A New Method Based on Machine Learning to Increase Efficiency in Wireless Sensor Networks

Baida'a Abdul Qader Khudor[1], Yousif Abdulwahab Kheerallah[2], Jawad Alkenani[3*]
Email: baidaa.khudur@uobasrah.edu.iq, yousif.abdalwahab@sa-uc.edu.iq, Jawadalkenani@sa-uc.edu.iq
[1]Department of Computer Science, College of Computer Science and Information Technology, University of Basrah, Basrah, IRAQ
[2,3]Department of Computer Science, Shatt Al-Arab University College, Basra, Iraq

Keywords: Data Aggregation, Kalman filter, Support vector machine, wireless sensor networks

Received: September 15, 2022

Wireless sensor networks (WSNs) contain many sensor nodes, and this network is used for many applications such as military, medical, and others. Accurate data aggregation and routing are critical in hostile environments, where sensors' energy consumption must be carefully monitored. There is, nevertheless, a substantial probability of duplicate data due to ambient circumstances and short-distance sensors. Large datasets include a variety of information, some of which is useful, while others are completely superfluous. This redundancy degrades performance in terms of computing cost and redundant transmission. Data aggregation, on the other hand, may eliminate redundant data in a network. In this paper new method called Kalman filter with Support vector machine (KF-SVM) is introduced to classify and data aggregate and get rid of noise in WSNs, which enhances network efficiency and extends its lifetime.

Povzetek: V prispevku je opisana izvirna metoda za agregiranje podatkov v senzorskih omrežjih, ki dela na osnovi Kalmanovega filtra in SVM.

1 Introduction

The creation of low-power sensors and the deployment of large-scale sensor networks are the results of developments in wireless communication and microelectronic devices. With the capacity for pervasive surveillance, sensor networks have received significant interest in a variety of application domains, including habitat monitoring, item tracking, environment monitoring, military, disaster management, and smart environments. In many applications, dependable real-time monitoring is an absolute must[1]. These applications generate an enormous volume of geographically dispersed, dynamic, and heterogeneous data. Data mining can help automated or human-driven tactical/strategic decision-making if this raw data is efficiently evaluated and turned into meaningful knowledge. Therefore, it is crucial to create techniques for mining sensor data for patterns so that intelligent judgments may be made quickly[2].

The fundamental purpose of topology control in WSNs is to guarantee the secure and dependable transfer of data acquired by sensor nodes. ensuring the integrity of the maximum connected graph is the primary strategy for preserving the created network topology. In recent years, the need for WSN application fields has increased steadily. The structure of the topology is produced by a self-organizing network. In many instances, not only perfect communication but also the scalability and universality of the network are necessary. In other words, a topology must accommodate the needs of many sorts of users. This architecture significantly enhances network functionality. The network needs both theoretical operability and real expanding operation, which places increased demands on the topology[3].

The heterogeneous data-transmission network created in this research is a network with a tree topology that exhibits excellent data transmission performance. The tree topology is superior to other network topologies in terms of data transmission and damage resistance. The benefits of the tree structure include efficient data transmission and data aggregation by non-leaf nodes. When designing the minimal tree topology in the references, energy and network latency requirements are thoroughly examined. The advantages and drawbacks of the existing network architecture are measured primarily by the node energy consumption and network throughput of WSNs. There are several network topologies designed for efficient data transmission, including cluster-based topology, tree-based topology, and others[1], [3]. The data mining community has recently paid a significant deal of attention to the extraction of knowledge from sensor data. On-sensor data, many techniques concentrating on clustering, association rules, common patterns, sequential patterns, and classification have proven effective[4]. However, the design and deployment of sensor networks present unique research challenges due to their large scale (up to thousands of sensor nodes), random and risky deployment, loss of communicating environment, limited power supply, and high failure rate. Because traditional data mining techniques are centralized, computationally expensive, and focused on disk-resident transactional data, they are inapplicable[5]. As a result, new algorithms have been

Table 1: Summarization table on the related works.

Ref	Methodology	Performance/Results
[10]	• Fuzzy Data Similarity	• The fuzzy data similarity (FDS) approach is introduced for determining the similarity between two texts. To illustrate the effectiveness of the suggested system, it was determined that the FDS was around 93% accurate. • Comparable approaches often involve distance measurements to determine the differences between two objects, and the proposed algorithm is compared to one of the most widely employed distance scales (Jaccard similarity, Cosine similarity, Overlap Coefficient).
[16]	• SCDRE protocol	• Using the right methods for data aggregation can help reduce data repetition and improve efficiency. We propose the Spatial Correlation based Data Redundancy Elimination for Data Aggregation (SCDRE) protocol, which makes use of statistical methods in sensor networks to perform redundant data elimination on two different levels: at the source level, using a simple data similarity function, and at the aggregator level, using a correlation coefficient. • • SCDRE exceeds other existing algorithms in terms of aggregation ratio, data accuracy, and energy consumption. the outcomes favor SCDRE over other methods, and the results back this up.
[26]	• REDA algorithm	• The data aggregation strategy emerges as a significant method for reducing the energy consumption of sensor nodes and enhancing bandwidth utilization. REDA is a Redundancy Elimination Data aggregation algorithm based on a pattern generation methodology. • The proposed pattern is unique to the sensed data and utilizes differential data collected from successive sensor node iterations. Consequently, redundant data transmission from sensor nodes within the same cluster to the respective cluster head (CH) is avoided throughout all iterations. Evaluation of performance demonstrates that the REDA algorithm reduces energy consumption by up to 44% compared to protocols without data aggregation methods. Moreover, in comparison to existing data aggregation algorithms, ESPDA and SRDA are superior..
[27]	• Extreme Learning Machine (ELM) • Kalman filter	• The clustering of nodes and extreme learning machine (ELM), a novel data aggregation scheme is proposed that efficiently reduces redundant and incorrect data. The projection stage of the ELM uses a distance-based radial basis function derived from the work of Mahalanobis to reduce the instability of the training procedure. The data at each sensor node is additionally filtered using the Kalman filter before being transmitted to the cluster head.
[28]	• Open-Pit Mining	• Open mining is presented as an efficient and cost-effective technique for data aggregation. This data mining technique employs numerous WSNs. Each has a central node around which a multitude of virtual pits collect and transmit information to the sink.
[29]	• 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.
[30]	• Kalman Filter	• lessen the burden on the environment, we've developed and deployed a hardware acceleration of the KF. Time to completion, power usage, and other metrics have been compared between the software and hardware versions of the method, and space requirements. The results demonstrate a reduction of approximately 97% in energy consumption and execution time with no discernible increase in area.
[31]	• 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.
[32]	• Mobile Sink Is For Data Aggregation	• They represented strategies for efficient data aggregation in HWSNs with several moveable troughs. When using the statically sink-based method, data packets are dropped through a multi-hop connection and distributed over the network. As a consequence, the fixed basin's energy use is inefficient. Utilizing a mobile sink to collect data conserves energy and extends the network's lifespan.

developed, and some data mining techniques have been changed, to process the data produced by sensor networks. In the past decade, numerous knowledge discovery strategies, techniques, and algorithms have been proposed[6].

Since data mining is such a large topic, it can be used for data from any domain; for more general surveys on data mining techniques [7], which investigated machine learning and data mining strategies for evaluating medical data. Since the classification of data mining techniques in this survey is based on frequent pattern mining, clustering, and classification, there are numerous studies on each of these techniques. For instance[8] describes frequent pattern mining over data streams. However, none of the aforementioned studies explored data mining techniques that concentrate on the extraction and analysis of information from WSN data[9].

Data gathering is one of the most critical challenges with WSNs since it is one of the most energy-hungry processes, if not the most one, that limits the network's lifetime. Thus, developing energy-efficient routing protocols has revolutionized the state of the art for two decades. Most of the designed protocols were aimed at extending the network's lifetime by balancing the energy depleted by the sensors' batteries during the multi-hop data dissemination pattern of flat and hierarchal networks. One of the most progress in such a field was developing a protocol for homogeneous WSNs called LEACH protocol which adopted the clustering technique to distribute the processes of collecting and sensing/transmitting among sensors periodically. Then, the use of heterogeneous WSNs was come out as a solution for the high energy depletion of the cluster head in the homogeneous WSNs[10].

Table 1, shows the Summarization of the Related Works. In this work, a new method hybrid is the Kalman filter [11], and the Support Vector Machine[12]. Kalman filter with Support Vector Machine (KF-SVM) is introduced for data classification, aggregation, and noise elimination in WSNs, which enhances network efficiency and extends network life.

This paper investigates algorithms and methodologies built specifically for WSNs data, leading not only to a distinct classification, evaluation, and discussion on various domains but also to a variety of solution options. The goal is how data mining algorithms will be used to intelligently develop sensor network applications. The rest of the paper is organized as follows: In the second section, the process of organizing the heterogeneous network in WSN is presented, with emphasis on the reasons for resorting to clustering and its importance. The third topic is a detailed explanation of the age of the sensor network and the problems that occur in it that lead to energy loss, with an explanation of the energy consumption calculation model. In the fourth section, the proposed model for data classification and aggregation with noise removal to increase network efficiency is presented. The fifth section presents the parameters for network simulation with the presentation of the results of the proposal. Finally, a summary of the research is presented in the sixth section.

2 Organization of HWSNs

Cluster partition is the process that partitions the S-sensors into several clusters depending on the number of CHs. In heterogeneous WSNs clustering, the key difficulties are informing the CHs about which S-sensors are within their clusters and informing the S-sensors whose clusters they belong to. The concept of cluster partitioning around CHs in white cells will be described further below. Based on the IDs assigned to the CHs, they will together broadcast a message that contains their position information. Each S-sensor would then generate a list of CHs from which it has listened where the signals that have been accurately received by the S-sensor will be favored to sort their source ID first. Consequently, each S-sensor will be aware of which CH it may correspond to and will pick the CHs at the top of the list as its ideal CH. Then, S-sensors will send a broadcast message that informs the chosen CHs that it has been picked to be their CH. Since the transmitter of the S-sensor is less than that of CHs, some of the S-sensor would not be able to inform that message to the desired CH. Thus, it will send it through other nearby S-sensors to successfully reach out to the desired CH. Additionally, some S-sensors may select their best CHs. However, they will not be able to connect with it due to the absence of nearby S-sensors, thus they would choose other CH stored in their generated list[13].

The exploration of raw data properties and the application of correlations constitute data aggregation. Using a strategy based on data aggregation, sensor nodes transform unprocessed data into a digest before sending it to the sink. Due to the lower side of the digest, data aggregation reduces transmission costs and network congestion. We contend that data aggregation is essential for lowering energy usage in WSNs. However, there are still various challenges to overcome before the performance of data aggregation can be enhanced. Existing contributions describe many aggregation algorithms that arrange sensor nodes based on raw data to aggregate data. Despite this, aberrant data arises regularly in raw data. Consequently, data inconsistency has a direct influence on the effectiveness of such methods.[14], [15].

Furthermore, several aggregation processes are designed for certain data kinds (e.g., temperature data) or network characteristics (e.g., grid network), limiting their flexibility. Consequently, we are motivated to develop data- and property-neutral aggregating strategies.

Without data aggregation in a WSN, sensor nodes continuously transmit all raw data to the sink. Although this data is usually redundant or related, it has the following disadvantages: 1) duplicated data is worthless to the application, 2) the likelihood of network congestion drastically increases, 3) network capacity is squandered, and 4) energy consumption rises proportionally. Prior research [16], [14], indicates that temporal and spatial correlations are frequently based on raw data. There is a temporal correlation between data gathered at distinct time instants for a particular sensor node, whereas there is a spatial correlation between data collected from neighboring sensor nodes. As seen in Fig. Figure 1. (a), when a sensor node is used to monitor the temperature in a particular location, the values obtained frequently remain constant for up to 20 minutes, if not an hour. Additionally, when two sensor nodes are used to measure temperature in the same room, as seen in Fig. Figure 1. (b), data obtained by one node is frequently comparable to, if not identical to, data gathered by another node[17].

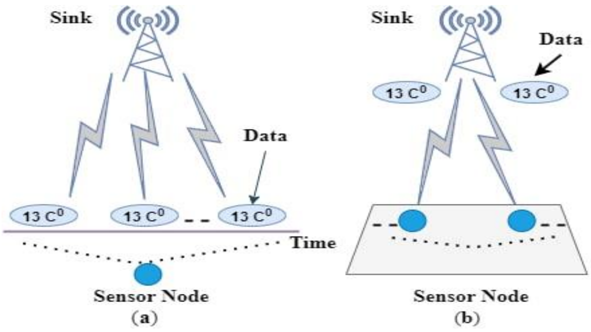

Figure 1**:** Data correlation: (a) Temporal correlation; (b) Spatial correlation[17]

As a general rule, an aggregation protocol should accomplish the following three basic goals[18], [19]:

Energy-saving: Data aggregation eliminates redundancy and linked communications in a network, which directly reduces the network's energy usage. Given that energy is the key constraint for WSNs, the data aggregation structure must prioritize energy-saving.

Data accuracy: The precision of the recovered data compared to the original data is referred to as data precision. Several pieces of information may be lost when raw data is aggregated into a digest by sensor nodes. It is thus acceptable to expect some divergence from the raw data while recovering the sink-side data to be present. An application's ability to reduce its energy consumption with acceptable precision is a universal need.

Network capacity saving: Due to sensor node bandwidth limits, WSNs' network capacity has also been widely explored in an attempt to conserve capacity. It is possible to conserve network resources by aggregating data.

3 Problems in WSNs

WSNs have limited energy since nodes are powered by low power capacity. Due to the extreme environment and inaccessibility of the deployment region, these batteries are seldom rechargeable. Thus, maximizing the WSN lifetime requires saving energy for as long as feasible [20], [21], [22], and [23].

In this regard, many routing protocols that offer lifetime enhancement have been developed for WSNs because a remarkable amount of energy is drained by communication. The purpose of energy-aware routing methods is to reduce energy consumption in the whole network. This can be achieved by considering different aspects, including 1) decreasing the total energy exhaustion in the network, 2) decreasing the amount (or the distance) of wireless data transmission, 3) keeping the maximum possible number of alive nodes to achieve better lifetime, and 4) equally distribute energy consumption over the nodes in the network to prevent premature network breakdown caused by certain sensors that become out of energy [8]. Once the restricted energy supply is depleted, nodes cease operation and are referred to as "dead." In this circumstance, the network may be unable to complete its designated task or operate at maximum capacity. As a result, network longevity is critical for assessing the effectiveness of routing strategies [21]. Usually, in many data routing methods, an optimal path is constructed for data forwarding from the sender node to the sink. If the same founded path is used for data forwarding over and over aiming for fast data transmission, then the sensors included in that routing channel will rapidly deplete their energy. The disadvantage of these routing strategies is that they reduce the total energy consumption of a wireless network, but at the price of energy inconsistency. These approaches result in a network partition problem (i.e., two or more parts of the network become unreachable to each other) after particular sensor nodes run out of their battery capacity. This phenomenon may impair the usefulness and effectiveness of the whole network. Additionally, using complex algorithms for routing may reduce energy consumption, but this can make much processing delay[24]. Figure 2 illustrates the network partition problem (a set of nodes becomes unable to communicate with the sink) caused by the death of certain nodes that are the only connectors between the partitioned part and the sink[25].

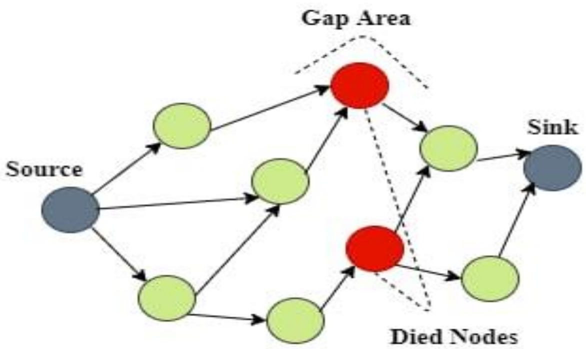

Figure 2: Network partition problems caused by the death of particular nodes[21]

In the routing phase, the radio transceiver consumes an amount of energy to release the data packet from the sender to the next Sensor nodes (or receiver). Furthermore, there will be a specific amount of energy that would be consumed to amplify the data packets to prevent the radio wave reflection and refraction in free-space propagation. All approaches in the forthcoming section were exposed to the same energy consumption model defined by W.B.Heinzelman which performs in the free space environment. The energy cost for transmitting an L-bits packet from any stationary sensor to its next hop is [25]:

$$E_{XT\,(L,\,d)} = LE_{elec} + LE_{amp}\,d^2 \tag{1}$$

The receiving cost can be computed using:

$$E_{RX\,(L)} = LE_{elec} \tag{2}$$

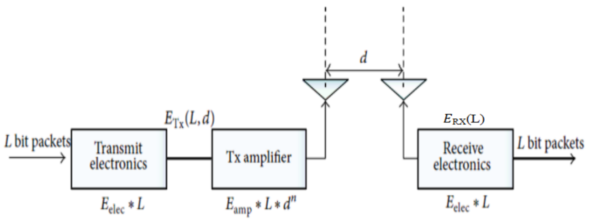

Figure 3: Energy Consumption Model[25]

Where: E_{TX}, E_{RX} represent the transmission energy and receiving energy respectively; L represents the length of the data packet that is desired to be transmitted by the sender or received by the receiver; d is the distance between the sender and the receiver circuits; E_{elec} represents the energy consumed in an electronic circuit; E_{amp} represents the energy depletion by amplifying the data signal. As shown in Fig. Figure 3, this illustrated the energy consumption model.

The power consumption model during the process of sending and receiving data shows that the data is directly proportional to the power consumption, which means that the higher the amount of data, the less power in the network and vice versa.

4 Proposed method

In the work proposed, we take into account a WSN with many sensors spread out in a specific diffusion area. At the time an event is detected, the CH is chosen using a heterogeneous network ordering technique [24], and the participating sensor nodes gather data about the surrounding environment and forward it to the CH. Having collected information, CHs must next send it wirelessly to the sink. Figure 4 is a high-level schematic depicting the process of data classification.

Sensor nodes Sensor1, Sensor2, and Sensor 3 collect Event1, Event2, And Event3 from the environment. Instead of sending each sensor node the data to the sink, it sends the data to the CH node's collection header. CH creates a single internal representation of the environment from its input. Then the individual actin is Submitted to sink.

There is a lot of data redundancy in WSN as we discussed in the previous section. To have a solution to this problem, we propose a data aggregation technique

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 10^3
S-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/m^2
	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/m^2
	Max. traffic in node's queue	50

that exploits the Support Vector Machine (KF- SVM) for a supervised learning model to eliminate redundancy, as shown in Figure 5. SVM performs two functions: one is to classify and one is to delete associated data. It uses a linear classifier method to represent the classification of participating sensor nodes and aggregation nodes (CH) using a hyperplane that divides it into two classes, redundant (class-1, red), and not redundant (class-1, blue).

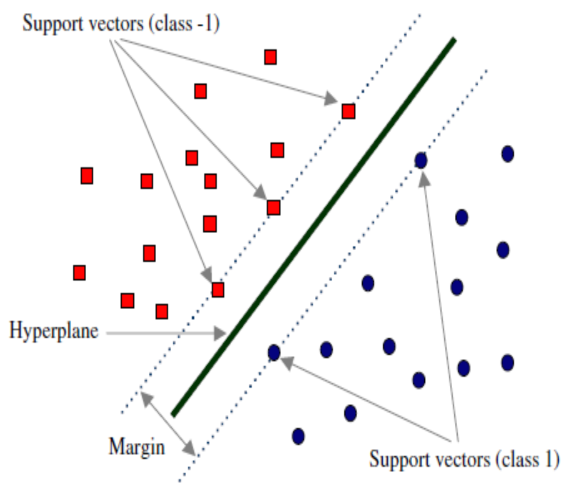

Figure 4: Support vector machine

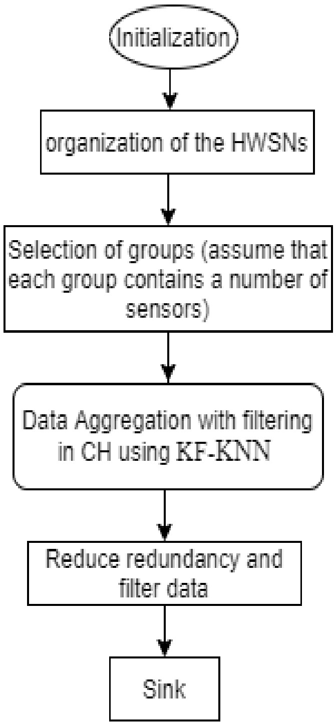

Figure 5: The proposed method of the KF-SVM

There are many different methods for fusing data, but KF is among the most used. It lessens background noise and provides precise estimates of the state vector that contains the relevant data. It has had widespread applications across a range of domains, including estimate, tracking, sensor fusion, etc. Prediction and sensor measurement updates derived from the system matrix of the initial state vector make up the KF framework.

The algorithm works in two stages. During the prediction stage, the Kalman filter produces estimates of the state variables and their associated uncertainty. These estimates are then updated using a weighted average, giving more weight to certain estimations, once the results of the following measurement (inevitably tainted with some inaccuracy, including random noise) are noticed. Thus, this proposed aggregation can help to obtain a better result in the power drain, reducing the delay, and improving the lifetime of WSNs.

5 Simulation and results

The lifetime of HWSNs can be extended by using the CHs for the KF-SVM hybrid data clustering and classification method and comparing it with the SVM. To see how well it works, it has been safely tested in terms of accuracy, error, time, and most importantly how much energy is left in the nodes, if the same routing metrics and the same environment are used in both.

We note that the proposed method KF-SVM is superior in performance compared to the previous work shown in Table 1, where the data were classified, aggregated, and noise removed, which led to an increase in the amount of energy remaining in WSN networks, which enhances the efficiency of the network and extends its life. As when you use SVM, it collects data and divides it into two important or non-important parts, and after the classification process, KF is used and thus filters the data collected in each CH, and sends it to the sink.

This section is divided into two parts, in the first part; the parameters used in the heterogeneous network are explained, with details of the simulation. In the second part, simulation results are presented.

5.1 Simulation setting

This subsection shows the parameters that should be set up to simulate the network to mimic the real-world network as possible. For all approaches, SVM, KF-SVM, 36 CHs, and 1000 S-sensors were randomly deployed in the topographical field with a dimension of 300x300 meters. Since this network is heterogeneous in terms of the initial energy levels, transmission range, and buffer size, Each S-sensor has initial energy of 0.5J while the batteries of CHs were equipped with an initial energy of 2.5J to be abundantly sufficient to fulfill the requirements of their role. Regarding the transmission range, the antenna of the CH can explore a radius of 80 meters while a radius of 20 meters is available with S-sensors. The buffer size of the S-sensor is [0,10], while the CH can buffer the packets within a size ranging from [0,50]. Each simulation for each approach has been executed for 2000 transmission rounds. In every round, all approaches generated one data packet which has a length of 2 kb. This simulation utilized a fixed area routing technique and has been designed with Matlab programming language. Table 3, shows more information about the simulation parameters. It is to be noted that the values of those simulation parameters are the same as the values that had been used in the original paper of the other method.

5.2 Simulation results

In this section, the KF-SVM proposed algorithm is put into every cluster head. Thus, we notice the effect of the algorithm on clusters only, instead of the normal sensors.

Clustering is extremely beneficial for applications that demand scalability to hundreds (or thousands) of sensors. In this context, scalability means the necessity for effective resource usage, data aggregation, and load balancing. Furthermore, several more routing protocols may utilize clustering to construct a hierarchy organization which leads to decreased path cost while conversing with the sink. The basic principle underlying clustering routing is that sensors only interact with a cluster head inside their cluster. Those CHs, which could be equipped with more specifications and less energy-constrained than "normal" sensors, then take the charge of transmitting data from their member sensors to the sink. This strategy can dramatically decrease the communication and energy overburdened at member S-sensors.

The network lifetime results obtained using the two methods are compared by counting the number of sensors that are kept alive after each data transmission round. On this point, Figure 6 demonstrates the ratio of the CHs, which are still alive in both the proposed

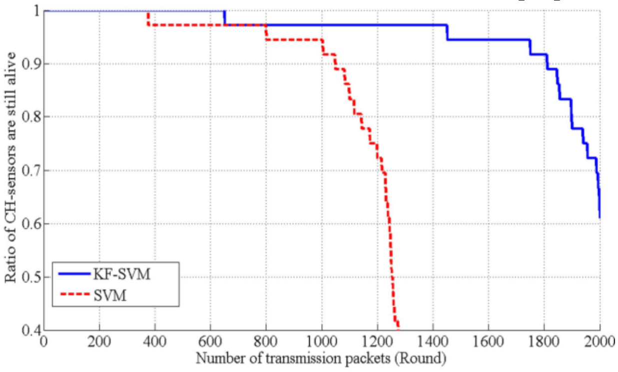

Figure 6: CHs ratio remains alive (three S-sensors sent at a time)

method (KF-SVM) and the SVM. As a result, the KF-SVM outperforms the SVM method based on the total number of nodes that are still alive in the network. Here and after sending (2000) packets to two sensors through the network, the result of the network lifetime achieved with the proposed is nearly (60%) more than that of SVM.

The proportion of leftover power in CHs varies with the number of transmission rounds depending on which of the two methods is employed. The proposed system outperformed the SVM in terms of overall performance and efficiency. Figure 7 illustrates how the ratio of residual energy for CHs, varies based on the transmission mode employed. As can see, the proposed method is better than the SVM while keeping the network stilling for an as long time as possible.

6 Conclusion

Sensor network data mining entails the accurate extraction of application-oriented models and patterns from a steady stream of rapidly changing data. In this case, all information must be processed immediately and none of it be stored. Data mining algorithms need to be able to process large amounts of data quickly. Multistep approaches and multicar mining algorithms are used by conventional data mining algorithms for the examination of static data sets. Typical data mining methods cannot be used for WSN data because of their massive size, high dimensionality, and dispersed nature. In this paper, a new method called Kalman filter with Support vector machine (KF-SVM) is introduced to classify and data aggregate and get rid of noise in WSNs, which enhances network efficiency and extends its lifetime. The simulation results of the proposed model indicate that KF-SVM outperformed in terms of greatly enhancing data latency reduction and lifetime maximization of the network. In future work, we propose a new technique using an intelligent timer protocol that controls the assembly process.

References

[1] J. Abdullah, M. K. Hussien, N. A. M. Alduais, M. I. Husni, and A. Jamil, "Data reduction algorithms based on computational intelligence for wireless

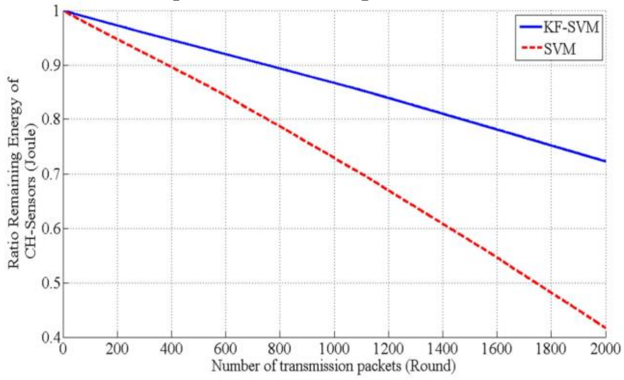

Figure 7: The energy ratio of the remaining CHs (three S-sensors sent

sensor networks applications," *ISCAIE 2019 - 2019 IEEE Symp. Comput. Appl. Ind. Electron.*, pp. 166–171, 2019, doi: 10.1109/ISCAIE.2019.8743665.

[2] J. Alkenani, K. A. Nassar, I. Technology, and C. Information, "Network Monitoring Measurements for Quality of Service : A Review," no. May, 2022, doi: 10.37917/ijeee.18.2.5.

[3] 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.

[4] 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.

[5] 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, doi: 10.1186/s13638-018-1176-4.

[6] J. Wang, L. Wu, S. Zeadally, M. K. Khan, and D. He, "Privacy-preserving Data Aggregation against Malicious Data Mining Attack for IoT-enabled Smart Grid," vol. 17, no. 3, 2021.

[7] 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, doi: 10.1109/ACCESS.2021.3051360.

[8] 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.

[9] L. N. Devi, G. K. Reddy, and A. N. Rao, "Live Demonstration on Smart Water Quality Monitoring System Using Wireless Sensor Networks," in *2018 IEEE SENSOR*, pp. 1–4, 2018.

[10] D. K. Altmemi and I. S. Alshawi, "Enhance Data

Similarity Using a Fuzzy Approach," *J. Posit. Sch. Psychol.*, pp. 1898–1909, 2022.

[11] Y. Wang, J. Wan, and J. Lai, "A Wireless Sensor Networks Positioning Method in NLOS Environment Based on TOA and Parallel Kalman Filter," in *2019 IEEE 19th International Conference on Communication Technology (ICCT)*, pp. 446–450, 2019.

[12] P. Patil and U. Kulkarni, "SVM based data redundancy elimination for data aggregation in wireless sensor networks," in *2013 international conference on advances in computing, communications and informatics (ICACCI)*, pp. 1309–1316, 2013.

[13] 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, doi: 10.12928/TELKOMNIKA.v20i1.20984.

[14] 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, doi: 10.1109/ABLAZE.2015.7155043.

[15] 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, doi: 10.1109/TNSM.2018.2884445.

[16] R. Maivizhi and P. Yogesh, "Spatial Correlation based Data Redundancy Elimination for Data Aggregation in Wireless Sensor Networks," *2020 Int. Conf. Innov. Trends Inf. Technol. ICITIIT 2020*, pp. 0–4, 2020, doi: 10.1109/ICITIIT49094.2020.9071535.

[17] 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, doi: 10.1109/IWCMC.2019.8766469.

[18] L. Krishnamachari, D. Estrin, and S. Wicker, "The impact of data aggregation in wireless sensor networks," in *Proceedings 22nd international conference on distributed computing systems workshops*, pp. 575–578, 2002.

[19] K. Maraiya, K. Kant, and N. Gupta, "Wireless sensor network: a review on data aggregation," *Int. J. Sci. Eng. Res.*, vol. 2, no. 4, pp. 1–6, 2011.

[20] Z. Nurlan, T. Zhukabayeva, M. Othman, A. Adamova, and N. Zhakiyev, "Wireless Sensor Network as a Mesh: Vision and Challenges," *IEEE Access*, vol. 10, pp. 46–67, 2021.

[21] 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.

[22] J. K. Alkenani and K. A. Nassar, "Network Performance Analysis Using Packets Probe For Passive Monitoring," *Informatica*, vol. 46, no. 7, 2022.

[23] J. Alkenani and K. A. Nassari, "Enhance work for java based network analyzer tool used to analyze network simulator files," vol. 29, no. 2, pp. 954–962, 2023, doi: 10.11591/ijeecs.v29.i2.pp954-962.

[24] M. S. Abdulridha, G. H. Adday, and I. S. Alshawi, "Fast simple flooding strategy in wireless sensor networks," *J. Southwest Jiaotong Univ.*, vol. 54, no. 6, 2019.

[25] 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. 10-pp, 2000.

[26] S. Khriji, G. Vinas Raventos, I. Kammoun, and O. Kanoun, "Redundancy elimination for data aggregation in wireless sensor networks," *2018 15th Int. Multi-Conference Syst. Signals Devices, SSD 2018*, pp. 28–33, 2018, doi: 10.1109/SSD.2018.8570459.

[27] I. Ullah and H. Yong, "Efficient data aggregation with node clustering and extreme learning machine for WSN," *J. Supercomput.*, no. 0123456789, 2020, doi: 10.1007/s11227-020-03236-8.

[28] 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.

[29] 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, doi: 10.1007/s11227-018-2642-9.

[30] F. Karray, M. Maalaoui, A. M. Obeid, A. Garcia-Ortiz, and M. Abid, "Hardware Acceleration of Kalman Filter for Leak Detection in Water Pipeline Systems using Wireless Sensor Network," in *2019 International Conference on High Performance Computing & Simulation (HPCS)*, pp. 77–83, 2019.

[31] 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*, pp. 866–870, 2013.

[32] 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, doi: 10.1007/s11277-015-2998-6.