INCREASE EFFICIENCY AND LIFETIME IN WSNS THROUGH DATA REDUCTION

Ali Abdulrazzaq Al-Sawad[1]
[1]Department of Computer Science, Shatt Al-Arab University College, Basra, Iraq

Article Info

Article history:

Received October 23, 2022
Revised November 28, 2022
Accepted December 1, 2022

Keywords:

Clustering
Data Collection
Kalman Filter
K-Nearest Neighbors
Routing

ABSTRACT

The mining and processing of data for wireless sensor networks (WSNs) have been a significant research area in a number of computer science disciplines, including distributed applications, database management systems, and information gathering. The main goal of implementing applications based on WSNs is to make real-time choices, which have proven very difficult due to the limitations of computing, communication, and data mining capacity for WSNs. Thus, due to the nature and unique characteristics of sensor data and the limitations of WSNs, traditional data mining methods cannot be easily applied to them. This paper presents a new approach, the Kalman Filter (KF) with K-Nearest Neighbors (KNN) named (KF-KNN) which is proposed for data classification and collection as well as noise elimination in WSNs, increasing the network efficiency and lifetime. The proposed technique is compared to KNN to illustrate the effectiveness of the recommended ways in increasing energy consumption and prolonging network lifetime.

Corresponding Author:

Ali Abdulrazzaq Abd Ali Al-Sawad
Department of Computer Science, Shatt Al-Arab University College, Basra, Iraq
Email:ali.abdulrazzaq@sa-uc.edu.iq

1. INTRODUCTION

The advancements in wireless communication and microelectronics have led to the creation of limited sensors and the application of large-scale sensor networks. The ability to conduct ubiquitous surveillance and sensor networks has garnered substantial attention in a wide range of application areas, including habitat monitoring, item tracking, environment monitoring, military, disaster management, and smart environments. In several applications, reliable real-time monitoring is a necessity [1]. These applications create a tremendous amount of dynamic, diverse, and geographically distributed data. Data mining may aid in automated or human-driven tactical/strategic decision-making if raw data is reviewed and transformed into relevant information. Therefore, it is essential to develop algorithms for mining sensor data for patterns so that rapid, intelligent decisions may be made [2].

In WSNs, computer power, communication capabilities, and power supply of sensor nodes are highly restricted, making it difficult or impossible to replace or recharge sensor batteries. In addition, big WSNs are installed randomly and intensively owing to the rise in data volume for two reasons. Because of the constant natural condition of the physical environment, the data collected by each sensor node are highly correlated and redundant. There is a substantial historical connection between the data of each consecutive sensor node. For instance, if sensor nodes acquire temperature data readings every five seconds, the temperature values may not change much. This means that it is not necessary to count the new reading every five seconds; otherwise, the previous reading will equal the actual reading. When sensor nodes are randomly and densely dispersed inside or near to a geographical phenomenon, a large volume of data is generated and made accessible for transmission as all sensor nodes in the area collect data. Each of these nodes sends a substantial amount of duplicate data in this circumstance. Another challenge and concern offered by WSNs is duplicated data. The similarity in the data perceived by a sensor is referred to as redundant data. Due to the redundant data process, sensor nodes squander the majority of their energy. However, numerous ways and procedures are employed to save energy. In general, data redundancy has a substantial impact on data quality [3].

The recent emphasis from the data mining community has been focused on the extraction of information from sensor data. On-sensor data, several clustering, association rule, common pattern, sequential pattern, and classification approaches have shown useful [3]. Because of their enormous size (up to tens of sensor network), random and hazardous deployment, loss of network scenario, restricted power supply, and high failure rate, the design and implementation of sensor networks provide unique obstacles to study. Traditional data mining approaches are inapplicable since they are centralized, computationally costly, and focused on disk-resident transactional data [4]. To process the data generated by sensor networks, new algorithms and modifications to certain data mining methods have been created. In the last decade, several methodologies, approaches, and algorithms for knowledge discovery have been proposed [5].

Because data mining is such a large topic, it may be used for data from any field; for more comprehensive studies on data mining techniques, see [6], which examined machine learning and data mining tactics for analyzing medical data. Given that this survey's categorization of data mining methods is based on frequent pattern mining, clustering, and classification, there are several types of research on each of these approaches. Specifically, [7] In calculating the performance of these networks' routing behavior, the lifespan factor is given significant weight. Now, researchers are focusing on strategies to lower network energy usage by proposing different kinds of routing protocols for WSNs. In order to choose the optimal network path for collecting and transmitting discovered data, it has been recommended that the hop distance of each route be lowered, also [8] describe frequent pattern mining across data streams. None of the aforementioned research, however, investigated data mining approaches that focus on the extraction and analysis of information from WSN data [9].

In this paper, a novel hybrid approach based on the Kalman Filter [10] and K-Nearest Neighbors [11] is presented called KF-KNN for data categorization, collection, and noise reduction in WSNs, which improves network efficiency and lifespan.

This paper analyzes algorithms and approaches designed exclusively for WSNs data, resulting in not only a separate categorization, assessment, and debate on diverse application areas, but also a diversity of solution alternatives. The objective is to determine how data mining methods will be utilized to construct sensor network applications intelligently. The remaining sections are grouped as follows: In the second part, previous work related to the work proposal is summarized. In the third part, the method of data collection in the WSN is described, with an emphasis on the need and significance of the collection. In the fourth part, a full description of the life of the sensor network and the challenges that contribute to energy loss are provided, along with a clarification of the calculation model for energy consumption. In the fifth part, the suggested model for data categorization and collection with noise suppression to improve network efficiency is provided. In the sixth part, parameters for the network simulation are presented, along with the results of the suggested method. In the last part, a summary of the research is offered.

2. RELATED WORK

Several academics have emphasized the issue of data categorization with routing in WSNs [6, 12]; the most challenging difficulty is finding methods to enhance energy efficiency so that the network can last for a much longer time [13], [14].

In [15], using the right methods for data collection can help reduce data repetition and improve efficiency. We propose the Spatial Correlation based Data Redundancy Elimination for Data Collection (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 collection ratio, data accuracy, and energy consumption. the outcomes favor SCDRE over other methods, and the results back this up. Nonetheless, energy consumption is mostly caused by the continual sensing and sending of data to sink nodes. Whether environmental items are moving swiftly or slowly, the data is the same or replicated in both circumstances, which increases transmission costs. To reduce transmission costs, several techniques or procedures, such as data aggregation and network hopping, are applied. Single hopping and multiple hopping are the two types of network hopping. In single hopping, sensors directly transfer data to the sink node. Due of distance, transmission costs grow. Single hop is inappropriate for expansive regions. Therefore, multi-hop is employed in expansive areas. Hierarchy routes, such as chain-based, tree-based, and group protocols, make extensive use of multi-hop. In [16], the authors proposed Multiple moveable troughs indicated methods for efficient data collection in HWSN. As a consequence, the fixed basin's energy use is inefficient. Utilizing a mobile sink to collect data conserves energy, hence extending the network's lifespan.

In [17], data from WSNs may be decreased using incremental Naive Bayes Prediction, while data collection has been recommended using compressive sensing technology in [4] and [18], where active sensor nodes are improved using particle swarm optimization to minimize the quantity of duplicate data. Fuzzy Dstar-Lite was suggested by the authors of [14] as the optimal information routing strategy for HWSNs. In addition, it emphasizes the importance of surpassing the blockage example and clarifies the UED issue. Open mining is proposed in [19] as a strategy for data aggregation that is both efficient and cost-effective. This data mining technique employs several. Each has a central node around which a multitude of virtual pits gather and transmit information to the sink. The proposed in [20], The data collection strategy emerges as a significant method for reducing the energy consumption of sensor nodes and enhancing bandwidth utilization. REDA is a Redundancy Elimination Data collection algorithm based on a pattern generation methodology. The 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 collection methods. Moreover, in comparison to existing data collection algorithms, ESPDA and SRDA are superior.

The proposed [21], Kalman Filter (KF), lessens 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. The proposed in [11] a fitness function has been developed to choose the optimal position of the sink node with high residual energy from neighboring sensor nodes, hence increasing network longevity. Eventually, experimental findings were obtained when sensor nodes were distributed at random inside the specified network region. The solution improves the network's energy usage by 11%, which enhances the network's lifespan[22].

3. DATA COLLECTION IN WSNs

This part addresses the problems associated with reducing data duplication in WSNs. Data redundancy occurs when a single item is repeated more than twice. The similarity is also known as the precise value. During the sensing process, redundancy is discovered when sensor nodes detect a physical item. There are approximate redundancy difficulties with WSNs, notably in hostile or extreme environments where sensors cannot be replaced or recharged. In contrast, another difficulty associated with WSNs is big data, since hundreds of sensors gather and aggregate data across a large region, producing a substantial amount of big data. WSNs are now one of the primary sources of big data in IoT because sensors collect a large quantity of data every minute before transferring it to the base station. Nevertheless, Big data processing is difficult to handle. The various degrees of redundancy included in the cluster-based design for WSNs are outlined. Clustering, processing, data analysis, and energy conservation are some of the most significant data difficulties in WSNs [23].

Data collection is the examination of the properties of raw data and the application of correlations. Using a data collection strategy, sensor nodes digest raw data before providing it to the sink. As a result of the digest's smaller size, data collecting decreases transmission costs and network congestion. We propose that collecting data is a crucial strategy for lowering energy usage in WSNs [24]. Before data collecting performance can be enhanced, however, various hurdles must be addressed.

Furthermore, several collection 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 collection 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 [15], [23], 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 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 30 minutes, if not an hour. Additionally, they are when two sensor nodes are used to measure temperature in the same room. Figure 1. (b), data obtained by one node is frequently comparable to, if not identical to, data gathered by another node[25].

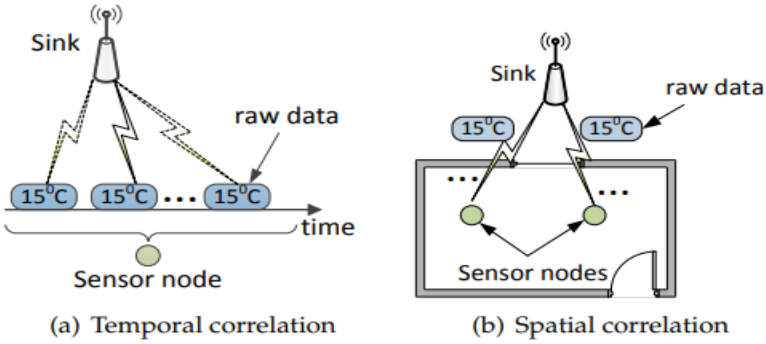

Figure 1. Data Correlation: (a) Temporal correlation; (b) Spatial correlation[26]

As a general rule, a collection protocol should accomplish the following three basic goals[27], [28]: Energy-saving: Data collection 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 collection 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.

4. THE LIFETIME 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 [29], [30].

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 [7]. 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 [30].

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. 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[31]. 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[32].

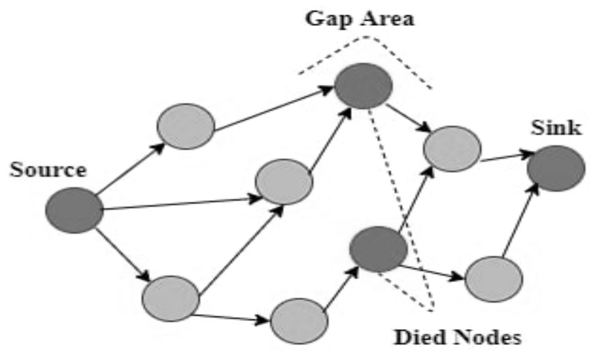

Figure 2. Network partition problems caused
by the death of particular nodes[30]

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-bit packet from any stationary sensor to its next hop [32], As shown in Figure 3, illustrated the energy consumption model.

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

The receiving cost can be computed using:

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

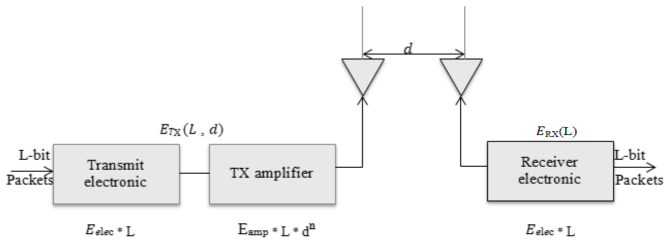

Figure 3. Energy consumption model [32]

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.

The energy consumption model during the process of transmitting and receiving data shows that the amount of data transmitted greatly affects the energy of the sensors, which means that the higher the amount of data, the less energy in the network.

5. 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.

Sensor nodes SN1, SN2, SN4, and SN4 collect Message1, Message2, Message3, and Message4 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.

Therefore, large datasets thus include a range of information, some of which is helpful while others are entirely unnecessary. Assuming that two sensors communicate four events simultaneously or near, this might lead to data redundancy difficulties. These issues can only be resolved via the most efficient usage of energy. Thus, the data are effectively aggregated in the cluster header before being sent to the 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 collection technique that exploits the (KF- KNN) for a supervised learning model to eliminate redundancy. The KNN algorithm believes that comparable objects are located nearby. In other words, comparable objects are close together. Similar data points are nearby. The usefulness of the KNN method depends on this assumption being true enough. KNN encapsulates the concept of similarity (sometimes called distance, proximity, or closeness). As shown in Figure 4, the proposed method of the KF- KNN.

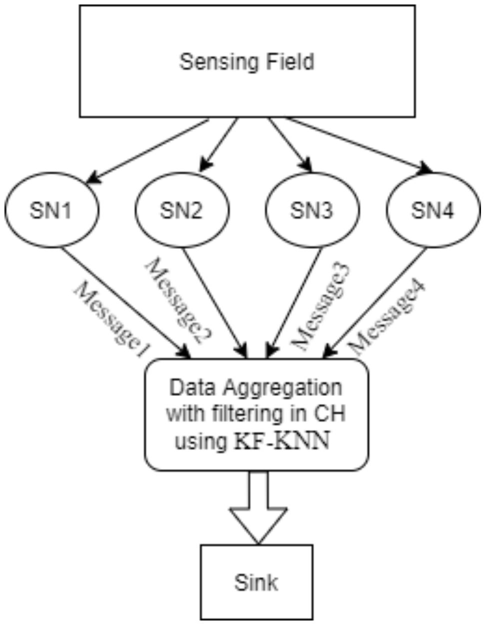

Figure 4. The proposed method of the KF- KNN

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 collection can help to obtain a better result in the power drain, reducing the delay, and improving the lifetime of WSNs.

6. PERFORMANCE EVALUATION

The lifetime of HWSNs can be extended by using the CHs for KF- KNN hybrid data clustering and classification method and comparing it with the KNN. To see how well it works has been safely tested nodes that survive and, most importantly how much energy is left in the nodes if the same routing metrics and the same environment are used in both. This section is divided into two parts, in the first part; the parameters used in the heterogeneous network are explained [13], with details of the simulation. In the second part, simulation results are presented.

6.1. Simulation Setup

MATLAB is used to run the simulations. Several system settings must be configured so that the network is as realistic as feasible. Table 6 displays a heterogeneous network with 1000 normal sensors and 36 CHs randomly positioned inside a 300 m × 300 m square region of topography.

The clustering strategy clusters normal sensors around CHs. Each system using the [32] radio paradigm has exhausted its broadcast cycles (2000). Each strategy generates a packet length of 2 KB. In contrast, all Normal sensors and CHs begin with identical initial energies of (0.5 J) and (2.5 J), with a sensed transmission of (20 m) and (50 m) (80 m). The normal sensors load should be generated at random from 0 to 10. Every CHs has a [0...50] range.

Table 1. Simulation parameters

Parameter		Value
Area of topographical (meters)		300 m x 300 m
Location of the sink (meters)		(0, 150)
Length of control packets		2k
No. of transmission packets (rounds)		2×10^3
Normal-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 the node's queue	10
Cluster-head-sensors	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 the node's queue	50

6.2. Simulation Results

In this part, the effort to offer an energy-efficient KF-KNN data-gathering approach employing cluster-based WSNs to quantify Spatial and temporal correlation data is described in depth. The duplex approach is used to discover duplicate data and minimize transmission inside the cluster. Then, sensors are organized into clusters to detect and transmit redundant data between sensor nodes. The inter-cluster transmission then begins as CH gathers data from all cluster member nodes using KF-KNN. Finally, the basin collects all the data received from the groups using the way described.

Consequently, we only see the influence of the algorithm on clusters and not on standard sensors. Counting the number of sensors that remain active after each cycle of data transmission is used to compare the network lifespan findings generated by the two techniques. At this stage, Figure 5 depicts the proportion of CHs that are still alive in both the suggested approach and the baseline technique (KF- KNN). Consequently, the KF-KNN algorithm beats the KNN algorithm based on the total number of nodes that are still active in the network. Here, and after delivering 2,000 packets to two sensors across the network, the network lifetime attained with the suggested is almost (52%) more than that of KNN.

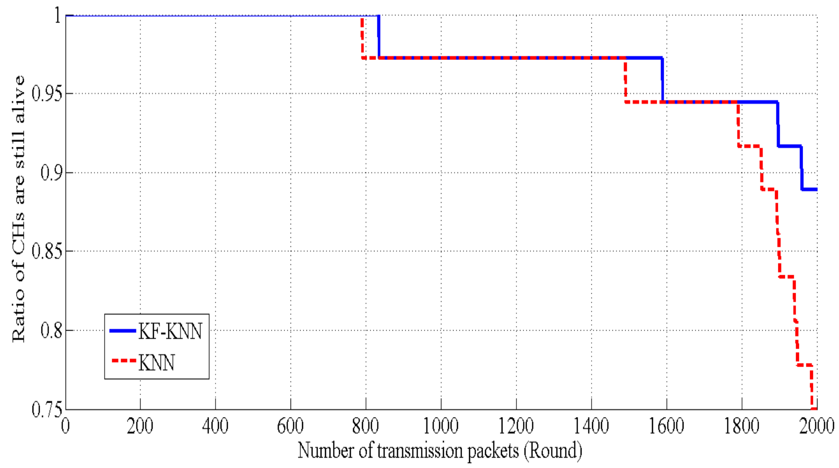

Figure 5. CHs ratio remains alive

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 KNN in terms of overall performance and efficiency. The cluster head is elected periodically to enhance the energy drainage, fairly distribute traffic load on sensor nodes, and to give better scalability, especially for WSNs with growing size. Another advantage of clustering is that the cluster head can perform aggregation on the gathered data to avoid redundant information from being sent to the sink, which makes the WSNs function more efficiently in terms of energy saving. Figure 6 depicts how the ratio of remaining energy for CHs fluctuates depending on the transmission mechanism used. Clearly, the suggested technique is superior to the KNN while maintaining the network's stability for as long as feasible.

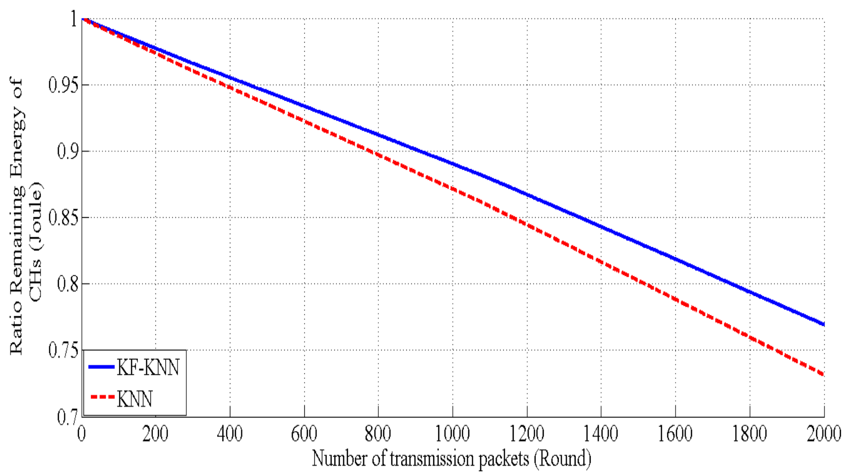

Figure 6. The energy ratio of the remaining CHs

7. CONCLUSION

Data mining in a wireless sensor network entails the precise extraction of application-oriented patterns and patterns from a steady stream of rapidly changing data. In this case, unnecessary data should be disposed of immediately and none of it is stored. Thus, data mining algorithms must be able to process large amounts of data quickly. Multi-step approaches and multi-component mining algorithms are used by traditional data mining algorithms to examine data sets. Typical data mining methods cannot be used for WSN data due to their massive size, high dimensionality, and sparse nature. In this paper, a new method called Kalman filter with K-Nearest Neighbors (KF-KNN) is presented to classify, aggregate data, and eliminate noise in WSNs, which enhances network efficiency and extends its life. The results of the proposed model reveal that KF-KNN greatly outperforms KNN in in reducing the amount of data and increasing the network efficiency and lifetime. 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 sensor networks applications,” *ISCAIE 2019 - 2019 IEEE Symp. Comput. Appl. Ind. Electron.*, pp. 166–171, 2019, doi: 10.1109/ISCAIE.2019.8743665.

[2] 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.

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

[4] 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.

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

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

[7] 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.

[8] D. K. Altmemi and I. S. Alshawi, “Enhance Data Similarity Using a Fuzzy Approach,” *J. Posit. Sch. Psychol.*, pp. 1898–1909, 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 SENSORS*, 2018, pp. 1–4.

[10] 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.

[11] M. M. Ahmed, A. Taha, A. E. Hassanien, and E. Hassanien, “An optimized k-nearest neighbor algorithm for extending wireless sensor network lifetime,” in *International conference on advanced machine learning technologies and applications*, pp. 506–515, 2018.

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

[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] 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. 1000–1010, 2020.

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

[16] A. Muthu Krishnan and P. Ganesh Kumar, “An Effective Clustering Approach with Data Collection 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.

[17] 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, doi: 10.1109/INFOP.2015.7489415.

[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 collection approach for WSNs based on open pits mining,” *Wirel. Networks*, vol. 27, no. 1, pp. 41–53, 2021.

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

[21] 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.

[22] 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.

[23] S. Kumar and S. Kumar, “Data collection 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.

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

[25] M. R. Choudhari and U. Rote, “Data Collection Approaches in WSNs,” *2021 Int. Conf. Comput. Commun. Informatics, ICCCI 2021*, pp. 27–32, 2021, doi: 10.1109/ICCCI50826.2021.9402430.

[26] 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.

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

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

[29] 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.

[30] 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.

[31] 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.

[32] 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 science*, pp. 10-pp, 2000.