Fuzzy Data Aggregation Approach to Enhance Energy-Efficient Routing Protocol for HWSNs

Asaad A Alhijaj[1], Baida'a Abdul Qader Khudor[2], Imad S. Alshawi[*3]
Email: asaad.abdulhassan@uobasrah.edu.iq[1], baidaa.khudur@uobasrah.edu.iq[2], emad.alshawi@uobasrah.edu.iq[3]
[*]Imad S. Alshawi
Department of Computer Science, College of Computer Science and Information Technology,
University of Basrah, Basrah, IRAQ

The sensor nodes' computing capability, communication capabilities, and power supply are severely constrained in WSNs, making sensor battery replacement or recharging difficult or even impossible. Therefore, energy is an important challenge to consider while creating WSNs. In hazardous circumstances, accurate data aggregation and routing are crucial, and the energy consumption of sensors must be closely controlled. Due to environmental conditions and short-distance sensors, however, there is a high possibility of duplicating data. Large datasets include a range of data, some of which are helpful while others are entirely unnecessary. This redundancy reduces performance in terms of redundant transmission and computation expense. Data aggregation, on the other hand, may reduce duplicate data in a network, hence reducing the volume of data sent and increasing the network's lifespan. In this context, two novel energy-conscious approaches called Fuzzy Data Aggregation with Spider monkey optimization (FDA-SMORP) for data aggregation in the cluster head and routing to the sink are presented. These strategies attempt to offset the energy consumption among all nodes in a wireless network such that these nodes exhaust all of their energy and die almost simultaneously. To demonstrate the efficacy of the suggested approaches in terms of minimizing delay caused by route planning, balancing energy usage, and extending network lifetime, the proposed methods are compared to some of the most well-known WSN systems.

Povzetek: Razvit je sistem za nadzorovanje potrošnje energije v senzorskih brezžičnih omrežjih.

1 Introduction

A WSN has a large number of nodes that can sense changes in the real-world environment. All aspects of human existence may benefit from a wirelessly networked sensor, such as smart buildings, the Internet environment, battlefields, industry, healthcare, and agriculture, and these are just a few of the uses of WSN[1]. The life of the network decreases as the sensors run out of power. These problems can only be solved if energy is used in the most efficient way possible. Because nodes create comparable data when placed close to each other or sent to data at the same time, this can cause data redundancy issues. This reduces network life energy consumption during processing, sending, and receiving data. To solve this problem, instead of sending each felt value to the sink separately, the data is first collected and aggregated using aggregate functions such as sum, average, etc., and it is then passed through routing protocols to deliver the data to the sink[2], [3].

Data aggregation is the analysis of raw data attributes and the application of correlations. Using a data aggregation approach, sensor nodes turn unprocessed data into a digest before delivering it to the sink. Data aggregation minimizes transmission costs and network overloading as a consequence of the decreased size of the digest. We argue that data aggregation is a critical method for reducing energy consumption in WSNs [4], [5]. However, there are still several obstacles to overcome before data aggregation performance can be improved. Existing contributions describe many aggregation algorithms that organize sensor nodes based on raw data to aggregate information. Nevertheless, aberrant data frequently emerges in raw data. Consequently, data instability has a direct impact on the efficiency of such approaches[6], [7]. In WSNs, there are a lot of ways to reduce the amount of data; like that each sensor collects before sending it to the sink, or while aggregating data in the cluster head (CH). Or use a way for the data packets to be routed like an efficient clustering solution with data aggregation, employing several mobile sinks for heterogeneous WSN[8].

Several researchers have highlighted the problem of data aggregation with routing in WSNs [9], [10]. When it comes to WSNs in general, the most difficult problem is finding ways to improve energy efficiency so that the network can last much longer [11], [8].

Table 1 summarizes the related works with their methodology, performance, and results. So, the current

work proposes a new energy-conscious protocol for HWSNs called Fuzzy Data Aggregation with Spider monkey Optimization Routing Protocol (FDA-SMORP). The new protocol can combine two approaches Fuzzy Data Aggregation[11] with Spider monkey Optimization

Routing Protocol [12]. So, FDA-SMORP is used to aggregate the sensing data inside the clusters by the cluster heads, which are used by the FDA, and to send the aggregated data through the optimal path to the sink for HWSNs by using the SMORP.

Table 1: Summarization table on the related works

Ref	Methodology	Performance/Results
[13]	• Ant Data Aggregation Algorithm	• A population-based approach such as the ant colony system allows researchers to naturally traverse research space in optimization settings in pursuit of the most useful data, and it is via data aggregation that wsns may reduce their power consumption. • In each cluster head, the sink node sends a unique seed vector that accounts for network dispersion. Clusters transmit measurement data to the sink node through a multi-hop routing tree.
[14]	• Support Vector Machine • Fisher's Discrimination Ratio	• His incremental support vector machine (SVM) training method aimed to eliminate unessential input. • Sets may be distinguished between data that has been aggregated and data that has been disseminated in a set by using Fisher's Discrimination Ratio (FDR). • The training of SVM is quicker since there are fewer data samples necessary.
[15]	• Mobile Sink Is For Data Aggregation	• They represented solutions for effective data aggregation with several movable troughs in HWSNs. When using the statically sink-based technique, data packets are dumped over a multi-hop connection and sent throughout the network. As a result, the fixed basin is inefficient in terms of its use of energy. • A mobile sink is utilized to gather data, which uses less power and hence prolongs the network's lifetime.
[16]	• Naive Bayes Prediction	• Data from wsns can be reduced using Naive Bayes Incremental Prediction, making the network last longer. And extract only the necessary data.
[17], [18]	• Particle Swarm Optimization	• Data aggregation has been suggested by utilizing compressive sensing technology, where active sensor nodes are optimized to decrease the amount of duplicate data using particle swarm optimization. As a result, they are efficient in terms of their use of energy.
[8]	• Fuzzy Dstar-Lite	• The authors proposed Fuzzy Dstar-Lite as a routing technique for producing the optimum information routing for HWSNs. Additionally, it brings up the point of outdoing the obstruction example and elucidates the Unbalanced Energy Depletion (UED) problem in the network.
[19]	• Open-Pit Mining	• Open mining is presented as a method for aggregating data that is both efficient and cost-effective to use. • This data mining method uses a lot of wsns. Each one has a center node around which many virtual pits collect and send data to the sink.
[20]	• Neural Network • Cosine Similarity	• the Reduce duplicate data and eliminate outliers by using a neural network of self-organized maps. • The use of cosine similarity in sensor node creation further simplifies the process based on the data's density and similarity.
[21], [22]	• Spider Monkey Optimization Routing Protocol	• The researchers described a novel technique for clustering the HWSNs approach that employed an efficient way of selecting the head of the cluster nodes, the degree of sensor nodes, and the remaining energy. Additionally, the chaining technique is used to collect and send the information package. • They proposed a swarm-based intelligence method called SMORP that was used in the homogeneous WSNs and the heterogeneous HWSNs. This method is used to find the optimal path in the network based on a set of routing criteria.
[12]	• Fuzzy Data Similarity	• A method called fuzzy data similarity (FDS) is presented to determine the similarity between two texts. To demonstrate the efficacy of the proposed method, the FDS was shown to be around 93% accurate. • Most comparable techniques employ distance measurements to evaluate the differences between a pair of objects, and the suggested algorithm is compared to one of the most used distance scales (Jaccard similarity, Cosine similarity, Overlap Coefficient).

This paper is organized as follows: In Section 2 presents a proposed smart data aggregation with a new routing protocol for HWSN. Section 3 shows the simulation results of the proposed method. Finally, the conclusion of this paper is presented in Section 4.

2   FDA with SMORP for HWSNs

The proposed method represents the process of aggregation and routing data for HWSNs. We assume that our network has two types of heterogeneous sensor nodes: the normal senses (N-sensor) and the high senses (CH-sensor). The N-sensors have limited resources, such as limited processing speed, storage capacity, and communication bandwidth. While the CH-sensors have high resources and represent the cluster heads in the network. The network is configured as follows: the N-sensors are deployed randomly, while the CH-sensors are deployed carefully. The cluster partition method [24] is used, in this paper, to organize the HWSNs as orderly clusters.

The SMORP selects the appropriate next hop to the sensor node based on the routing criteria (maximum remaining energy, fewest hops, and lowest traffic load). This work supposes: (i) All N-sensors have the same transmission range and begin with the same amount of battery power. Each N-sensor in (ii) is aware of its position, as well as that of its CH and neighbors. (iii) All CHs have the same transmission range and start-up power from the battery (iv). Each CH is aware of its position and also of its neighbors, namely the other CHs and the sink location.

2.1   Network Model

The goal of the proposed model is to ensure that when some of the sensors send an event at the same time, there is a high probability that the same event will repeat, increasing the amount of data that occupies high space and lowering energy in the network. CH's FDA is used to effectively aggregate data based on redundancy elimination, extract useful information, and then send it via an improved spider monkey protocol, which reduces the power consumption of sensor nodes and thus extends the life of the network. Figure 1 shows the data aggregation with routing in HWSNs.

The routing protocol is one of the major concerns in extending the lifetime of HWSNs. If any sensor node (N-sensor or CH-sensor) runs out of energy during the routing protocol, the information exchange between (N-sensor and CH) and (CH and the sink) will likewise be broken. Typically, this results in a shortage of HWSNs over their lifetime. The amount of power each sensor in an HWSN gets affects how long it lasts, it is very important to save power in those sensors so that the network as a whole can last as long as possible. In this light, the SMORP is capable of extending the lifetime of HWSNs by lowering energy expenditures and evenly distributing energy usage.
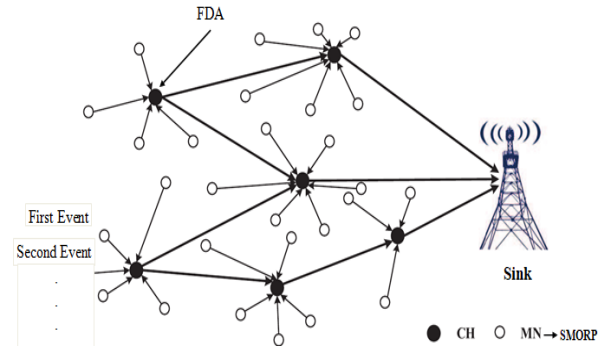

Figure 1 :Data aggregation with routing in HWSNs

2.2   FDA-SMORP proposed

FDA-SMORP is used to aggregate the sensing data inside the clusters by the cluster heads, the FDA has recommended a method for aggregating data that eliminates redundancies and extracts relevant information. A similarity measure in the context of data mining is a distance whose dimensions indicate object properties. Thus, if the distance between two data points is small, the objects will be highly similar, and vice versa[25]. The majority of aggregation techniques use distance measurements to evaluate the differences between a pair of items [11].

After that, the SMO method evaluates a tree structure in the course of (N, Fit), where N is the candidate node set in the forwarding route and Fit is the fitness functions set that each candidate node n ∈ N is assigned a fitness function value fit(n). The tree node will explore depending on its fitness function.

In SMORP, the created routing route is used repeatedly (rounds), and the status of each node along the way is evaluated to decide if the same path should be used for the next round. According to the previous assumption, the sink has access to current information on each node's battery energy, position coordinates, and network traffic load. Eq. (1) is used to determine the fitness of a contiguous node (n_i).

$$fit(n_i) = fuzzy(RE(n_i), TL(n_i), D(n_i)) \qquad (1)$$

Where $RE(n_i)$, $TL(n_i)$, and $D(n_i)$ are the remaining energy, traffic load, and the distance to the destination for node n_i, respectively. All these parameters are the inputs that will calculate the fitness value to the node n. After that, the GLSM assesses the information gathered from all of LLSM's neighbor nodes and chooses the optimal node with the greatest probability P with the probability value specified by Eq. (2):

$$P(n_i) = \frac{fit(n_i)}{\sum_{j=1}^{N} fit(n_j)} \qquad (2)$$

Where P(n_i) is the probability associated with node n_i, fit (n_i) is the fitness associated with node n, and N is the number of neighbor nodes.

Figure 2 shows how data is aggregated within the routing protocol FDA-SMORP in each cluster head effectively.

3   Performance Evaluation of FDA-SMORP

The primary goal of this paper is to develop the SMORP [11]. In this paper, we assume that three sensors send the events at the same time. Thus, the network is optimized by the assembly process in each cluster. The simulation results for the proposed method are compared over three scenarios.

3.1 Simulation Setting

Simulations are carried out in MATLAB R2010a (version 7.10) under Windows 7 (32 bits). The experiments are performed on a PC (ThinkPad T410i, China) with an Intel ^{R} Core ^{TM} i3 Processor running at 2.4 GHz and 2 GB of RAM. To make the network as realistic as possible, some parameters must be set in the

system. Table 2 depicts a heterogeneous network with *1000* N-sensors and *36* CHs randomly arranged within a *300 m x 300 m* square topographical area. Both systems are used the clustering method to group the N-sensors around CH-sensors. Also, they used a radio model [26] and exhausted their transmission cycles *(2000)*. Each system produces a *2 KB* packet length. All N-sensors and CH-sensors start with the same starting energy of *(0.5 J)* and *(2.5 J)* with a sensed transmission of *(20 m)* and *(80 m)*, respectively. The traffic load, in each node, is assumed to be generated randomly between [0...10] and [0...50] for the N-sensors and the CH-sensors, respectively.

3.2 Simulation Results

The life of HWSN can be extended by using a CH fuzzy data collection method called FDA with a routing protocol called SMORP that has been optimized in to increase energy efficiency. To see how well it worked, it was tested in three different scenarios, if the same routing metrics and the same environment were used in both.

To validate the operation of the proposed model, three scenarios are applied to the model. Assuming the
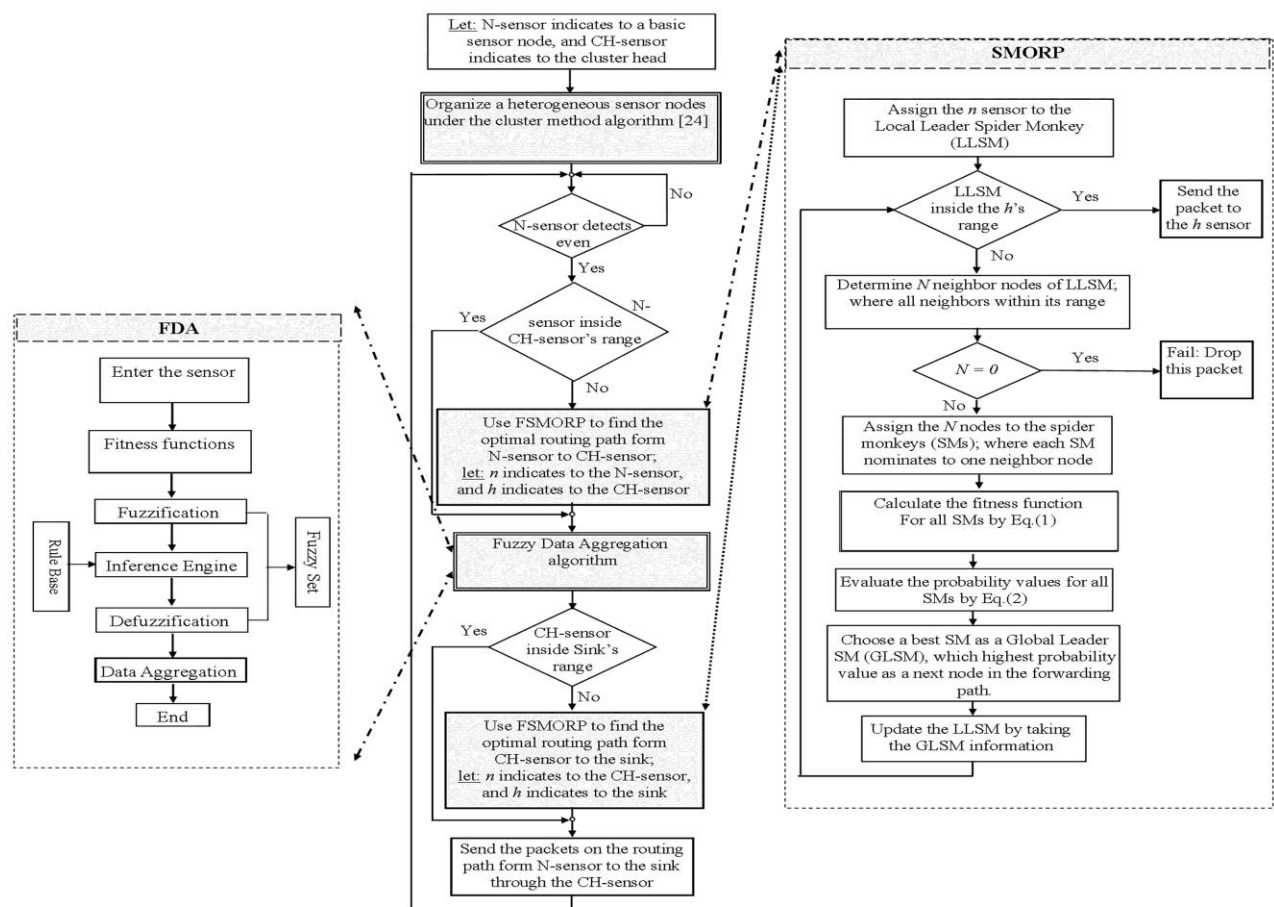

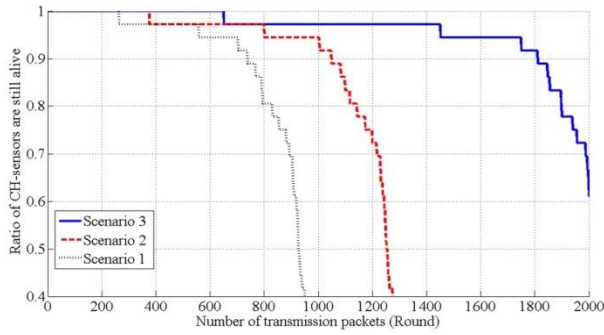Figure 2: The flow chart of the FDA-SMORP proposed

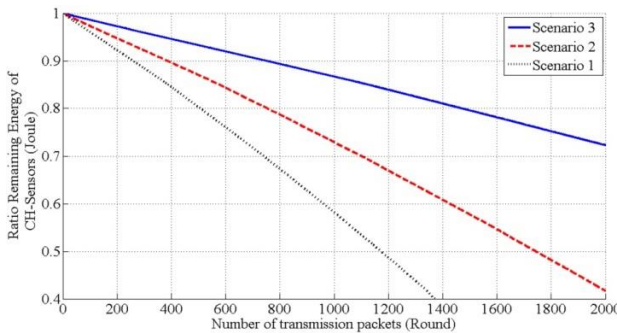Figure 3: The CH-sensors ratio remains alive (three N-sensors sent at a time).

Figure 4: The energy ratio of the remaining CH-sensors (three N-sensors sent at a time).

Table 2: Simulation parameters

Parameter		Value
Area of topographical (meters)		300m x 300m
Location of the sink (meters)		(0, 150)
Length of control packets		2k
No. of transmission packets (rounds)		2 x 10³
N-sensors	Number of nodes	1000
	Limit of transmission distance	20 m
	Initial energy	0.5 J
	E_{elec}	50 nJ/bit
	E_{amp}	100 pJ/bit/m²
	Max. traffic in node's queue	10
CHs	No. of nodes	36
	Limit of transmission distance	80 m
	Initial energy	2.5 J
	E_{elec}	100 nJ/bit
	E_{amp}	200 pJ/bit/m²
	Max. traffic in node's queue	50

packet size is 2k, and then considers setting a high similarity threshold. That is, the greater the similarity of the detected events, the less the amount of data transmitted to the sink. FDA proposed algorithm is put into every CH-sensor. Thus, we notice the effect of the algorithm on clusters only, instead of the N-sensors

a) First Scenario

In this scenario, assuming the head of the cluster receives different messages from the three sensors, the data is all aggregated and transmitted.

b) Second Scenario

In the second scenario, assuming that the head of the cluster receives two similar messages from the three sensors, the messages are aggregated removed from the similarity, and sent to the sink.

c) Third Scenario

In the third scenario, assuming that the head of the cluster receives three identical messages from the three sensors, the messages are aggregated removed from the similarity and sent to the sink.

The network lifetime results obtained using three scenarios are compared by counting the number of sensors that remain alive after each data round. At this point, Figure 3 shows the proportion of CH sensors, which are still alive in each scenario. As a result, the performance of the third scenario outperforms the performance of both the first and second scenarios, meaning that the more the detected events are similar, the smaller the amount of data sent. In light of this, we note that the amount of energy consumed in the third scenario is small compared to the first and second scenarios based

on the total number of nodes that are still alive in the network. Here, after sending (2000) packets to two sensors over the network, the result of the network life achieved in the third scenario is approximately (60%) more than in scenario two and approximately (80%) more than in scenario first.

The percentage of energy remaining in the CH sensors varies with the number of transfer cycles depending on the system used. The third scenario outperforms the first and second proposed scenarios in terms of overall performance and efficiency.

Figure 4 shows how the percentage of residual power for the CH sensors varies based on the transfer mode used. As you can see, the third scenario is better than the first and second scenarios by maintaining the stability of the network for as long as possible.

4   Conclusion

Many routing protocols have been used in WSNs for saving energy. Nevertheless, just saving energy is inadequate to prolong the life of the networks. Large datasets include a variety of information, some of which is useful, while others are completely superfluous. Assuming that some of the sensors transmit an event at the same time or when they are close to each other, this can cause data redundancy issues these problems can only be solved if energy is used in the most efficient way possible. As a result, two novel energy-conscious approaches called Fuzzy Data Aggregation with Spider monkey optimization (FDA-SMORP) for data

aggregation in the cluster head and routing to the sink are presented. These strategies attempt to offset the energy consumption among all nodes in a wireless network such that these nodes exhaust all of their energy and die almost simultaneously. The simulation results of the proposed model indicate that FDA-SMORP outperformed in terms of greatly enhancing data latency reduction and lifetime maximization of the network.

References

[1] C. Nakas, D. Kandris, and G. Visvardis, "Energy efficient routing in wireless sensor networks: A comprehensive survey," *Algorithms*, vol. 13, no. 3, p. 72, 2020. https://doi.org/10.3390/a13030072.

[2] M. D. Aljubaily and I. S. Alshawi, "Energy sink-holes avoidance method based on fuzzy system in wireless sensor networks.," *Int. J. Electr. Comput. Eng.*,vol.12,no.2,2022.http://doi.org/10.11591/ijece.v12i2.

[3] I. S. Alshawi, "Balancing Energy Consumption in Wireless Sensor Networks Using Fuzzy Artificial Bee Colony Routing Protocol," *Int. J. Manag. Inf. Technol.*, vol. 7, no. 2, pp. 1018–1032, 2013. https://doi.org/10.24297/ijmit.v7i2.3354.

[4] S. Kumar and S. Kumar, "Data aggregation using spatial and temporal data correlation," *2015 1st Int. Conf. Futur. Trends Comput. Anal. Knowl. Manag. ABLAZE 2015*, no. Ablaze, pp. 479–483, 2015. https://doi.org/10.3390/s18092840.

[5] N. Nguyen, B. Liu, S. Chu, and H. Weng, "Challenges , Designs , and Performances of a Distributed Algorithm for Minimum-Latency of Data-Aggregation in Multi-Channel WSNs," *IEEE Trans. Netw. Serv. Manag.*, vol. PP, no. c, p. 1, 2018.https://doi.org/10.1109/TNSM.2018.2884445.

[6] M. R. Choudhari and U. Rote, "Data Aggregation Approaches in WSNs," *2021 Int. Conf. Comput. Commun. Informatics, ICCCI 2021*, pp. 27–32, 2021.https://doi.org/10.1109/ICCCI50826.2021.9402430.

[7] A. Karaki, A. Nasser, C. A. Jaoude, and H. Harb, "An adaptive sampling technique for massive data collection in distributed sensor networks," *2019 15th Int. Wirel. Commun. Mob. Comput. Conf. IWCMC 2019*, pp. 1255–1260, 2019. https://doi.org/10.1109/IWCMC.2019.8766469.

[8] I. S. Alshawi, A.-K. Y. Abdulla, and A. A. Alhijaj, "Fuzzy dstar-lite routing method for energy-efficient heterogeneous wireless sensor networks," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 19, no. 2,pp.1010,2020,http://doi.org/10.11591/ijeecs.v19.i2.pp906-916.

[9] W. K. Yun and S. J. Yoo, "Q-Learning-based data-aggregation-aware energy-efficient routing protocol for wireless sensor networks," *IEEE Access*, vol. 9, pp.10737–10750,2021. https://doi.org/10.1109/access.2021.3051360.

[10] N. Chandnani and C. N. Khairnar, "Efficient Data Aggregation and Routing Algorithm for IoT Wireless Sensor Networks," *IFIP Int. Conf. Wirel. Opt. Commun. Networks, WOCN*, vol. 2019-Decem,2019,https://doi.org/10.1109/WOCN45266.2019.8995074.

[11] I. S. Alshawi, Z. A. Abbood, and A. A. Alhijaj, "Extending lifetime of heterogeneous wireless sensor networks using spider monkey optimization routing protocol," vol. 20, no. 1, pp. 212–220, 2022.http://doi.org/10.12928/telkomnika.v20i1.20984.

[12] D. K. Altmemi and I. S. Alshawi, "Enhance Data Similarity Using a Fuzzy Approach," *J. Posit. Sch. Psychol.*, pp. 1898–1909, 2022.

[13] R. Misra and C. Mandal, "Ant-aggregation: ant colony algorithm for optimal data aggregation in wireless sensor networks," in *2006 IFIP International Conference on Wireless and Optical Communications Networks*, 2006, pp. 5-pp. http://doi.org/10.1109/WOCN.2006.1666600.

[14] X. Shen, Z. Li, Z. Jiang, and Y. Zhan, "Distributed SVM classification with redundant data removing," in *2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing*, 2013, pp. 866–870. http://doi.org/10.1109/GreenCom-iThings-CPSCom.2013.152.

[15] A. Muthu Krishnan and P. Ganesh Kumar, "An Effective Clustering Approach with Data Aggregation Using Multiple Mobile Sinks for Heterogeneous WSN," *Wirel. Pers. Commun.*, vol. 90, no. 2, pp. 423–434, 2016 https://doi.org/10.1007/s11277-015-2998-6.

[16] P. D. Ganjewar, S. Barani, and S. J. Wagh, "Data reduction using incremental Naive Bayes Prediction (INBP) in WSN," *Proc. - IEEE Int. Conf. Inf. Process. ICIP 2015*, pp. 398–403, 2016, http://doi.org/10.1109/INFOP.2015.7489415.

[17] D. gan Zhang, T. Zhang, J. Zhang, Y. Dong, and X. dan Zhang, "A kind of effective data aggregating method based on compressive sensing for wireless sensor network," *Eurasip J. Wirel. Commun. Netw.*, vol. 2018, no. 1, 2018, https://doi.org/10.1186/s13638-018-1176-4.

[18] M. I. Adawy, S. A. Nor, and M. Mahmuddin, "Data redundancy reduction in wireless sensor network," *J. Telecommun. Electron. Comput. Eng.*, vol. 10, no. 1–11, pp. 1–6, 2018.

[19] H. Ramezanifar, M. Ghazvini, and M. Shojaei, "A new data aggregation approach for WSNs based on open pits mining," *Wirel. Networks*, vol. 27, no. 1, pp. 41–53, 2021. https://doi.org/10.1007/s11276-020-02442-9.

[20] I. Ullah and H. Y. Youn, "A novel data aggregation scheme based on self-organized map for WSN," *J. Supercomput.*, vol. 75, no. 7, pp. 3975–3996, 2019, https://doi.org/10.1007/s11227-018-2642-9.

[21] M. Pandey, L. K. Vishwakarma, and A. Bhagat,

"An energy efficient clustering algorithm for increasing lifespan of heterogeneous wireless sensor networks," in *International Conference on Next Generation Computing Technologies*, pp. 263–277 2017, https://doi.org/10.1007/978-981-10-8660-1_20.

[22] M. Al Mazaideh and J. Levendovszky, "A multi-hop routing algorithm for WSNs based on compressive sensing and multiple objective genetic algorithm," *J. Commun. Networks*, no. 99, pp. 1–10, 2021, https://doi.org/0.23919/JCN.2021.000003.

[23] A. H. Jabbar and I. S. Alshawi, "Spider monkey optimization routing protocol for wireless sensor networks.," *Int. J. Electr. Comput. Eng.*, vol. 11, no.3,2021,http://doi.org/10.11591/ijece.v11i3.pp2432-2442.

[24] I. S. AlShawi, L. Yan, W. Pan, and B. Luo, "Fuzzy chessboard clustering and artificial bee colony routing method for energy-efficient heterogeneous wireless sensor networks," *Int. J. Commun. Syst.*, vol. 27, no. 12, pp. 3581–3599, 2014, https://doi.org/10.1002/dac.2560.

[25] G. Sahar, K. A. Bakar, F. T. Zuhra, S. Rahim, T. Bibi, and S. H. H. Madni, "Data Redundancy Reduction for Energy-Efficiency in Wireless Sensor Networks: A Comprehensive Review," *IEEE Access*, 2021, https://doi.org/10.1109/ACCESS.2021.3128353.

[26] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *Proceedings of the 33rd annual Hawaii international conference on system sciences*, pp. 10pp,2000,https://doi.org/10.1109/HICSS.2000.926982.