological shifts are some of the characteristics of this type of network. The current approach chooses the cluster heads at the beginning of each cycle at random, without taking into account the amount of energy that is still contained in these nodes. Given that they most likely have the lowest energy levels of all the nodes in the network, the recently chosen CH nodes will quickly leave the system. When communicating over long distances between cluster heads and the BS, which happens frequently in a large network, a significant amount of energy is wasted. This would imply that the lifespan of WSNs is limited in some way. Therefore, to boost the network's longevity as well as its efficiency, we suggest a strategy that is based on machine learning and is dubbed energy-proficient crossbreed clustering methodology (ECCM). The results of the experiments show that the ECCM is superior to the LEACH technique because it increases the amount of residual energy by 35%, extends the lifetime of the network by 37%, and increases throughput by 15%.
It is possible to test the provided approaches on any real-world application dataset and obtain reliable results. In the future, deep learning approaches can be used to solve the problems associated with mobility management.

Conflicts of Interest: "The authors declare no conflict of interest".

References

[1] Abdullah, M & Ehsan, A 2015, 'Routing Protocols for WSNs: Classifications and Challenges', Journal of Electronics and Communication Engineering Research', vol. 2, pp. 5–15.
[2] Alaybeyoglu, A 2015, 'A distributed fuzzy logic-based root selection algorithm for WSNs', Journal of Electrical and Computer Engineering, vol. 41, pp. 216–225.
[3] Amutha J, Sandeep Sharma, Jaiprakash Nagar (2020), "WSN Strategies Based on Sensors, Deployment, Sensing Models, Coverage and Energy Efficiency: Review, Approaches and Open Issues", Wireless Personal Communications Volume 111, pp:1089–1115.
[4] Asha Jerlin Manuel, Ganesh Gopal Deverajan, Rizwan Patan & Amir H Gandomi 2020, 'Optimization of Routing-Based Clustering Approaches in WSN: Review and Open Research Issues', Electronics, vol. 9, pp. 1-29.

[5] Batra, PK & Kant, K 2016, 'LEACH-MAC: a new cluster head selection algorithm for WSNs', Wireless Networks, vol. 22, issue 1, pp. 49–60.
[6] Gherbi, C, Aliouat, Z & Benmohammed, M 2017, 'A survey on clustering routing protocols in WSNs', Sensor Review, vol. 37, issue 1, pp. 12–25.
[7] Mezouary, RE, Choukri, A, Kobbane, A & Koutbi, ME 2016, 'An energy-aware clustering approach based on the K-means method for WSNs', Advances in Ubiquitous Networking, Springer, pp. 325–337.
[8] Praveen Kumar, D, Amgoth, T & Annavarapu, CSR 2019, 'Machine learning algorithms for WSNs: A survey', Information Fusion, vol. 29, pp. 1-25.
[9] Rao, PCS, Jana, PK & Banka, H 2017, 'A particle swarm optimization based energy efficient cluster head selection algorithm for WSNs', Wireless Networks, vol. 23, pp. 2005–2020.
[10] Ray, A & De, D 2016, 'Energy efficient clustering protocol based on k-means (EECPK-means)-midpoint algorithm for enhanced network lifetime in WSN', IET Wireless Sensor Systems, vol. 6, issue 6, pp. 181–191.
[11] Shahraki, A, Taherkordi, A, Haugen, O & Eliassen, F 2020, 'Clustering Objectives in WSNs: A Survey and Research Direction Analysis', Computer Networks, vol. 107376.
[12] Shi, C, Wei, B & Wei, S 2021, 'A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm', EURASIP Journal on Wireless Communication and Networking, vol. 2021.
[13] Tripathi, A, Gupta, HP, Dutta, T, Mishra, R, Shukla, KK & Jit, S 2018, 'Coverage and Connectivity in WSNs: A Survey, Research Issues and Challenges', IEEE Access, vol. 6, pp. 26971–26992.
[14] Yuan, C & Yang, H 2019, 'Research on K-Value Selection Method of KMeans Clustering Algorithm', J Journal, vol. 2, issue 2, pp. 226–235.
[15] Hung, L. Energy-Efficient Cooperative Routing Scheme for Heterogeneous WSNs. IEEE Access 2020, 8, 56321–56332.
[16] Lai, Z.; Xu, J.; Bowen, C.; Zhou, S. Self-powered and self-sensing devices based on human motion. Joule 2022, 6, 1501–1565.
[17] Ma, D.; Lan, G.; Hassan, M. Sensing, computing, and communications for energy harvesting IoTs: A survey. IEEE Commun. Surv. Tutor. 2019, 22, 1222–1250.
[18] Sachithanantham, N.C.; Jaiganesh, V. An Enhanced Efficient Leach Protocol (EELP) with Novel Cross Layer Technique to Forward the Data Packets in WSNs. Int. J. Innov. Res. Appl. Sci. Eng. 2021, 4, 841–849.
[19] Devising, R.S.S.; Kumar, A. A Study of Energy Efficient Load Balancing Multipath Routing Scheme for WSN. J. Compos. Theory 2019, 12, 1479–1492.
[20] Takale, S.; Lokhande, S. Quality of Service Based Routing Algorithm for WSN. J. Univ. Shanghai Sci. Technol. 2021, 23, 156–164.
[21] Simeon, O. Application of K-Means Clustering Algorithm for Selection of Relay Nodes in WSN. Int. Multiling. J. Sci. Technol. 2020, 5, 2811–2818.
[22] Purushothaman, R.; Narmadha, R. Scalable and Detect Link-Failure Traffic Balancing Network Using Adaptive Filter. Biosci. Biotech. Res. Commun. 2020, 13, 32–38.
[23] Singh, J.; Kaur, R.; Singh, D. A survey and taxonomy on energy management schemes in WSNs. J. Syst. Archit. 2020, 111, 101782.
[24] Yıldırım, K.S.; Carli, R.; Schenato, L. Adaptive Proportional–Integral Clock Synchronization in WSNs. IEEE Trans. Control Syst. Technol. 2018, 26, 610–623.
[25] Cheour, R.; Khriji, S.; Gotz, M.; Abid, M.; Kanoun, O. Accurate Dynamic Voltage and Frequency Scaling Measurement for Low-Power Microcontrollers in WSNs. Microelectron. J. 2020, 105, 104874.