International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.7043
Article Received: 08 March 2023 Revised: 30 April 2023 Accepted: 08 May 2023

Survey on Various Aspects of Clustering in Wireless Sensor Networks Employing Classical, Optimization, and Machine Learning Techniques

Ms. Monica Shivaji Gunjal[1], Pramodkumar H. Kulkarni[2]
[1]Research Scholar,
[1]Department of Electronics & Telecommunication Engineering,
Dr.D.Y.Patil Institute of Technology,Pimpri,
Pune, Maharashtra
monicagunjal@gmail.com
[2]Department of Electronics & Telecommunication Engineering,
Dr.D.Y.Patil Institute of Technology,Pimpri,
Pune, Maharashtra
Pramod.kulkarni@dypvp.edu.in

Abstract: A wide range of academic scholars, engineers, scientific and technology communities are interested in energy utilization of Wireless Sensor Networks (WSNs). Their extensive research is going on in areas like scalability, coverage, energy efficiency, data communication, connection, load balancing, security, reliability and network lifespan. Individual researchers are searching for affordable methods to enhance the solutions to existing problems that show unique techniques, protocols, concepts, and algorithms in the wanted domain. Review studies typically offer complete, simple access or a solution to these problems. Taking into account this motivating factor and the effect of clustering on the decline of energy, this article focuses on clustering techniques using various wireless sensor networks aspects. The important contribution of this paper is to give a succinct overview of clustering.

Keywords: Wireless Sensor Networks (WSNs), Clustering techniques, Energy Efficiency, Classical Methods, Optimization Methods, Machine Learning Methods.

I. INTRODUCTION

The size, simplicity, cost-effectiveness, and ease of use of Wireless Sensor Networks (WSN) make them noteworthy and promising for deployment. It is based on wireless sensor nodes placed in an area of interest to examine and record environmental information from the vibration, sound, temperature and light transmits the data to the base station as a result [1-3]. Thus, these node sets have the benefit of ensuring accurate sensing. WSNs are used in numerous areas, including healthcare, transportation, industrial, public safety, military systems, and environmental monitoring. Each node uses a to interact with the Base Station (BS) for data transport via a single hop or many hops [8,9]. Cluster Heads (CH) are discovered within each cluster as a result of the clustering process, which involves gatherings of sensor nodes into clusters. The information is then gathered and shared by CH to the base station via data transport through a single or more hop to the BS [10]. Data transfer in single-hop, the CHs send the acquired data straight away to the base station whereas, in multi-hop data transfer, CHs transfer the data to next-level CHs and finally reach the BS. The Clustering decreases interference and cuts down on power use which improves network performance. However, based on various methodologies and optimization, a thorough overview of clustering techniques like Classical, Optimization and Machine learning for WSNs is proposed.

II. RELATED WORK

Table 1 lists the current survey publications that concentrate on clustering in WSNs. The main contributions, protocols/algorithms, work done, and limitations are all examined. The general classification of clustering in wireless sensor networks includes different aspects that are summarized in Fig. 1.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.7043
***Article Received**: 08 March 2023 **Revised**: 30 April 2023 **Accepted**: 08 May 2023*

Table1: List of the current survey publications that concentrate on clustering

Ref	Year	Work done	Algorithms/Protocols	Limitation
[1]	2017	There is a functional alternative to cognitive radio-assisted wireless sensor networks (CR-WSN) that can be used for several WSN applications, including real-time surveillance, indoor sensing and heterogeneous multi-class sensing.	ACO, FLA, ANN, NPGA, SPEA, PSO, RL, FA	Different methods are investigated based on both conventional and Optimization approaches. However, research priorities does not emphasize multi-hop dynamic transmissions installation of nodes and finding Optimizing issues within 3D networks
[2]	2018	Extensive Simulation of proposed ABCO based LEACH algorithm in different 4 scenarios is tested. Comparison of traditional LEACH Algo. and the proposed algorithm is done.	LEACH,ABCO-based LEACH	Only 4 scenarios are considered. Other scenarios could consider for testing traditional LEACH algo. and proposed algo. Even comparison can be done with other optimization algorithms.
[3]	2018	Classification of clustering is done based on different categories and each category is elaborated separately, which helps future researchers.	Optimization algorithms, LEACH algorithms, fuzzy based algorithms etc.	Very few application specific implemented papers are considered for the literature survey.
[4]	2020	A comprehensive survey done on various clustering algorithms. About 215 important clustering algorithms are extracted for study and summarized.	Total 215 clustering Techniques.	More efforts are required to address the problem of mobility and heterogeneity.
[5]	2020	The Localization problem is WSN is focused and metaheuristic algorithm used to solve this problem	ABC, Bat algo., elephant herding algorithm	Need to design a new optimization algorithm to solve the problem of localization.
[6]	2021	Comparative Study of all LEACH variants done. Pros and cons of all variants discussed.	LEA, LEACH-B ,I-LEACH, V-LEACH, W-LEACH	Parameters like fault tolerance, high security & QOS need to be considered. And deployment of the node can be considered to be fixed.
[7]	2022	Extensive review of different clustering algorithms is done. New method proposed to improve energy efficiency, scalability & data correlation in WSN	LEACH.PSO, Genetic Algo.,ABC, ACO, Fuzzy Logic based algorithm	The newly proposed method can be compared with machine learning based algorithms for clustering.
[8]	2022	Machine Learning methods are added with optimization algorithms to improve energy efficiency & network lifetime of the WSN-IoT network..	AI techniques, Metaheuristic optimization algorithms	This paper does not focus on specific application based paper.

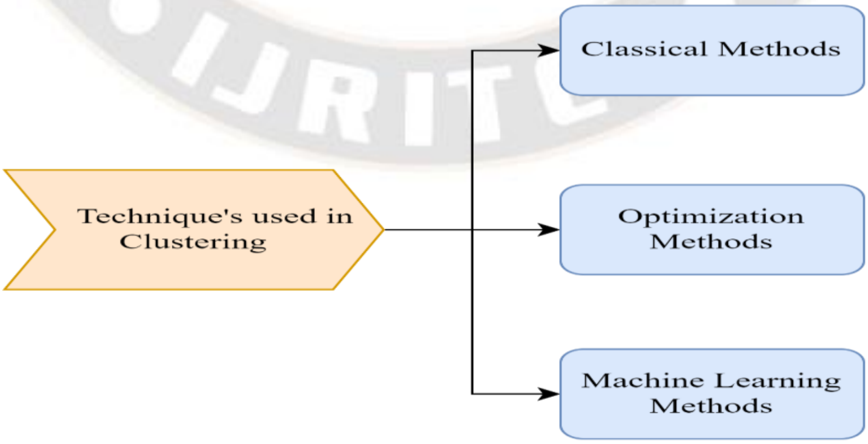

Fig.1. Different Clustering Aspects

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.7043
***Article Received**: 08 March 2023 **Revised**: 30 April 2023 **Accepted**: 08 May 2023*

The life of wireless sensor networks can be prolonged, by using data fusion. It can successfully lower the amount of network energy usage and data transmission. The current data fusion solutions, however, result in increased power and delay overhead. In WSNs, a fusion approach based on hybrid delay-aware clustering (HDC) is suggested. [9] Due to resource limitations, the Internet of Things is built using billions of next-generation sensors (IoT). Because of the widespread use of IoT devices and their current security measures fall short of the requirements for harmful attacks under dynamic conditions. Additionally, due to the open-access and unstable nature of the Internet medium, conventional routing techniques are compromised in terms of data security and exposed to a variety of dangerous threats. this research introduces a new lightweight structure-based Data Aggregation Routing (LSDAR) protocol for the Internet of Things Next-generation sensor networks integrated for node level energy routing performance improvement [10].In Wireless Sensor Networks, there is an opportunity for attackers to inject forged data and delete some blocks during data fragmentation. Therefore, it is necessary to confirm both the accuracy of the data and the integrity of these broken blocks. DoS attacks must also be avoided. The author proposed a design to take into account the a aforementioned points. Data validation and integrity checking are used in a trust-based data aggregation process in the WSN. This protocol uses the trust value, which is calculated for each node based on the trust attributes. [11]

In WSN, safe data aggregation increased interest in the scientific field. Data aggregation is necessary for efficiently using constrained sensor node resources. A model is essential which improves data accuracy and decreases the transmission of unnecessary and redundant data. Consequently, the rate and overhead of energy use are reduced. As a result, to balance the data on energy efficiency processing in a large-scale WSN with secure data aggregation the author has developed an optimal security model employing improved complete homomorphism encryption (OSM-EFHE). [12]

An agile data fusion algorithm based on hybrid delay-aware clustering (HDC) in WSNs is proposed to increase the performance of these wireless sensor networks. HDC combines the benefits of single-layer and multi-layer cluster structures and adaptively selects the clustering patterns by the decision function to achieve the trade-off between network delay and energy consumption. A theoretical examination of the delay and energy consumption of single-layer clusters and multi-layer clusters is offered along with the HDC network model. Additionally, the dynamic cluster head re-selection technique and the energy-efficient clustering algorithm are suggested to maximize network load balancing and energy usage. [13]

In this study, the authors present a clustering algorithm that chooses cluster heads using an enhanced artificial bee colony (ABC) algorithm. The clustering problem in WSN is solved by introducing the network cluster head energy, cluster head density, cluster head position, and other comparable characteristics into the modified ABC algorithm theory, which is based on the conventional ABC algorithm. For intra-cluster communication, a polling control mechanism based on busy/idle nodes was implemented [14]. The limitations of the conventional LEACH protocol are highlighted by the author, and additional improvements are made to lengthen network lifespan by lowering energy usage. This research highlights the shortcomings of the present LEACH[15].

This study proposes GACRP, an enhanced genetic algorithm-based annulus sector clustering routing protocol. In According to GACRP, the sectors in the circular network are the same size for each annulus. To determine the minimum energy consumption of the system, a cluster is formed by each annulus sector, and the node that is best in each annulus sector is chosen to be the cluster head. [16]

The majority of the heuristics algorithms in use today for finding the provision of effective routing and transportation solutions are constrained by the optimal routing strategies for clustering in bigger search areas. Consequently, when the search space expands the probability of finding the ideal clustering and routing solution grows exponentially as the resources of the sensor node are gradually depleted by an optimized operation. To overcome the issues and restrictions posed by the current routing systems, two new heuristics algorithms have been proposed by the author I)A gravitational approach based clustering method II) A clustered gravitational routing algorithm. [17] In this paper, a new adaptive coding routing protocol is developed that incorporates RS and LDPC codes into routing protocol to increase the reliability of the communication link and reduce the required transmission energy. Comparison is done with LEACH and BRE- LEACH, in terms of throughput, network longevity, stability, and usage [19]. In this paper, author focuses on reducing the power consumption of sensor nodes to increase their lifespan during the data transmission and reception processes by utilizing the firefly algorithm for head selection. An acceptable CH is chosen using a grid-based clustering technique from networks clusters. For routing Proficient Routing based Power-Efficient Gathering in Sensor Information (PR-PEGASIS) hierarchical routing protocol is used. Active Distortion Control Artificial Neural Network (ADC-ANN) is used to recognize the distortion in the network. [20]. This study provides a comprehensive explanation of how several published optimization methods for the WSN industry. [21]

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.7043
***Article Received**: 08 March 2023 **Revised**: 30 April 2023 **Accepted**: 08 May 2023*

A clustering-based protocol's requirements rely mostly on choosing the most suitable influencing elements that demand an optimal cluster head nodes (CHs) and next-hop nodes (NHs). To address this decision criterion a fuzzy inference system can be used (FIS). Even while FISs can offer satisfactory choices for CHs and NHs in clustering procedures. FIS elements like fuzzy input variables, fuzzy membership functions, and fuzzy rules are defined manually. The author implemented, an improved shuffled frog leaping algorithm (ISFLA) and an enhanced fuzzy-based clustering protocol in this study. The suggested protocol, known as EFC-ISFLA, develops a fuzzy-based clustering protocol that has been optimized to keep the network's lifespan intact. [18]. This study shows the study of Mayfly Optimization Algorithm (MOA) in conjunction with an energy-efficient routing protocol to manage Cluster Head (CH) selection. As a result, data delivery between CH and BS takes place. It Reduces energy usage by rotating CH selection based on optimization throughout the network. The proposed Mobile Sensor Node Energy Coherent MOA (MECMOA) compared with Low Energy Adaptive Clustering Hierarchy (LEACH). The authors of this research suggest employing three different algorithms to enhance the performance of a sensor network in terms of sink node location as well as route building and optimization using computational techniques inspired by natural processes. Opportunistic coding is also utilized at possible relays to cut down on transmissions. The three algorithms that make up the suggested implementation combine their strengths for considerably improved data transmission. The placement of the sink node is done first using particle swarm optimization, and the creation of the route using the sensors and sink of the specific cluster comes next. Before broadcasting to neighbors, the wiener-spanning tree undergoes opportunistic packet amalgamation, and an artificial bee colony approach for further optimization is used.[22]

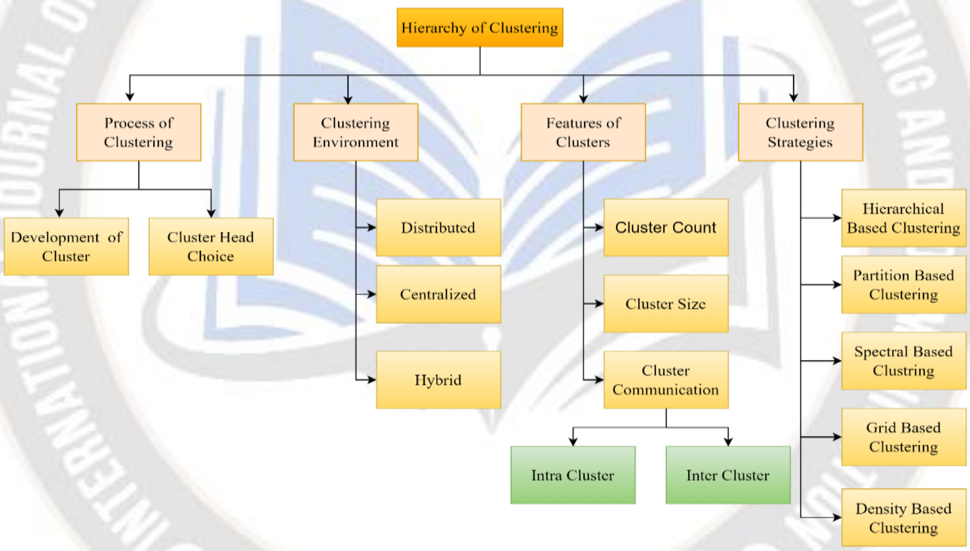

Fig.2. Hierarchy of Clustering

III. THE HIERARCHY OF CLUSTERING

Fig.2 illustrates the hierarchy of cluster analysis based on different strategies, including environment, features, methodologies, process. The examined clustering tactics are analyzed and divided into several categories using a variety of variables, including the type of nodes, mobility pattern of node, node deployment criteria, performance metrics, and sink node positions.

3.1. Clustering environment

The WSNs can either be distributed, centralized or hybrid in nature depending on the clustering environment. In distributed clustering, the sensor nodes collaborate to create routes or join a cluster that has already been formed without access to global information such as cluster location and energy level. By running its algorithm, each sensor node can complete its job and decide whether to become the cluster head or not. To reduce energy use, WSNs employ distributed clustering techniques. In centralized clustering, the base station or cluster node needs comprehensive knowledge of the network to manage the entire procedure. The nodes are organized into clusters using these methods, and each cluster is given a CH. Based on the transmission of trustworthy data, scalability, and minimization of network ratio, centralized clustering restricts the network's performance.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.7043
***Article Received**: 08 March 2023 **Revised**: 30 April 2023 **Accepted**: 08 May 2023*

3.2 Features of the cluster

Clustering technique includes features like cluster size, cluster count, and cluster communication that are related to the cluster's internal structure. Clustering scheme variability was the focus of the two types of cluster counts, fixed and variable. The cluster size is divided into two categories: equal and unequal, with the same and differing cluster sizes, respectively. The network's energy consumption is more evenly distributed because of the unequal clustering, which also eliminates the energy-hole issue. Clustering allows for both intra- and inter-cluster communication between sensor nodes. Communication takes place between the sensor nodes and the CH during intra-cluster communication. On the other hand, in inter-cluster communication, the CHs and the BS or sink node communicate.

3.3. Clustering strategies

Numerous methods, including grid-based, partition-based hierarchical, spectral, and density-based clustering, can be used to carry out the clustering process. In top-down or bottom-up methods, a tree structure is used in hierarchical clustering. It is ideal for point-to-point communication and very versatile [16]. The clusters in partition-based clustering are divided into several divisions, with each division representing a cluster that is suitable for a small number of nodes [13].

3.4. Process of clustering

The clustering process in WSN consists mostly two phases cluster creation and cluster head selection. It maximizes the network life duration.

3.4.1. Development of clusters

By having fewer cluster members in this phase, the stress on CHs close to the sink node is reduced. Regarding the Received Signal Strength Indication, each member node joins the closest cluster leader. Each node is eligible to serve as a cluster head by computing its radius based on the local data it has received [34].

3.4.2. Cluster head choice

The choice of CH is essential for maximizing energy usage because CH is largely used for collecting and sending data to the sink node [14]. The WSN lifetime is improved by the appropriate CH selection, but [29]. In a cluster-based network, hotspot issues arise because CH close to the base station soon uses up all of its energy. To solve this issue, unequal clustering methods are used [30]. The proper CH selection emphasizes on a variety of factors, including cluster head mobility, Cluster Head selection, communication, and the role of Cluster Head. The cluster head (CH) selection criteria place particular emphasis on metrics such as cluster density, residual energy, cluster head and sink nodes and RSSI.

IV. TECHNIQUES EMPLOYED IN VARIOUS CLUSTERING PROCESS

The clustering process are carried out using strategies using classical methods, optimization strategies, and machine learning techniques, as shown in Fig.1. The Various aspect like establishment of clusters, the choice of cluster leaders, and routing protocols, security, dependability, and uneven clustering are considered.

4.1. Classical Method

The classical method is used in traditional ways to calculate the number of clusters. Additionally, traditional methods are employed to replace the function of CH that determines the nodes' remaining energy or energy consumption depending on the threshold, which initiates re-clustering of the network [16]. Traditional techniques emphasize the method used in cluster-based routing systems to choose cluster heads. Because cluster heads are chosen at random, traffic between them is erratic. WSNs apps can use the selection process, but they still need to deal with challenges such connectivity, load balancing, coverage, and scalability [28]. Each node has an equal opportunity to become a CH under the LEACH [29] protocol. however, it cannot be a CH in subsequent rounds, managing the responsibilities shared by CHs MECH evaluates the energy and the quantity of hops while sending data to the BS. Comparatively, TL-LEACH reduces the size of the data because there are two aggregation phases. T-LEACH does not typically compare CHs in terms of the established cutoff point. The overlapping parameters used by FBR prevent the nodes from being re-clustered. Next in HEED multihop routing is used when CH transfer data to the destination node. LEACH enhances TEEN, a reactive clustering routing technology. Each cluster's cluster head (CH) gathers information from its cluster members. Data is combined, processed, and sent to the BS or a higher level CH by CHs.

4.1.1. Applications

For event-detecting WSNs, periodic reporting scenarios, high mobility scenarios, periodic data gathering applications, as well as proactive and reactive applications, classical approaches are more suited. In asynchronous sensor network applications, it makes easier for BSs to get node interpretations and make decisions.

4.1.2. Benefits

It is difficult to choose the same node to serve as the CH in a subsequent round so that all nodes can share the same load. Cluster members limit the further evaporation of energy when

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.7043
***Article Received**: 08 March 2023 **Revised**: 30 April 2023 **Accepted**: 08 May 2023*

their appointed times have passed. The traditional technique creates balanced cluster information using the location of a node. This increases intra-cluster communication and cluster size. Simply it does not depends on the size and topology of the network. Low range of transmission required for long-distance communication are produced by the two-level clustering strategy, which decreases the total amount of energy used. The number of CHs being reduced efficiently covers the entire network and reduces the overlapping of the cluster, enhancing the potency of the operating algorithms at the CHs level.

4.1.3. Limitation

Most traditional approaches, like LEACH, only do single-hop inter-cluster communication, which is insufficient for a large-scale system. Too much energy is consumed in the network environment. The different coverage, energy hole, and hot spot difficulties occur because many nodes with low energy levels serve as CHs. Higher energy nodes prematurely perish during similar rounds. HEED incurs significant costs due to the creation of clusters require numerous assessments, calculating the number of cluster instances while accounting for the cost of energy connection and utilization are difficult. A large-scale setting results in significant energy usage because from the source to the BS of the inter-cluster routing. The concern nodes CH selection without initial energy levels do not fulfilled load balancing issues.

4.2 Optimization Method

Optimization in the Wireless Sensor Network (WSN) is required to cut down on redundant processes and energy use. Data aggregation is necessary for wireless sensor network optimization for secure data transfer at both the cluster head and base station. The standard methods fail to tackle NP-hard problems within a given amount of time. The optimization methodology is utilized to find a solution. Heuristic or meta-heuristic optimization are both possible. Heuristic methods focus on a single issue at a time, because of their gluttony and inclination toward wasting time, the remedies stuck in a neighborhood minimal. Additionally, the heuristic makes use of mistakes in trying to solve difficult issues. Use of meta-heuristics finds a global minimum on the set of issues one problem at a time. We might not always get the best answer via meta-heuristics, but offers an ideal answer the majority of the time. Fig. 3 Shows Categorization of optimization method. Meta-heuristic Techniques and clustering methods are coupled to achieve a network that uses less energy by choosing the best options.

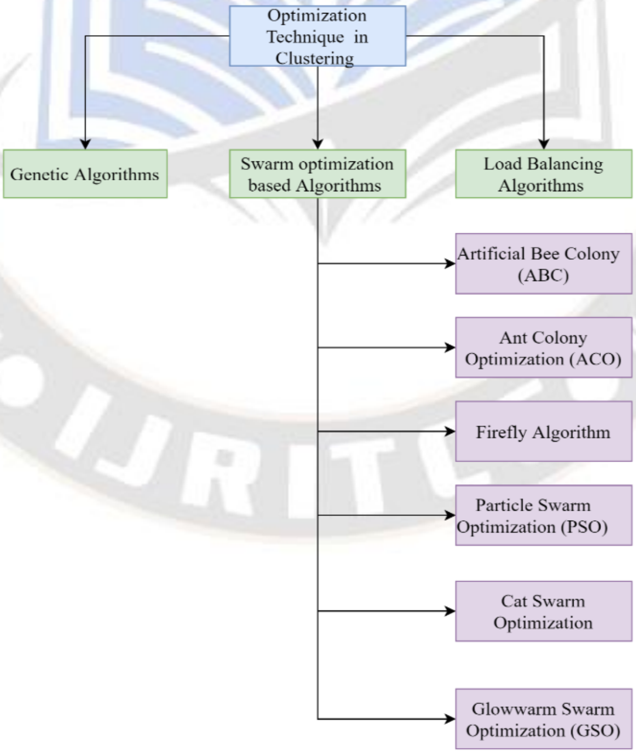

Fig.3.Optimization Techniques in Clustering

In terms of imitating behavior, proposed solutions and limitations comparisons between various optimization techniques employed in WSNs are made available in Table 2.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.7043
Article Received: 08 March 2023 Revised: 30 April 2023 Accepted: 08 May 2023

Table 2: Tabular survey of optimization approaches in WSN

References	Year	Algorithm	Imitate	Proposed Solution	Limitation
[23]	2017	ACO	Ants searching for food sources	finds the best route	Inadequate in terms of energy use
[27]	2016	ABC	Bees, in search of a food source	efficient for issues with low convergence	Slow data-transmission efficiency
[25]	2020	Firefly Algorithm	Fireflies use flashlights to help them hunt for prey.	helpful for multi-objective issues	Identification of a compromised path and data transmission
[26]	2020	Grasshopper	Swarms of grasshoppers are looking for food	able to tackle real-world issues in unexplored search areas	All search agents must be involved in determining the next position
[33]	2020	PSO	Birds are searching for food sources.	determines the best option based on energy usage	Increased network overhead
[36]	2019	CSO	based on the behavior of cats	Seeking Mode and Tracking Mode are used to find the optimal solution	premature convergence problem

4.2.1. A Genetic algorithm

The Genetic Algorithm (GA) is an evolutionary algorithm that is primarily utilized in WSNs for routing and clustering schemes. GA is applied to lengthen the lifespan of cluster heads during clustering.[34].A combined approach using GA and bee colony method, the clustering process is enhanced and improved. The quantity of energy used by each round also minimizes data by shortening the overall distance. The evolutionary algorithms can resolve non-continuous problems. It has some drawbacks for multi-dimensional, non-differential, and non-parametric problems. These include: longer Running times are unable to manage a huge population or guarantee consistent optimized response times.

4.2.2. Swarm Based Optimization

Another field, known as swarm intelligence (SI), which develops algorithms by drawing inspiration from animal or insect social structures. The Grasshopper, Firefly Algorithm, Artificial Bee Colony, Particle Swarm Optimization and Ant Colony Optimization are a few of them.

The artificial bee colony was inspired by the foraging behavior of bee colonies. There are three "bees" in the ABC algorithm. These three bees in the "colony" include observers, scouts, and working bees. Bee populations are used by the ABC algorithm to detect the ideal route. A bee waiting to make a decision on the "dancing" area and an observer is a food source, and a bee aimlessly hunting is a heading to a previously visited food source with a scout and a worker bee. Food supplier’s locations are a representation of potential remedies for the optimization issue and the quantity of a food source's "nectar" reflects its quality.

The ACO method can be easily modified to tackle both WSN problems. Load balancing and routing issues that ACO can be used to resolve.

The firefly algorithm is biologically inspired algorithm. It works on firefly's flashing characteristic. The next three guidelines are incorporated into this algorithm.1. The opposite mate attracts the fireflies.2. Using the brightness value, the attractiveness of the fireflies’ calculated .3.The objective function utilized to determine the illumination of the firefly.

In the Particle Swarm Optimization (PSO) method, each "bird" or particle stands for a distinct solution. This particle has a fitness value that measures the effectiveness of the answer [33]. In the Binary PSO (BPSO) algorithm the cluster head is chosen ideally using the PSO method and it is a two-tier PSO.

In CSO algorithm, cues taken from cats tracing and napping habits. Cats appear to be slothful creatures that prefer to lie around and sleep. However, while they were resting they have a very high level of consciousness and are very aware of currently happening around them. As a result, they are always watching the environment thoughtfully and intelligently, and when as they spot a target, they swiftly begin migrating in that direction. Consequently, the CSO algorithm is built on combining these two primary mannerisms of cats. The CSO algorithm has two modes, namely, modes of tracking and seeking. Each kitten stands for a solution with a position, a fitness rating, and a flag of its own. [24]

Improved Genetic Algorithm (IGA) and Binary Ant Colony Algorithm (BACA) are used in a hybrid meta-heuristic technique that ensures optimal coverage, reduces the sensing of redundant information, and optimizes the multi-objective

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.7043
Article Received: 08 March 2023 Revised: 30 April 2023 Accepted: 08 May 2023

function by selecting the smallest number of sensors synthetic bee. WSNs use a path for routing, by modifying the sleep/awake scheduling of Grasshopper. The Fractional Grasshopper Optimization Algorithm (Fractional-GOA) approach reduces energy consumption. Load-balanced clustering-based approximation methods used because of aggregation, communication to the sink node and data processing during clustering, the cluster heads use more energy. Therefore, load balancing is crucial to the operation of WSN. According to approximation algorithms, there are two load-balanced clustering schemes: The Greedy Load Balanced Clustering Algorithm (GLBCA) and the Load-Balanced Clustering Algorithm (LBCA).

4.2.3. Applications

Applications where changes in mobility and topology are of the utmost importance can benefit from swarm intelligence. Evolutionary Algorithms can be used to solve some centralized issues like energy-aware routing, scheduling, localization, sensor fusion, and clustering. Continuous parameter estimation issues related to laboratory experiments are addressed using evolutionary strategies. A chromosome provides the optimal answer to the particular problem in the continuous or integer optimization issues that the GAs tackle. Based on a variety of requirements, the GA application and fitness function define the dependability of the chromosomes. Additionally, a GA-based technique is used to solve clustering, categorization, and configuration management.

4.2.4. Benefits

Ant colony optimization is a routing technique that has been recognized as a powerful and scalable method. The likelihood of enabling nodes to maintain strong connectivity to the BS is taken into account while choosing the route by ACO-based routing algorithms. ACO allows for multi-path information transfer in WSNs and it is versatile, secure, and offers routing and data aggregation. It also reduces energy usage and network traffic. By simultaneously using a fitness assessment function and integrating partial solutions through a crossover, the GA algorithm efficiently analyses the search space. A variety of potential solutions are examined and manipulated by genetic algorithms. GA uses multi-criteria optimization to solve problems while keeping a search threshold. The more specific benefit of GA is its capacity to provide rule-based, workable solutions to a variety of issues using machine learning and pattern recognition.

4.2.5. Limitations

Evolutionary algorithms are not adaptable to changes in the environment. They have significant memory and processing needs. Only per-deployed clustering schemes the majority of which are suggested for homogeneous networks are practical for clustering in WSNs. ACO algorithms have a number of drawbacks, including the performance being very cycle-dependent. As a result, the ACO algorithm uses a lot of overhead communication to handle the routes separately. The ABC algorithm's drawbacks include its slow convergence rate and tendency to become quickly stuck in the local optimum. If the path's reliability deteriorates significantly, the forager will stop employing fresh swarms, according to social insect capabilities. The clustering procedures are entirely centralized using the firefly algorithm.

4.3 Machine learning Methods

Machine learning (ML) is a method that gains knowledge from practice or research. Machine learning is an integral part of many interdisciplinary professions, including engineering, medicine, and computers. Recent developments in machine learning help to address a variety of WSN related problems. Without ML, accessing enormous amounts of sensor data in WSN and gleaning insights from the gathered data is not an easy operation. ML automates the analysis of complex data and improves the efficiency and cost-effectiveness of computer processes. It can be broadly divided into three categories: supervised, unsupervised, and reinforced learning. The categorization of ML techniques in WSN is shown in Fig 4.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.7043
***Article Received**: 08 March 2023 **Revised**: 30 April 2023 **Accepted**: 08 May 2023*

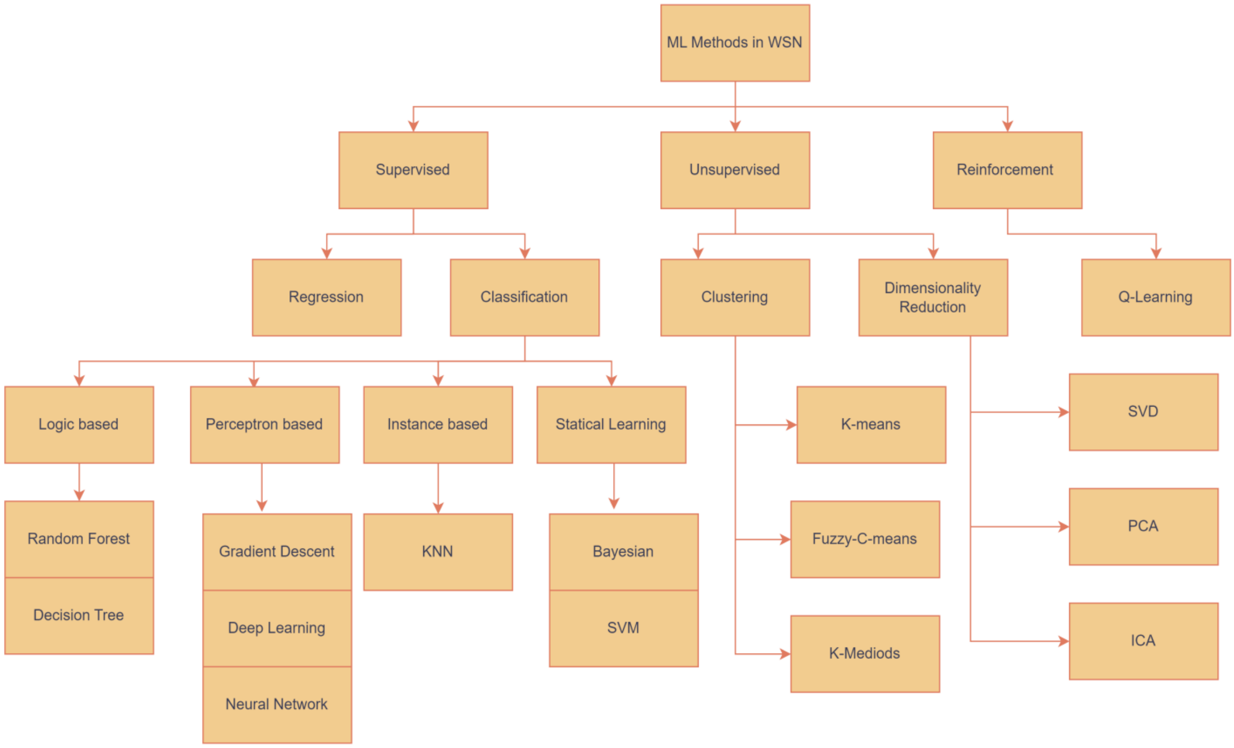

Fig.4. ML techniques used in WSN

4.3.1 Supervised Learning

A variety of inputs and outputs are offered in supervised learning. The goal of algorithms is to predict objective results and build a model that represents the relationships and dependencies between the input features. Regression and classification are two categories for supervised learning algorithms. In terms of minimum mistakes, regression estimates precision performance. i)Logic-based algorithms for classification that employ Random Forest (RF) and Decision Tree (DT) methodologies (ii) Perception-based algorithms that make use of Gradient Descent (GD),Deep Learning (DL), Artificial Neural Networks (ANNs), and Neural Networks (NNs). (iii) Instance-based algorithms that employ k-Nearest Neighbor (k-NN) methods .iv) Bayesian and Support Vector Machine (SVM)-based statistical learning algorithms

4.3.2. Unsupervised learning

Unsupervised learning produces no output that is related to the inputs, even the model that is developed extracts the relationship. Using this learning strategy, clusters of similar patterns are identified. WSN difficulties including routing, data aggregation, connectivity concerns, and anomaly detection are all resolved through unsupervised learning. Dimensionality reduction and clustering are the other divisions. Fuzzy-C-Means (36), K-Means, Hierarchical-based, and K-Medoids algorithms are clustering approaches. Singular Value Decomposition (SVD), Principal Component Analysis (PCA), and Independent Component Analysis are dimensionality reduction techniques (ICA).

4.3.3. Reinforcement Learning

The algorithm of reinforcement learning schemes continuously picks up new information from the interaction of the environments. It chooses the best outcome, which contributes to maximizing performance. In fact, WSN employs Q-learning as one of its reinforcement learning strategies. The following details explain why machine learning algorithms are utilized in WSNs: (i)Data transmission overhead is reduced at the sensor or CH level ii) The WSN performance is improved by energy harvesting and long-term maintenance (iii) Network efficiency is improved by separating faulty nodes from normal nodes and (iii) Accurate localization of the sensor node can be determined quickly, even if their locations are changed due to internal factors

4.3.4. Applications

Clustering, scheduling, and routing are examples of complex multi-objective problems for that fuzzy logic is ideally suited to define and resolve. A branch of mathematics called fuzzy logic (FL) expresses approximations in human reasoning. Pattern recognition, digital image processing, and several control systems, including power systems, elevators, and home appliances, all uses fuzzy logic. In a variety of disciplines including bio informatics ,pattern recognition, and business

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.7043
Article Received: 08 March 2023 Revised: 30 April 2023 Accepted: 08 May 2023

intelligence, fuzzy-c means clustering technique is employed. Neural networks can be used to address centralized problems including sensor fusion, clustering, data mining, routing, and scheduling. Additional viable neural network applications include localization, sink location, and optimal sensor nodes. Due to its greater processing requirements, ANN is used in many real-time WSN applications.

4.3.5. Benefits

The fuzzy logic is the best choice to perform clustering heuristics and routing optimization to simultaneously achieve several goals. Reducing uncertainty in decision-making and facilitate methodical research. Hyper spectral data is best classified using the Random Forest (RF) classifier. ANN classifies difficult and non-linear data sets relatively quickly. Deep learning techniques deal with methods of layer-based data processing. The Support Vector Machine (SVM) chooses the best hyper plane to categories the data. The K-Nearest Neighbor approach reduces dimensionality while simultaneously identifying any missing values.

4.3.6. Limitation

Although ML in WSNs has many benefits and a wide range of applications, it has several drawbacks. The FL method yields a non-optimal result, and fuzzy rules must be used to relearn topological alterations. The limited understanding of future information is the main problem with routing algorithms that emphasize reinforcement learning. As a result, the algorithms are inadequate for highly dynamic circumstances because it takes them a while to find the best route [31]. Since ML algorithms need to learn from past data, they cannot automatically produce accurate predictions. Performance is based on past data, if the data set is too huge, processing will consume a significant amount of energy. For example, connectivity, routing, energy harvesting, data aggregation just a few of the problems that machine learning techniques are used to solve in WSN (Table 3).

Table 3: Machine learning methods in WSN

Machine Learning Algorithm	Routing issue	Connectivity issue	Energy harvesting issue	Data Aggregation issue
Regression	×	√	√	√
Decision Tree	×	√	×	√
Random Forest	×	×	×	×
ANN	√	×	×	×
K-NN	×	×	×	√
SVM	×	√	×	×
Deep Learning	√	×	×	×
K-means	×	×	×	×
PCA	×	×	×	√
ICA	×	×	×	×

V. SUMMARY

In this article the hierarchy of cluster analysis based on different strategies, including environment, features of the cluster, clustering strategies, Process of clustering (Development of clusters, Cluster head choice) are discussed. Various techniques employed in clustering process like classical methods, Optimization methods and Machine Learning methods are discussed along with advantages, limitations and applications. For ease of reference, the exact comparisons are made using the proper criteria for each aspect. Comparing survey tables are from almost last decade papers.

VI. CONCLUSION

This survey focus on different clustering techniques and cluster oriented algorithms owing to the energy efficiency of nodes in WSNs and the significance of clustering as an effective option for decreasing energy consumption. The different features, such as cluster creation, cluster head selection, routing, dependability, security, and unequal clustering are all addressed by these algorithms. Different approaches are analyzed using traditional, optimization, and machine learning techniques.

Statement of Declaration:

No Such statement and declaration.

No Such competing interest and funding.

REFERENCES

[1] Ramya R, Brindha T. A Comprehensive Review on Optimal Cluster Head Selection in WSN-IoT. Advances in Engineering Software. 2022 Sep 1;171:103170.

[2] Al-Sulaifanie AI, Al-Sulaifanie BK, Biswas S. Recent trends in clustering algorithms for wireless sensor networks: A comprehensive review. Computer Communications. 2022 May 21.

[3] Sharma N, Gupta V. Meta-heuristic based optimization of WSNs Localisation Problem-a Survey. Procedia Computer Science. 2020 Jan 1;173:36-45.

[4] Daanoune I, Abdennaceur B, Ballouk A. A comprehensive survey on LEACH-based clustering routing protocols in Wireless Sensor Networks. Ad Hoc Networks. 2021 Apr 1;114:102409.

[5] Shahraki A, Taherkordi A, Haugen , Eliassen F. Clustering objectives in wireless sensor networks: A survey and research direction analysis. Computer Networks. 2020 Oct 24;180:107376.

[6] Gambhir A, Payal A, Arya R. Performance analysis of artificial bee colony optimization based clustering protocol in various scenarios of WSN. Procedia computer science. 2018 Jan 1;132:183-8.

[7] Fanian F, Rafsanjani MK. Cluster-based routing protocols in wireless sensor networks: A survey based on methodology. Journal of Network and Computer Applications. 2019 Sep 15;142:111-42.

[8] Rawat P, Chauhan S. Clustering protocols in wireless sensor network: A survey, classification, issues, and future directions. Computer Science Review. 2021 May 1;40:100396.

[9] Liu X, Zhu R, Anjum A, Wang J, Zhang H, Ma M. Intelligent data fusion algorithm based on hybrid delay-aware adaptive clustering in wireless sensor networks. Future Generation Computer Systems. 2020 Mar 1;104:1-4.

[10] Haseeb K, Islam N, Saba T, Rehman A, Mehmood Z. LSDAR: A light-weight structure based data aggregation routing protocol with secure internet of things integrated next-generation sensor networks. Sustainable Cities and Society. 2020 Mar 1;54:101995.

[11] Daniel D A, Roslin SE. Data validation and integrity verification for trust based data aggregation protocol in WSN.

[12] Shobana M, Sabitha R, Karthik S. An enhanced soft computing-based formulation for secure data aggregation and efficient data processing in large-scale wireless sensor network. Soft Computing. 2020 Aug;24(16):12541-52.

[13] Liu X, Zhu R, Anjum A, Wang J, Zhang H, Ma M. Intelligent data fusion algorithm based on hybrid delay-aware adaptive clustering in wireless sensor networks. Future Generation Computer Systems. 2020 Mar 1;104:1-4.

[14] Wang Z, Ding H, Li B, Bao L, Yang Z. An energy efficient routing protocol based on improved artificial bee colony algorithm for wireless sensor networks. IEEE Access. 2020 Jul 20;8:133577-96.

[15] Khan MA, Awan AA. Intelligent on demand clustering routing protocol for wireless sensor networks. Wireless Communications and Mobile Computing. 2022 Mar 29;2022.

[16] Chu-hang W, Xiao-li L, You-jia H, Huang-shui H, Sha-sha W. An Improved Genetic Algorithm Based Annulus-Sector Clustering Routing Protocol for Wireless Sensor Networks. Wireless Personal Communications. 2022 Apr;123(4):3623-44.

[17] Selvi M, Santhosh Kumar SV, Ganapathy S, Ayyanar A, Khanna Nehemiah H, Kannan A. An energy efficient clustered gravitational and fuzzy based routing algorithm in WSNs. Wireless Personal Communications. 2021 Jan;116(1):61-90.

[18] Kongsorot Y, Musikawan P, Muneesawang P, So-In C. An enhanced fuzzy-based clustering protocol with an improved shuffled frog leaping algorithm for WSNs. Expert Systems with Applications. 2022 Jul 15;198:116767.

[19] Daanoune I, Baghdad A, Ullah W. Adaptive coding clustered routing protocol for energy efficient and reliable WSN. Physical Communication. 2022 Jun 1;52:101705.

[20] Daanoune I, Baghdad A, Ullah W. Adaptive coding clustered routing protocol for energy efficient and reliable WSN. Physical Communication. 2022 Jun 1;52:101705.

[21] Parwekar P, Rodda S, Kalla N. A study of the optimization techniques for wireless sensor networks (WSNs). InInformation systems design and intelligent applications 2018 (pp. 909-915). Springer, Singapore.

[22] Singh A, Nagaraju A. Low latency and energy efficient routing-aware network coding-based data transmission in multi-hop and multi-sink WSN. Ad Hoc Networks. 2020 Oct 1;107:102182.

[23] Nasir HJ, Ku-Mahamud KR, Kamioka E. Ant Colony Optimization approaches in wireless sensor network: performance evaluation. Journal of Computer Science. 2017 Jun 24;13(6):153-64.

[24] Ahmed AM, Rashid TA, Saeed SA. Cat swarm optimization algorithm: a survey and performance evaluation. Computational intelligence and neuroscience. 2020 Jan 22;2020.

[25] Pitchaimanickam B, Murugaboopathi G. A hybrid firefly algorithm with particle swarm optimization for energy efficient optimal cluster head selection in wireless sensor networks. Neural Computing and Applications. 2020 Jun;32(12):7709-23.

[26]]Abualigah L, Diabat A. A comprehensive survey of the Grasshopper optimization algorithm: results, variants, and applications. Neural Computing and Applications. 2020 Oct;32(19):15533-56.

[27] Yue Y, Li J, Fan H, Qin Q. Optimization-based artificial bee colony algorithm for data collection in large-scale mobile wireless sensor networks. Journal of Sensors. 2016 Feb;2016.

[28] Krishnan M, Lim Y. Reinforcement learning-based dynamic routing using mobile sink for data collection in WSNs and IoT applications. Journal of Network and Computer Applications. 2021 Nov 15;194:103223.

[29] Abu-Baker A, Alshamali A, Shawaheen Y. Energy-Efficient Cluster-Based Wireless Sensor Networks Using Adaptive Modulation: Performance Analysis. IEEE Access. 2021 Oct 8;9:141766-77.

[30] Amutha J, Sharma S, Sharma SK. Strategies based on various aspects of clustering in wireless sensor networks using classical, optimization and machine learning techniques: Review,

ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.7043
***Article Received**: 08 March 2023 **Revised**: 30 April 2023 **Accepted**: 08 May 2023*

taxonomy, research findings, challenges and future directions. Computer Science Review. 2021 May 1;40:100376.
[31] Kulkarni PK, Malathi Jesudason P. Multipath data transmission in WSN using exponential cat swarm and fuzzy optimisation. IET Communications. 2019 Jul;13(11):1685-95.
[32] Nayak P, Swetha GK, Gupta S, Madhavi K. Routing in wireless sensor networks using machine learning techniques: Challenges and opportunities. Measurement. 2021 Jun 1;178:108974.
[33] Marini F, Walczak B. Particle swarm optimization (PSO). A tutorial. Chemometrics and Intelligent Laboratory Systems. 2015 Dec 15;149:153-65.
[34] Hussain S, Matin AW, Islam O. Genetic algorithm for energy efficient clusters in wireless sensor networks. InFourth International Conference on Information Technology (ITNG'07) 2007 Apr 2 (pp. 147-154). IEEE.
[35] Kulkarni PH, Malathi P. PFuzzyACO: fuzzy-based optimization approach for energy-aware cluster head selection in WSN. Journal of Internet Technology. 2019 Nov 1;20(6):1787-800.
[36] Kulkarni PK, Malathi Jesudason P. Multipath data transmission in WSN using exponential cat swarm and fuzzy optimisation. IET Communications. 2019 Jul;13(11):1685-95.