Turkish Journal of Engineering

https://dergipark.org.tr/en/pub/tuje

e-ISSN 2587-1366

Scalable Multi-Clustering Aggregation Scheme in WSN Using Machine Learning

Ramesh Dennison [*1], **Thirassama Jaya**[2]

[1]*Research Scholar, Anna University, Information and Communication Engineering, India, ramesh4773@gmail.com*

[2] *Professor, Saveetha Engineering College, Anna University, Electronics and Communication Engineering, India, jayacsiramesh@gmail.com*

Cite this study: Ramesh D., & Dr. Jaya T. (2025). Scalable Multi-Clustering Aggregation Scheme in WSN Using Machine Learning, Turkish Journal of Engineering, 9(1), 103-115.

https://doi.org/10.31127/ tuje.1488192

Keywords
Data Aggregation,
Machine Learning,
Energy Efficient,
Network life span,
Clustering,
reliability

Research Article

Received:23.05.2024
Revised:26.06.2024
Accepted:27.06.2024
Published:20,01,2025

Abstract

Due to Limited resource constraints in WSN data packets collide while being routed to sink, redundant data can be eliminated by data aggregation, which minimizes overall amount of data transferred and increases network's lifespan. Minimizing energy consumption and boost data aggregation rate is most crucial factor in WSN. A Scalable Multi-Clustering Aggregation utilizing Machine Learning (SMCA-ML) focuses on data aggregation approach of heterogeneous wireless sensor networks, using neurons as wireless sensor network nodes in a machine learning method. Machine Learning method accumulates the captured data collected by senor nodes and integrates the accumulated data with multi-clustering route. Threshold value of hidden layer and weight of input layer are randomly generated by the proposed method prior to training. This results in an unstable output that affects the efficiency of data aggregation and causes a long delay. More crucially, distinct threshold settings were made in accordance with the features of uneven energy consumption in wireless sensor networks (WSNs) allow data packets more swiftly by setting smaller threshold in far sink with enough energy. To maximize data aggregation, close sink area with tight energy employs a greater threshold. This way, the combination can result in high data fusion, efficient energy consumption, and little delay. The results of simulation suggested that SMCA-ML based data aggregation algorithm can significantly extend lifespan of network, substantially decrease energy consumption, enhance network energy, expand network performance and improve data aggregation efficiency when compared to conventional Stable Election Protocol (SEP), Back Propagation algorithm, and Extreme Learning Machine.

1. Introduction

With the development of 5G technology and recent advancements in wireless communications, anything may now be connected to a network. This change has led to the utilization of multimodal wireless sensor networks for a wide range of applications and environments in monitoring different activities. The amount of data that the network must analyse, store, and transfer is increasing dramatically due to the rapid proliferation of networked devices and the rise of intelligent transportation. Sensor nodes capture real time information and report back to the base station with accurate outcomes as additional sensors are installed. Data aggregation and energy consumption are issues associated with data transmission due to the rapidly growing amount of data and the limited bandwidth resources available. Consequently, in order to increase the precision of data monitoring and energy usage, sensor nodes are typically placed in large quantities. On the other hand, the densely deployed nodes may lead to numerous nodes detecting the same abnormality at the same time and producing a lot of data. There are some redundancies in data gathered by many sensors; in fact, some sensor nodes may also detect false information, which would reduce the effectiveness of data aggregation. The essential data must be processed and transferred, which is costly and requires a lot of bandwidth, energy, and time. Transmission delay and quality of service are negatively impacted by congestion brought on by a large volume of data being transmitted from sensor nodes to the base station [1]. Massive amounts of data are sent at top speed to the base station in order to find critical information in an instantaneous setting. Therefore, in order to lower the processing load on the cluster head and data redundancies at clustered level of WSN, an efficient and reliable structure for processing local data needs to be established [2].

Since the arrangement of nodes in heterogeneous network is somewhat random and has a high density of elements, the data that is gathered has redundancy, complementarity, real-time, and uniqueness. It is a perfect and efficient approach to decrease the quantity of data transmitted without expanding the infrastructure in order to deal with the rapid development of data. The quantity of data transferred over a network can be reduced via data integration.

This revolutionary technology is gaining traction and spreading over all facets related to environmental processing and surveillance. These systems primary benefit is their ability to be deployed in dangerous and isolated areas, giving users adaptable organizational options and easing data access. A wireless sensor network (WSN) comprises of abundant amount of sensor nodes that are widely dispersed in typically unreachable locations and create data transmission networks. Their primary responsibilities include process observation, data collection, and transmission to a base station for processing [3]. The concept of sensor networks which rely on the cooperation of numerous sensors has been brought about by sensors. Recent years have seen a rapid development of sensor networks, and the deployment of these networks offers new applications a cutting edge. The amount of data gathered and processed increased as a result of the widespread use of WSN apps and the variety of the fields concerned. Recent years have seen a rapid development of sensor networks, and the deployment of these networks offers new applications a cutting edge. The amount of data gathered and processed increased as a result of the widespread use of WSN apps and the variety of the fields concerned.

The activities of comprehensive data gathering, data propagation, filter redundancies, association among other data, evaluation, and extraction of data provided by numerous resources are one of the definition of data aggregation technology that is now more widely acknowledged. Furthermore, the data Aggregation approach integrates time and geographical dependence, eliminates useless and incorrect information, and retains significant and legitimate components in order to create a group of multivalent cohesive and precise data descriptions of occurrences. Data aggregations primary functions consist of reductions in energy use, enhancing data security, minimizing transmission latency, optimizing network resources, and lowering processing overhead. Data aggregation technique reduce the amount of energy used by the network while simultaneously guaranteeing that nodes use energy in a fair manner, avoiding energy holes, and enhancing lifespan of individual node which improve the efficiency of the network. When using heterogeneous sensor networks in practice, data aggregation technology aggregates the processed data from multiple sensor nodes to provide information to base station [4-6]. This method's primary benefit is that it uses less energy and bandwidth for communication. This is because lot of redundancies in gathered information. Amount of data transmitted to the base station can be minimized by eliminating unnecessary data leads to efficient data transmission. Information integration techniques help minimize energy use, improve data security, minimize transmission delays, and optimize network resources which leads to the improvement of overall network performance.

Additionally, it will make information gathering more efficient. Without information aggregation technology, scheduling data from various sensor nodes will be more challenging for the link layer, and collisions will occur more frequently. These factors will lower communication efficiency and interfere with timely information collecting. In order to reduce network traffic, WSNs use data aggregation. This involves an intermediary sensor node, also known as Cluster Head Node, aggregating the data it receives from its cluster nodes rather than simply transmitting it all at once to its parent node. This strategy has resulted in the creation of a number of data aggregation systems that maximize sensor node energy consumption by clustering and carefully choosing a Cluster Head (CH) to handle aggregation of data to enhance network lifespan [7-9]. The goal of data integration method is to maximize these data properties, eliminate any extraneous information, and improve mutual assistance, credibility, and fault tolerance. The main aim of this research work is to improve the data aggregation, thereby increasing the energy efficiency of the network.

2. Problem Statement

Sensors in operational network zone typically have an elevated number of nodes and an unequal distribution due to small size and resource constraints, which leads to nodes in operational region capturing duplicate data. Energy consumption of sensor nodes is too high if data from source node is transmitted directly to the Sink node, which drastically reduces networks lifespan. To evade aforementioned problems, sensor network use data aggregation technologies all along the data acquisition technique. With this technique, network communication capacity is preserved while data transmission throughout the network is reduced and node energy consumption is minimized.

The ultimate purpose of information aggregation is to increase system efficacy by utilizing multi-sensor cooperative operation. The neural network and wireless sensor network (WSN) models are comparable. The neurons in a neural network are related to nodes in wireless sensor networks (WSNs) that collect information about their surroundings. Similar to how a neural network must realize information transfer through synapses using protocols. The entire purpose of a wireless sensor network is to process copious amounts of data and derive the data's characteristics essentially, the same function as a neural network data fusion system. As a result, wireless sensor network data fusion can be done using neural networks. A neural network model is built for every wireless sensor network cluster. Cluster head gathers the information from the cluster member node and transmit the aggregated information to the sink node. Reduced amount of data transmission in network leads to increased network lifespan, minimize energy consumption during data transmission and enhances data transmission efficiency to a greater extent. The goal of a wireless sensor network is to process a lot of data in order to determine its properties, which are the same as those of a neural network data fusion function.

As a result, information integration from wireless sensor networks can be done using neural networks. A neural network model is built for every wireless sensor network cluster. By combining the data from the cluster members, the cluster head can extract the feature vector from the data in the cluster and send it to the aggregation node. This decreases the amount of data transmitted during communication, increases data transmission efficiency, uses less energy during communication, and extends the life of the network.

This paper proposes a new machine learning-based SMCA algorithm for information aggregation in heterogeneous sensor networks. The primary contributions of our work in this research, when compared to the existing general selection procedures, can be summed up as follows: Demonstrate the problems with a data aggregation approach for heterogeneous sensor networks and state the data aggregation algorithms problem. Describe a novel machine learning-based SMCA algorithm to improve the data integration technique and energy efficiency. To illustrate the application and effectiveness of the suggested data aggregation algorithm, present comprehensive simulation results. Compare the suggested algorithms performance to that of alternative data aggregation algorithms.

3. Related Work

Attacks on wireless sensor networks are frequent because of wireless transmission. Consequently, it is extremely data security in the data aggregation is crucial procedure [10]. In order to aggregate information, a number of aggregation algorithms in the contributions that are currently available group sensor nodes based on the raw data. But anomalous data frequently shows up in unprocessed data. Therefore, there is no doubt that the data instability affects how well these methods work. Furthermore, the adaptability of these functions is restricted because multiple aggregation functions are defined for a specific type of network characteristic (such as a grid network) or a particular type of data (such as temperature data). We are hence driven to suggest property-independent and data-independent aggregation techniques. There are two main categories of responses produced by data aggregation algorithms: query based and event based. Three components make up the query based data aggregation algorithm: propagation of data, relevance aggregation, and clustering [11]. The data aggregation mechanism's responsiveness can occasionally lag significantly. The majority of research being done now focuses on responsive aggregation algorithms. Three categories cluster hierarchy, tree-based aggregation structure, and planar based aggregation structure define the data aggregation path depending on response. All sensor nodes are initially clustered via the cluster-based aggregation technique [12]. In a hierarchical clustering grid, the clustering data aggregation approach is typically applied. The sensor network as a whole is divided into several sections by the hierarchical method of data aggregation. After acquiring monitoring information from the outside world, other nodes will transmit event detected to cluster heads, which initially elected as cluster head in each zone. Gateway node receives data that cluster heads have synthesized from other nodes. Cluster members transmit data packets detected from an event to cluster heads, cluster members elect a cluster head to handle packet transmission.

IBRE-LEACH [13] considers residual energy when choosing CHs, preventing the assignment of CH responsibilities to nodes with low energy levels. A fixed threshold limits the number of members when clusters form. Overstepping this limit prevents a node from joining its closest cluster, which could result in it becoming an abandoned node (AN). To further lessen BS overload and energy consumption, the node that has the most energy left and is closest to the BS is named the root. Next, each CH and AN creates an ascending routing table with the distances to other CHs, ANs, the root, and the BS. In addition to being in charge of data transmission and routing, cluster heads consume more energy when compared to other nodes. As a result, two most important aspects of cluster based data aggregation approach are selection of cluster head and routine topology maintenance.

Consumption of sensor energy is minimized based on data aggregation method considered, in [14], researcher developed an enhanced method of clustering nodes in sensor network to improve efficiency of network. In a wireless multi-sensor system, Jan et al. [15] introduced a unique method for accumulating and cluster component selection using a hybrid algorithm. The suggested strategy effectively decreased the number of blind broadcast messages while simultaneously reducing the signal congestion as a result of establishing clusters. The raw data is frequently the basis for temporal and spatial correlations, according to earlier research [16-18]. Data acquired from contiguous sensor nodes exhibits spatial correlation, but data received at different times instants for a given sensor node exhibits temporal correlation. In-network data aggregation reduces degree of correlation for sensor networks, in [18] proposed a cluster based hierarchical network to perform aggregation while removing redundant information from detected data in an energy proficient manner. A Novel Mobile Target Detection Algorithm (NMTDA) based on information theory was suggested by Zhang et al. [19] to create local and global data aggregation rules using the Bayesian approach and Neyman-Pearson lemma, based on the likelihood of ratio test statistics. Gigi et.al [20] By merging IoT sensors with an appropriate clustering technique, sensor networks' energy performance can be improved. However, clustering necessitates more work, such figuring out the cluster head and cluster creation. The study methodology (E2CRN) presents a novel technique for ecological energy responsive cluster for IoT-MWSNs: Environmental Energy Attentive Clustering with Remote Nodes. In E2ACRN, weight is the only factor that determines the cluster head (CH). The weight is computed using the local average energy of all IoT sensors in the cluster and the residual energy of each IoT sensor. Unsuitable clustering approaches with improper planning could cause nodes to be too far apart from CH. These far-off nodes use more energy to interact with sink. Rajesh et.al [21] present fuzzy assumption algorithm to determine Outstanding-Cluster Head. Cluster Heads (CHs) can send messages to the base station and selected on an alternating basis based on probability threshold costs. Under the suggested FCHIDS approach.

RiPPAS accumulate secrecy to transmission of information through the use of false identities, which provide each node with a distinctive key and set of pennames, security is increased. The uniqueness of designating the group of pennames to each node at the time of deployment determines the network's reliability, as the BS uses the received penname transmitted with the cipher text to identify the decryption key. Transmitting the data without a source ID ensures anonymity with minimal transmission overhead. The intruder is unable to perform long-term analysis to extract valuable information about the penname, which is required to obtain the key, due to the randomness of parent selection and the mystery surrounding the sender ID. A complicated incremental PH-based data aggregation (LIPDA) for the hierarchical cluster topology WSN. Each cluster receives a unique pair of keys using RC4, which offers a risk of key leakage by opening the door for node compromise attacks. One key in the pair is used to create the imaginary component, which represents the data as a complex number. The complex number that has been formed is encrypted using a different key. The hypothetical portion aids in confirming the accuracy of data exchange at the BS.

After data acquisition, sensor node uses reversible multicast fusion tree and multi-hop mode to transmit data to sink node. With data gathered, each intermediary node in this architecture processes data fusion. First of all, it creates multicast trees among cluster heads, elects cluster leaders, and automatically organizes a big number of clusters [22]. In this manner, data is delivered from source node to cluster heads and process information. Data received by cluster heads are then transmitted to destination by reverse multicast tree aggregation. Certain works [16,19,23,24] combine cluster and tree-based aggregation. In response to novelty of information in Internet of Things, author of [19] proposed a data fusion clustertree structure approach based on incident driven (DFCTA) and data integration architecture of integration tree comprised of aggregating nodes. Fog computing environment empowered a data aggregation ensemble strategy to deal medicinal data gathered from BSNs [25]. The author achieved 98% accuracy when the number of estimators was 40, tree complexity was equivalent to 15, and 8 features were taken into consideration for the prediction task. Aydın etl al. [26] proposes a CNN based data classification model to aggregate data in order to increase data accuracy collected from sensors.

Information detection rates based on tree data aggregation was proposed by Lin et al. [27] in consideration of the data's correlation and redundancy, which enhanced the data generation rate in wireless networks. The author of [17] proposed a hierarchical routing scheme-based routing model called energy efficient clustering (ENEFC) in order to extend lifespan of networks while improving energy efficiency at the node level. Analysis and results from experiments demonstrate that ENEFC scheme outperform others in the area of energy efficiency. WSNs using the Adaptive Data Aggregation (ADA) approach. With consideration to packet size and data amount within a cluster, this technique aims to minimize the aggregated data throughout execution time. This procedure minimizes payload size for varying cluster sizes; the overhead of this method is computation at each node. Data Aggregation Technique based on tree searching algorithm in Wireless Sensor Networks (WSNs) use the Fitting Function to reduce redundant data by assigning a predetermined threshold value [28]. Subsequently, the combined data was placed into a transmission equation. As a result, less energy is used for transmission, extending the battery life of the sensor node. Because it requires extra work, fitting aggregated data to the equation has an overhead.

IoT applications that prioritize trust derivation can benefit from an energy efficient, trust-aware approach in WSNs that use a game theoretical strategy. An ideal number of recommendations that meet the network's security requirements are found using this risk strategy approach. Trust Derivation Dilemma Game (TDDG) is launched with the purpose of acquiring trust based on these suggestions. Both the network's latency and energy consumption decrease as a result. Despite the computational burden, the protocol determines the trust node by sending a trust request packet each time. Using mixed integer programming, data gathering trees are evaluated taking into account three types of aggregator node information: full, non-aggregator, and hybrid aggregators. The data was directed to the sink node by the star data aggregation tree, as demonstrated by the author. As a result, every node can speak with the sink node directly. Consequently, less energy is used for data relay, extending the life of the network. Because the aggregator node is validated using the threshold level computation, the overhead rises. In WSNs, the effective secure in-network data aggregation mechanism is called Trusted Secure Processing Tree (TSPT). Data is gathered by this protocol using a hierarchical cluster design, with the cluster leader acting as an aggregator. Based on the reputation value that is kept up to date in the popularity tables of every one of the sensor nodes and cluster heads, the reliable aggregator is recognized.

Based on the Euclidean distance between CH and the deployed node and leftover energy, the EDIT protocol determines the aggregator node. Two forms of distance are taken into account when calculating the routing delay or energy level compromise: hop-count and Euclidean distance. As a result, it shortens latency and extends network lifetime. All intermediate nodes create link from the source to destination, source-based tree aggregation technique in suggested intermediary nodes require highest possible energy. This process is known as energy-aware tree formation. The author of [29] presented a novel Cluster Tree Routing System Data Gathering (CTRS-DG), which performs better than baseline techniques already in use in terms of network lifetime and stability period. Source nodes in a form of tree topology, with intermediary nodes at intersection branches of tree doing data aggregation. Reducing energy usage by improving a data aggregation tree's structure is one of the key features of tree protocols. Adaptive Application Independent Data Aggregation (AIDA) illustrated by He et al. [30], an intermediary layer operating between data link layer (MAC) and network layer makes all aggregation decisions without regard to the application. Because it adapts varying network activities in an appropriate way and no need to rebuild the application specific aggregation logic. This suggested aggregation layer is composed of two units. The first is an aggregation function unit, which chooses packet formats from broadcast, anycast, multicast or unicast after aggregating the packets in an aggregation pool. Adaptive Energy aware Data aggregation Tree (AEDT) approach consider the node having highest energy as aggregator node. The aggregator and sink nodes are the only ones that enter the sleep mode, which lowers energy usage. When traffic reaches a certain threshold, packets are chosen based on the parent capacity. An overloaded notification is sent to the network if the expected traffic load exceeds the parent node's transmission capacity. All of the paths that have been found are also kept in a memory table. The suggested tree is updated and, if necessary, a new aggregator node is elected after time t seconds.

In a planar topology, the traditional flooding fusion approach involves the source node simultaneously transmitting data to the destination through a neighbouring node. Numerous copies of packets are transmitted to intermediate node when utilizing the flooded routing strategy. By using nodes data packet mechanisms for caching, duplicate data from implosion are eliminated. However, it continues to support certain algorithms due to its ease of use, efficacy, and high resilience. Planar routing makes it difficult to build an efficient filtering of information mechanism and makes it tough to set filter points. A robust information aggregation technique centred on geographical and temporal correlations for sensor networks are recommended to enhance network aggregating proficiency and resilience over disruption from noise, as well as to assess the level of data destruction in the face of additive attacks. Machine learning technology gives data aggregation a new direction and increases its availability and intelligence. In order to increase network performance, Wang et al. [31] introduced a smart Data Gathering Schema-Data Fusion (DGS-DF) to conduct data aggregation with neural network. The suggested technique could effectively preserve energy and lengthen the network's lifespan. Because serial structure free techniques perform better small and medium range networks, fundamental route establishment algorithms can be made better by minimizing the number of communications required and decreasing the visiting path even further.

But the veracity of the data is not taken into consideration by the aforementioned algorithms. False data from sensors may be generated by the sensing node in the real environment. For instance, the sensor node may detect false data due to issues like hardware failure or external noise [32]. In this research, we present a unique SMCA approach for heterogeneous sensor networks utilizes machine learning to examine network energy consumption and operation under various data aggregation strategies, based on a summary of prior studies. The network is first divided into a number of segments by the data aggregation method, and each segment has a distinct monitoring weight. The suggested technique collects original data attributes from sensors for information aggregation. Enhanced machine learning learns, optimize the weight and threshold of neural network. Subsequently, Sink receives the combined data from the cluster head. Intra-network data aggregation reduces the propagation of superfluous data and boosts data aggregation accuracy in the WSNs data aggregation process.

4. Methodology

In contrast to the conventional neural network theory previously discussed, every neural network parameters needs to be adjusted. A single hidden layer is applied in SMCA in conjunction with feed forward neural network and machine learning. The fact that there is just one concealed node layer is its primary characteristic. Machine learning approach has a very fast training speed, a very swift convergence speed, and an improved capacity for generalization and learning speed due to its random selection of the number of hidden layer nodes. Computation of network output weight, threshold are performed monopoly by a hidden layer network. As a result, learning speed is boosted by numerous times in contrast to Back Propagation neural networks, Radial Basis Function neural networks, and Support Vector Machine intelligent approaches [33]. Machine learning techniques are broadly used in image classification, face recognition, authenticity applications, and health diagnostics has been greatly expedited by this achievement. [34]. The SMCA algorithm is appropriate for forecasting short-term loads and can produce strong prediction results with less training samples. Prior to learning, the SMCA algorithm produces invisible layer criteria and input layer weights at random.

The following benefits of using an ML algorithm over traditional neural networks: Short momentary learning. The algorithm is easy to implement, and the parameter setting is minimal. The capacity for generalization is improved. The machine learning's invisible layer mapping mechanism goes beyond the conventional BP network gradient approach. Alterations in mapping function will update results without affecting the algorithm itself. ML approach create weights of input layers and thresholds related to hidden layer at random fashion, whereas step down method train classic neural networks. Training process does not require iterative modifications, and it proceeds quickly. In this paper, we propose a new machine learning-based approach for mobile sensor network data integration. Our approach addresses the issue that machine learning randomly generates the input layer weight and the hidden layer threshold prior to training, resulting in unstable output that affects information integration efficiency and causes long delays.

A new approach to sensor network data aggregation based on SMCA - Machine Learning is proposed. With the above technique, a challenge of neural networks generating input layer weight and hidden layer threshold at random before learning is addressed. Randomly generated input layer weight and threshold generates unstable output, which reduces data aggregation efficiency, network lifespan, consume more amount of energy and causes a considerable delay. The suggested algorithm effectively increases the network coverage area, expands the network capacity, reduces network consumption, balances the distribution of energy consumption, and extends the network's lifetime by taking into account the sensor node's remaining energy, the data transmission distance, the energy balance, and other factors. The suggested algorithm can increase network performance and efficiency in addition to being universal and adjustable.

5. SMCA Algorithm

Our proposed technique consists of initial parameters such as energy, nodeID, individual keys for data sharing and localization parameters. Sensor nodes are deployed in a Euclidian distance plane. Sensing nodes, or nodes operating in the sensing mode, interact with other nodes in the network and perceive events before sending data to the sink node. Relay nodes, or nodes operating in relay mode, exclusively transmit data from perceiving nodes to the sink. Traditional Neural network uses multiple hidden layers and parameters are adjusted to its convenient. In sensing models, certain studies make assumptions about multiple range of sensing features. Since signal intensity (such as that of radios, acoustics, and seismic waves) diminishes with increasing distance, the practical dependence of sensing precision on destination distance is observed. Researchers made hypothesis that the multimodal sensing model applies, in which detection efficiency solely hinges on distance between two nodes is within threshold. The Euclidean distance between source and destination point is given by

$$D(s,d) = \sqrt{\sum_{i=1}^{n}(s_i - d_i)^2} \tag{1}$$

where the node in the interest region is denoted by d_i and the source sensing node is represented by s_i. The multimodal sensing model's detection probability is shown as $\omega_i(j)$ following the distance calculation $D(s,d)$contained within detection range r. Interest zone is certainly detected by source sensing node s_i if D(s,d) is within the threshold value of r.

$$\omega_s(d) = \begin{cases} 1, & if\ D(s,d) < r \\ 0, & otherwise \end{cases} \tag{2}$$

Sensor nodes maintain a geographical region that is observed or covered by sensor nodes without any gaps or duplication in sensor node lifespan, battery life, and maintenance schedules to ensure persistent and accurate monitoring over time, in accordance with the spatiotemporal correlation in the process of information gathering and event recognition. Equation-(3) represents the spatial area that sensor nodes observe or cover, battery life, and connectivity between the sensor nodes to ensure reliable and consistent surveillance over time.

$$f(N_s, N_c) = \sum_{i \in N_s}((E_s + E_c) * \eta_i) + \sum_{i \in N_c}(E_c * \eta_i) \tag{3}$$

Where N_s represents sensing nodes, N_c cluster heads in data transmission session and η_i weight of input nodes. E_s is the energy consumed by sensing node in event identification, E_c energy cost of cluster head for a packet transmission from one cluster head to another cluster head. The weighted sum of energies is our spatial-temporal coverage function is instead of an energy sum. Energy has an exponential increase in value as the quantity of residual energy declines with two variables E_i and z. E_i is the residual energy of sensor nodes and z is uniform in (0,1) and is represented as

$$\eta_i = z^{E_i} \tag{4}$$

The fact that there is only one concealed node layer is its most notable characteristic. Subsequently, universal feedforward neural network in single hidden layer is implemented using this topology. As a result of SMCA arbitrarily selecting the total number of nodes in hidden layer neural network and computations of output weights, the threshold for the neural network in a hierarchical process. Machine learning method trains quickly, improving the machine learning network's generalization capacity as well as its learning speed. Consequently, learning speed is multiplied by thousand when compared to existing intelligence methods for Support Vector Machines, Radial Bias Function neural networks, and Back Propagation neural networks.

SMCA algorithm along machine Learning generates precise estimation result with respect to training samples for predictions. Consider there are n sensor nodes then the activation function f(x) can be calculated as

$$f(x) = \sum_{i=1}^{N}(x_i * w_j + b_i) \qquad j = 1 \ldots\ldots n \tag{5}$$

Where N is number of hidden layer nodes, x_i input layer weight w_j input matrix b_i input threshold.

The output Φ_i of the network with a set of N input layers is given by

$$\Phi_i = \Psi_i f(x) \tag{6}$$

$$\Phi_i = \sum_{i=1}^{N} f\left(x_i * w_j + b_i\right) \Psi_i \quad (7)$$

Where Ψ_i is interconnection between output and hidden layer

In comparison to conventional neural networks SMCA-Machine Learning has mapping functions with respect to hidden layers is greater than conventional BP gradient method. It also has a shorter learning time, a compact parameter setting, an easy-to-implement algorithm, and improved generalization ability. High prediction accuracy, excellent adaptability, and superior generalization performance are benefits of the SMCA machine learning approach.

A unique search algorithm with specific benefits and features in various areas has been proposed. Different kinds of sensor nodes release signals with distinct properties at different frequencies, and even the same kind of sensor node might release signals with distinct characteristics depending on the event detection strategy used. Moreover, the bandwidth of their signals varies depending on the type, frequently being increased by adding extra harmonics. Suppose the position of the sensor nodes with time t is P_t and movement frequency of sensor node is denoted by S_t then the position P_{t+1} is given in Equation (9) and updated movement frequency position of sensor in given in Equation (8)

$$S_{t+1} = \omega S_t + (c - P_t) w_f \quad (8)$$

$$P_{t+1} = \rho P_t + \lambda S_{t+1} \quad (9)$$

Where λ, ρ, are weight coefficients with a value 1 the updated position of the sensors is given by

$$P_{t+1} = P_t + \omega S_t + (c - P_t) w_f \quad (10)$$

When the current optimal position is reached by the sensors then the equation is given by

$$Q_{t+1} = R Q_t + W c \quad (11)$$

Matrix R governs the behaviour of this dynamical system. Evidently, if the algorithm converges and the iterations continue, the entire sensor density should move towards c (P_t = c as t = ∞) resulting in zero speed (S_t = 0 as t = ∞). Where c is random vector between –1 and 1

The proposed approach has frequency tuning in comparison to other searching techniques, and it can dynamically manage the correlation between global and local search by proactively shifting from global search to local search when the requirement is met. The algorithm's features include a straightforward model, quick integration, parallel processing, ease of use, and adaptability in parallel. The SMCA machine learning algorithm uses the least norm square approach to generate the output weights and randomized selection of hidden layer variables to increase learning speed. Even though the gradient descent algorithm's shortcomings are partially mitigated by these machine learning features, number of nodes in hidden layer is pre-allocated, setting parameters selected at random stay constant throughout the training process, a large number of non-optimized nodes persist. The process of minimising cost function is not greatly aided by these nodes. An excessive number of nodes in hidden layer complicates network but also lowers capacity of algorithm for generalization. This research proposes the SMCA Machine Learning technique for minimum threshold and weight for heterogeneous wireless sensor networks' data aggregation strategy. Our proposal in this paper is to maximize efficiency of information aggregation problem on heterogeneous sensor networks via SMCA machine learning technique. The suggested algorithm takes into account all relevant factors, including the sensor node's residual energy, distance of data to be transmitted, energy equilibrium and other factors, As a result lifespan of the network, its coverage area and capacity are effectively increased. Additionally, networks energy consumption is reduced and the overall distribution of energy consumption is balanced.

SMCA data aggregation algorithm is predicated on sensor network SEP clustering. All the nodes in event zone are first clustered by traditional clustered SEP routing protocol assigns a value between 0 and 1. The sensor node becomes cluster heads and members together create a steady state cluster arrangement if value is less than a predetermined threshold [35]. The data is pre-processed by cluster members before being sent to cluster head. Aggregated data from cluster heads eliminates superfluous and unnecessary details before transmitting to sink. SMCA data aggregation methodology combines SEP clustering routing protocol with an enhanced machine learning technique. Every iteration of cluster protocol in sensor network requires reconstructing the cluster and altering its topology, which raises energy overhead of network while creating new clusters. Taking this into account, the suggested technique clusters sensor nodes maintain the same topology. Node having highest energy is chosen as cluster head whereas cluster head is interchanged, which can lower energy consumption during clustering.

5.1 Steps of SMCA Algorithm

The data aggregation process using SMCA Algorithm is as follows. Initially, set up every element of heterogeneous sensor network ascertain current state of network common, routing, and cluster heads. Node initialization Phase, we provide a novel idea termed sensor couple, wherein one sensor node locates nodes to be event detecting nodes, while the other is responsible for transmitting data from source to sink. Position vector in each dimension are either 0 or 1 in the case of the typical binary SMCA. As a result, this SMCA can distinguish between the sensing node and cluster head states of each node. Because there are an increasing number of states for each sensor node, it becomes more challenging for the typical binary SMCA to represent all of the nodes' states. As a result, for the two types of sensors, all vectors pertaining to sensor information, such as position and frequency, should be extended. SMCA predominantly makes use of probabilistic sensing model. Taking into account the ranges of each sensor, nodes may choose to designate cluster heads and cluster members. Position can be thought of as the set of accessibility to each sensor node since each sensor's position is represented by an N-dimensional vector, where N is the number of sensor nodes in the network. Nodes associated elements with a value of 1 in the vectors nodes are allocated responsibilities of sensing node and cluster head node, respectively, after determining the optimal location of sensors using Equation (3).

Every sensor pair's frequency and position are initialized using our suggested method. This is due to the fact that a greedy initialization technique leads the algorithm to converge optimal solution more quickly than a random one. The sensor regions are divided into protected and exposed zones based on the coverage range. One sensor node is chosen at random for each exposed area, and it is then activated. The associated element in the position is then set to 1. By turning on a new sensor node in the preceding stage, those whose detection probability in the exposed region over the threshold value are eliminated. In the clustering phase the monitoring region is constructed based on the position of the sensing node. A random selection of cluster heads is made inside a cluster; it gathers all data from the cluster members. Following the stabilization of the WSN clustering, cluster heads transmit data about cluster members to sink node and creates a table for path discovery from source to sink. After network has reached equilibrium, WSNs data aggregation technique is first trained to define count of sensors, network output weights, and threshold properties of concealed learning. Because of their energy constraints, WSNs common nodes handle sensing as well as sending and receiving duties. The data aggregation model is trained using SMCA - Machine learning in sink node of heterogeneous sensor network to minimize energy consumption of typical nodes and enhance the lifespan of network. After training, sink node obtains pertinent information from the relevant data aggregation model and builds the network structure based on the information received.

Sink node notifies relevant cluster head of data aggregation model network characteristics i.e number of nodes in hidden layer, output of network weight, and threshold. Trained data aggregation model is used by cluster head to combine data transmitted by cluster members, extract information, eliminate redundancies, avoid transmission of unnecessary information, subsequently store combined data, and transmit to sink node.

Implementation Steps for SMCA – ML data aggregation method is given below.

Step 1: Frequency and speed of the sensor nodes are initialized.

Step 2: Initialize the population of the sensor nodes and the parameters such as security, location of sensors, number of hidden layers in machine learning and initialization of network.

Step 3: The data aggregation model's input weights and maximum threshold limits are set initially.

Step 4: The estimated error and output weight matrix for every sensor node is calculated.

Step 5: Split the Sensor nodes into sensing node and relay nodes i.e. cluster heads

Step 6: In order to differentiate sensing and cluster head node, node in network region increases the frequency of communication to find best fit of cluster heads.

Step 7: Repeat step 6 until the best fit solution is reached.

Step 8: Find the best solution among sensor in order to decide cluster head from cluster members and mark the region as covered region.

Step 9: Cluster Head transmits cluster member details to base station.

Step10: Base station establishes data aggregation methodology and train data sample to cluster nodes to acquire network parameters using machine Learning.

Step 11: Sensor Nodes Continue the execution while the termination condition of training data set has not yet reached.

Step 12: Cluster heads perform data aggregation on received data from cluster member nodes using SMCA – ML method. Cluster head transmits aggregated data to sink.

Step 13: Once the data transmission is over repeat the steps from 1 to 12 for the next session.

6. Results and Discussion

The base station (BS) and sensor nodes that comprise up the suggested network are dispersed randomly and have equal access to power resources, and processing capacity. These nodes gather event-related information and send it to BS. When it comes to power and resources, the BS can handle more than other sensor nodes in the network. Simulations has been performed to evaluate the suggested method's performance against others. Two types of comparison approaches were created. The first group assumed that all sensor nodes had enough communication range to send their data straight to the sink node, and they merely took into account the target coverage. Compared the computational complexity and energy consumption of our approach. Conversely, the second group considered the data aggregation rate and network life time. Simulation's primary goals are to track network lifespan, minimize energy usage, and enable safe data aggregation and communication. This method's extensive experimental evaluation guarantees that the discriminator network can reliably safeguard the network against redundant data. It increases network energy efficiency without sacrificing network latency.

For both methods, the NS2 simulator is utilized with comparable settings. The network topology measures 600X600 m^2. There are 300 sensor network, and their transmission range is 40 meters. The nodes' starting energy is fixed at one joule and then base station is set to 50 J. The sensor nodes' maximum energy usage is set to 50 nj/bit for data reception (Rx) and 10 pj/bit for data transmission (Tx). The simulation can last up to 30 minutes at most. The hidden layer node count of the chosen network starts at 10 and cycle count is incremented by 20 cycles to attain a specified value of 300. This value assigned to limit learning machine attribute. The hardlim function is chosen for learning by the SMCA-ML algorithm's implicit layer activation function.

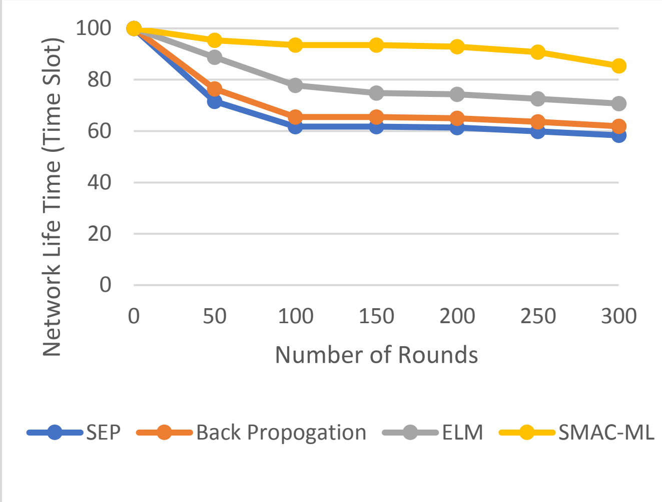

Figure 1 Network Lifetime according to Various Existing Algorithms

The findings presented in this paper are based on an average of 100 investigations, which helps to strengthen the model's conclusions.

One of the key metrics that represents the network's performance is its life span. As simulated event rises, proposed algorithm increasingly demonstrates its superiority. Consequently, prolonged network lifespan, higher data aggregation efficiency. Figure 1 displays the network life span of the four algorithms using the sensor network's network life span as a measure. Figure 1 illustrates that the SMCA-ML data aggregation approach suggested in this paper has a network lifetime per polling network that is significantly higher than that of the SEP techniques, as well as higher than that of methods based on back propagation neural network technique and ELM approach. Simultaneously, network lifespan per polling is rapidly reducing as number of round rises. When the number of rounds increases the network life span of SMCA – ML algorithm is about 90 % whereas the existing techniques SEP is about 58%, Back Propagation neural network is 61% and Extreme Learning Machine is about 71%

The sensor network's utilization of energy, reliability, and connectivity effectiveness are significantly influenced by quantity of cluster heads. A portion of network's performance can be estimated from ideal number of cluster heads and periodic assessment index of network. Typically, proportion of cluster heads to total sensor nodes is kept between 5% and 12%. Compared to previous schemes, this range has less changes to wireless sensor network and an increased amount of cluster heads.

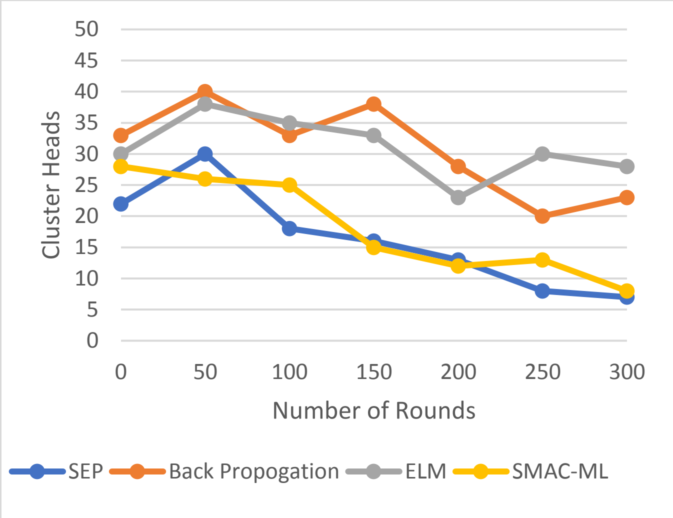

Figure 2 Comparison of Cluster Heads with Existing Approaches

Figure 2 compares availability of network cluster heads for each of four techniques.

The SMCA-ML technique presented in this paper has between 8 and 28 cluster head nodes, while the SEP protocol has between 7 and 22 cluster heads. The quantity of nodes varies drastically, after 250 cycles cluster head count falls below 10. In BP neural network, cluster head nodes vary greatly, ranging from 20 to 33. However, estimated amount of cluster heads is quite near relative to this paper's approach. In the ELM algorithm, estimated amount of cluster heads varies from 30 to 23. Figure 2 illustrates range variations gets comparatively larger as iteration increases. Both ideal number of cluster heads and their reliability are increased. It is evident that the suggested algorithm works the most effectively.

Figure 3 illustrates average utilization of energy of cluster head node decreases as iteration increases. Notably, SEP approach exhibits highest reduction, of almost 85%. This paper's suggested SMCA-ML approach has the smallest decrease. In addition, the ELM algorithm uses less energy than the BP neural network, cluster heads of SEP approach uses more energy than the other algorithms, and the algorithm suggested in this research uses the least energy overall. The SMCA - ML algorithm uses 85%, 80%, and 75.6% less energy than the cluster heads of SEP algorithm, Back Propagation neural network algorithm, and ELM methodology, respectively. Due to lack of data aggregation and cluster head node's sharp decline in energy consumption as iteration increases, SEP algorithm uses a lot more energy than the other three algorithms combined.

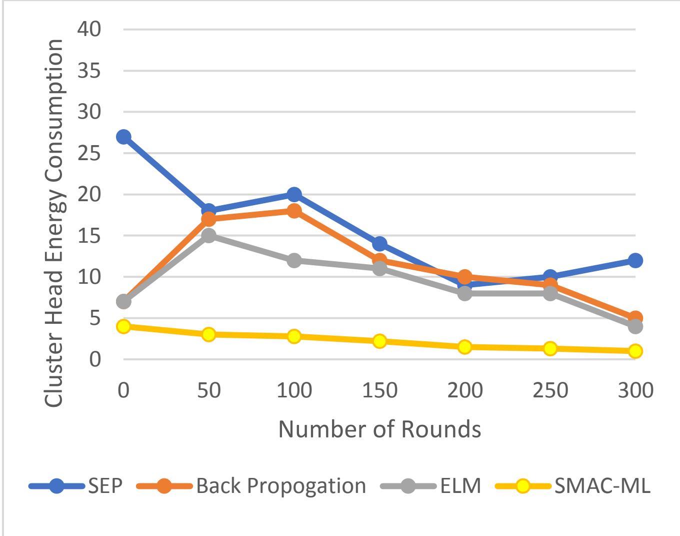

Figure 3 Comparison of Cluster Head Energy Consumption Versus No. of Rounds

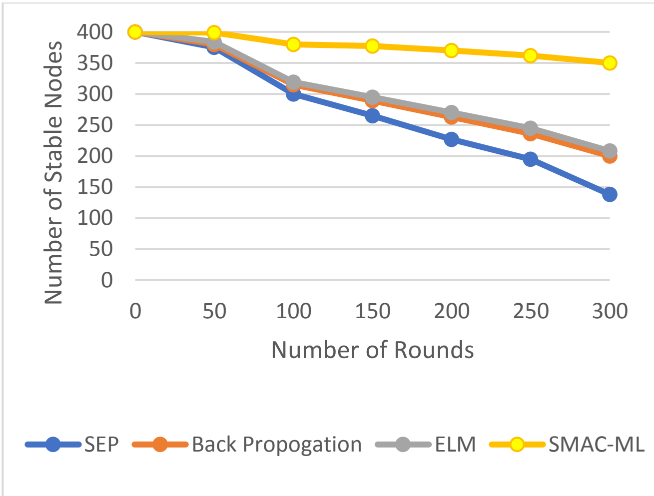

Figure 4 Comparison of Alive Nodes with respect to Number of rounds

One of the key metrics used to assess the effectiveness of wireless sensor networks is quantity of network nodes that have survived. The longer the lifespan of the network, the more stable the sensing system and capable of performing monitoring roles. With an increase in iteration of simulation, a curve representing network's surviving node evolves according to various data gathering strategies. Due to the utilization of energy, the number of nodes that persist in network gradually decreases as the number of cycles increases. According to Figure 4, node mortality has reached 67% and existence rate of sensor in network is only 33% at the 300th cycle of the SEP algorithm. On the other hand, the node survival rates of the Back Propagation neural network, ELM neural network, and SMCA-ML approach are greater. The BP neural network's node lifespan in the 300th round was 50%, whereas the ELM neural network's node existence rate was 52.7%. The transmission path and data aggregation proficiency is higher due to an algorithm suggested in this paper's improvement of the ML algorithm. Consequently, in the 300th cycle, the SMCA-ML algorithm's node persistence rate stays at 87.5%. The SMCA-ML-based heterogeneous wireless sensor network features more nodes, a greater existence rate, and an increases network lifespan than other three methods.

7. Conclusion

The SMCA-ML approach is used for data aggregation in heterogeneous sensor network efficiently decrease information redundancies in network and increase data aggregation efficiency. Nevertheless, the algorithm's capacity for generalization is limited by overall amount of nodes in hidden layer, selection of hidden layer parameters by random method and parameters are not varied throughout the training process. Consequently, we suggest SMCA-ML data aggregation approach to solve aforementioned issues. By employing the aforementioned technique, energy consumption of network is decreased, which has effect of extending network's lifespan. The suggested data aggregation method of heterogeneous sensor network based on SMCA-ML offers a larger performance increase, according to simulation testing. The subsequent study centres on the subsequent two facets: Initially, a complete heterogeneous wireless sensor network that accounts for residual energy, sensor node distance, and other characteristics can streamline the data aggregation process and enhance network performance. Secondly, the network's communication capability maintained continuously by including additional sensors when a node dies, and this approach of introducing additional nodes deserves more research.

Acknowledgments

Nil

Author Contributions

Ramesh Dennison Conceptualization, Methodology, Data curation, Writing-Original draft preparation, Validation, analysis and interpretation of results, and manuscript preparation.

Thirassama Jaya Visualization, Supervision, Guiding, Assisting, Reviewing and Editing

Conflict of Interest

The Authors declare no conflict of interest.

References

1. Giji Kiruba, D., & Benita, J. (2022). A Survey of Secured Cluster Head: SCH Based Routing Scheme for IOT Based Mobile Wireless Sensor Network. *ECS Trans.* Vol. 107, pp. 16725.
2. Ji, S., Tan C., Yang, P., Sun, Y.J., Fu D., & Wang, J. (2018). Compressive sampling and data fusion-based structural damage monitoring in wireless sensor network. J Supercomputing, vol. 74, no. 3, pp. 1108_1131.
3. Rajesh, D., & Rajanna, G. S. (2023). Energy-efficient CH selection protocol for mobile smart dust network. e-Prime - Advances in Electrical Engineering, Electronics and Energy, Vol. 4, 100176.
4. Rajesh, D., & Kiruba, D.G. (2021). A probability-based energy competent cluster based secured ch selection routing EC2SR protocol for smart dust. Peer-to-Peer Netw. Appl. Vol. 14, pp. 1976–1987.
5. Rajesh, D., & Jaya, T. (2021). ECIGC-MWSN: Energy capable information gathering in clustered secured CH based routing in MWSN. Materials Today: Proceedings, Volume 43, Part 6, pp. 3457-3462.
6. Rajesh, D., & Rajanna, G.S. (2023). Energy Aware Data Harvesting Strategy Based on Optimal Node Selection for Extended Network Lifecycle in Smart Dust. Journal of Intelligent & Fuzzy Systems, vol. 44, no. 1, pp. 939-949.
7. Zhou, Z., Liao, H., Gu, B., Huq, K. M. S., Mumtaz, S., & Rodriguez, J. (2018). Robust mobile crowd sensing: When deep learning meets edge computing. IEEE Netw., vol. 32, no. 4, pp. 54_60.
8. Zhou, Z., Chen, X., & Gu, B. (2019). Multi-scale dynamic allocation of licensed and unlicensed spectrum in software defined HetNets. IEEE Netw., vol. 33, no. 4, pp. 9_15.
9. Giji Kiruba, D., Benita, J., & Rajesh, D. (2024). Energy Efficient Clustering Mechanism for Malicious Sensor Nodes in IOT Based MWSN. International Research Journal of Multidisciplinary Scope (IRJMS), 5(1):50-61.
10. Giji Kiruba, D., Benita, J., & Rajesh, D. (2023). A Proficient Obtrusion Recognition Clustered Mechanism for Malicious Sensor Nodes in a Mobile Wireless Sensor Network. Indian Journal of Information Sources and Services, 13(2), 53–63.
11. Yang, A.M., Yang, X.L., Chang, J.C., Bai, B., Kong , F.B., & Ran, Q.B. (2018). Research on a fusion scheme of cellular network and wireless sensor for cyber physical social systems. IEEE Access, vol. 6, pp. 18786_18794.
12. Rajesh, & Giji Kiruba (2022). A Comparative Study On Energy Efficient Secured Clustered Approaches for IOT Based MWSN. Suranaree J. Sci. Technol. Vol. 29, no. 4, pp. 010151(1-18).
13. Ikram Daanoune & Abdennaceur Baghdad. (2022). IBRE-LEACH: Improving the Performance of the BRE-LEACH for Wireless Sensor Networks. Wireless Personal Communications: An International JournalVolume 126Issue 4, pp 3495–3513.

14. Din,S., Ahmad, A., Paul, A., Ullah Rathore, M. M., & Jeon, G. (2017). A cluster based data fusion technique to analyze big data in wireless multi-sensor system. IEEE Access, vol. 5, pp. 5069_5083.
15. Jan, S. R. U., Jan, M. A., Khan, R., Ullah, H., Alam, M., & Usman, M. (2019). An energy-efficient and congestion control data-driven approach for clusterbased sensor network. Mobile Netw Appl, vol. 24, no. 4, pp. 1295_1305.
16. Lu, Y., & Sun, N. (2018). A resilient data aggregation method based on Spatio- Temporal correlation for wireless sensor networks. EURASIP J. Wireless Commun. Netw., vol. 2018, no. 1, pp. 157_165.
17. Muthukumaran, K., Chitra, K., & Selvakumar, C. (2018). An energy efficient clustering scheme using multilevel routing for wireless sensor network. Computers & Electrical Engineering, Volume 69, Pages 642-652, ISSN 0045-7906.
18. Huafeng Wu, Jiangfeng Xian, Xiaojun Mei, Yuanyuan Zhang, Jun Wang, Junkuo Cao, & Prasant Mohapatra. (2019). Efficient target detection in maritime search and rescue wireless sensor network using data fusion. Computer Communications, Volume 136, Pages 53-62, ISSN 0140-3664.
19. Zhang, W., Yang, J., Su, H. et al. (2018). Medical data fusion algorithm based on Internet of things. Pers Ubiquit Comput 22, 895–902.
20. Giji Kiruba, & Benita. (2021). Energy capable clustering method for extend the duration of IoT based mobile wireless sensor network with remote nodes. Energy Harvesting and Systems, vol. 8, no. 1, pp. 55-61.
21. Rajesh, D., & Jaya, T. (2022). Enhancement of network lifetime by fuzzy based secure CH clustered routing protocol for mobile wireless sensor network. J Ambient Intell Human Comput vol. 13, pp. 2795–2805.
22. Rajesh, D., Giji Kiruba, D., & Ramesh, D. (2023). Energy Proficient Secure Clustered Protocol in Mobile Wireless Sensor Network Utilizing Blue Brain Technology. Indian Journal of Information Sources and Services, 13(2): 30–38.
23. Biradar, S. P., & Vishwanath, D. T. S. (2018). Network lifetime maximization of sensor network based on energy aware source tree routing. Int. J. Adv. Netw. Appl., vol. 10, no. 02, pp. 3788_3793.
24. Saeed Mehrjoo, & Farshad Khunjush. (2018). Optimal data aggregation tree in wireless sensor networks based on improved river formation dynamics. Comput. Intell., vol. 34, no. 3, pp. 802_820.
25. Rajesh, D., & Jaya, T. (2019). Exploration on Cluster Related Energy Proficient Routing in Mobile Wireless Sensor Network. International Journal of Innovative Technology and Exploring Engineering (IJITEE), Vol. 8, no. 4, pp. 93-97.
26. Aydın, V. A. (2024). Comparison of CNN-based methods for yoga pose classification. Turkish Journal of Engineering, 8 (1), 65-75
27. Lin, H., Bai, D., & Liu, Y. (2019). Maximum data collection rate routing for data gather trees with data aggregation in rechargeable wireless sensor networks,.Cluster Comput., vol. 22, no. S1, pp. 597_607.
28. Dennison, R., Dennison, R., Dasebenezer, G.K., & Chinnathurai, E.S. (2023). Enhancing lifespan and energy efficiency in mobile smart dust networks. Ingénierie des Systèmes d'Information., 28(5): 1317-1323
29. Osamy, W., Khedr, A. M., & Aziz, A. (2018). El-Sawy Cluster-tree routing based entropy scheme for data gathering in wireless sensor networks. IEEE Access, vol. 6, pp. 77372_77387.
30. Osaba, E., Yang, X.S., Fister, I., Del Ser, J., Lopez Garcia, P., & Vazquez Pardavila, A. J. (2019). A discrete and improved bat algorithm for solving a medical goods distribution problem with pharmacological waste collection. Swarm Evol. Comput., vol. 44, pp. 273_286.
31. Wang, J., Gao, Y., Liu, W., Sangaiah, A. K., & Kim H.J. (2019). An intelligent data gathering schema with data fusion supported for mobile sink in wireless sensor networks. Int. J. Distrib. Sensor Netw., vol. 15, no. 3, Art. no. 155014771983958.
32. Dennison, R., Dasebenezer, G. K., & Dennison, R. (2024). Energy capable protocol for heterogeneous blue brain network. Turkish Journal of Engineering, 8(1), 152-161
33. Chaturvedi, I., Ragusa, E., Gastaldo, P., Zunino, R., & Cambria, E. (2018). Bayesian network based extreme learning machine for subjectivity detection. J. Franklin Inst., vol. 355, no. 4, pp. 1780_1797.
34. Sattar, A. M. A., Ertugrul, Ö. F., Gharabaghi, B., Mcbean, E. A., & Cao, J. (2019). Extreme learning machine model for water network management. Neural Comput. Appl., vol. 31, no. 1, pp. 157_169.
35. Ji, S., Tan, C., Yang, P., Sun, Y.J., Fu, D., & Wang, J. (2018). Compressive sampling and data fusion-based structural damage monitoring in wireless sensor network. J Supercomput., vol. 74, no. 3, pp. 1108_1131.