Hybrid Kalman Filter and Optimization-Based Routing for Energy Efficiency in Heterogeneous Wireless Sensor Networks

Ali K. Marzook[1], Jawad Alkenani[2]*
[1] University of Basrah- College of Engineering- Department of Electrical Engineering
[2] Department of Computer Science, Shatt Al-Arab University College, Basra, Iraq
E-mail: ali.marzook@uobasrah.edu.iq[1], Jawadalkenani@sa-uc.edu.iq[2]
*Corresponding author

Keywords KF-BA, network lifetime, routing, BA, wireless Sensor Networks

Received: September 15, 2024

Several significant research studies have been done in distributed applications, database management systems, and information collecting in computer science concerning data mining and processing for wireless sensor networks (WSNs). As a result of WSNs' limited computation, networking, and data mining capabilities, the primary objective of creating WSN-based applications has proven to be extremely challenging: making decisions in real time. As such, typical data mining techniques are difficult to apply to sensor data due to their nature, peculiarities, and the constraints of wireless sensor networks. This work introduces a novel method for data mining and gathering and noise removal in wireless sensor networks (WSNs), dubbed KF-BA. This methodology increases network lifetime and efficiency by combining the Kalman Filter (KF) with the Bat Algorithm (BA). The recommended methods and BA contrast the provided techniques for raising energy consumption and prolonging network lifetime. The BA algorithm is not as effective as the KF-BA approach. The suggested strategy yields network longevity that is almost (57%) longer than BA in this instance, and that is after 2,000 packets are sent to two sensors spread over the network. The results were simulated using the MATLAB environment.

Povzetek: Predlagana metoda KF-BA združuje Kalmanov filter in Bat algoritem za izboljšanje energetske učinkovitosti in podaljšanje življenjske dobe brezžičnih senzorskih omrežij. Rezultati kažejo na 57 % daljšo življenjsko dobo omrežja v primerjavi z BA pristopom.

1 Introduction

Wireless communication and microelectronics developments have led to the creation of small-scale sensors and large-scale sensor networks. The ability to conduct ubiquitous surveillance and sensor networks has attracted a lot of interest. Reliable real-time monitoring is necessary for several applications [1]. An enormous amount of dynamic, varied, and widely dispersed data is produced by these applications. When raw data is analyzed and turned into pertinent information, data mining may help with both automated and human-driven tactical and strategic decision-making Consequently, intelligence algorithms for pattern recognition in sensor data must be developed to enable quick decision-making.[2]. WSNs are an attractive field of research with the objective of great performance in terms of technical expression and proportional upgrading. The tactics or approaches for reducing data redundancy in IoT, cloud computing, and smart networks should be as straightforward as feasible. They should need less calculation, processing, and transmission to minimize energy consumption and optimize data accuracy. There are several difficulties and restrictions. IoT applications also improve wireless sensor networks' longevity and energy efficiency by supporting scalability, mobility, node localization, dynamic environments, and user satisfaction.

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. Each sensor node collects data, which is highly correlated and redundant because of the physical environment's unchanging natural state. The data from each subsequent sensor node has a significant historical correlation with one another. For example, temperature values may not vary significantly if sensor nodes obtain temperature data readings every five seconds. This indicates that the prior reading will equal the current reading if the new reading is not counted every five seconds. A significant amount of data is produced and made available for transmission when sensor nodes are placed randomly and densely inside or close to a geographical phenomenon. This is because every sensor node in the vicinity gathers data. Each node 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 most of their energy.

Nonetheless, a variety of techniques and methods are used to conserve energy. Data redundancy significantly affects data quality [3]. In the data mining community, information extraction from sensor data has received a lot of interest lately [3]. Because traditional data mining techniques are limited to disk-resident transactional data, computationally expensive, and centralized, they are not suitable [4]. Some data mining techniques have been modified, and new algorithms have been developed to handle the data produced by sensor networks. Numerous techniques, strategies, and algorithms for knowledge discovery have been put forth throughout the past ten years. [5].

Since data mining is such an expansive field, it may be used for any kind of data; for further in-depth studies on data mining techniques, see [6], which examined data mining techniques for medical data analysis as well as machine learning. There have been several forms of study on each of the three data mining method categories in this survey, which include frequent pattern mining, clustering, and classification.

In paper [7] Both network users and network service providers must comprehend the quality of service (QoS) provided by networks to evaluate how effectively. Next-generation monitoring systems are required due to strict network requirements. These systems must be able to detect quickly falling network performance as well as the underlying reason for poor service quality. To address this problem, a novel fuzzy logic-based method is developed. As a result, the method was assessed and contrasted using Bayesian classification, probabilistic neural networks (PNNs), and network performance metrics including packet loss, delay, and jitter. While the fuzzy technique typically performed better than the PNN and Bayesian approaches, all approaches properly identified the QoS categories. 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.[8]. Explain regular pattern mining in several data sources. Nevertheless, none of the research mentioned above examined data mining techniques that focus on obtaining information and examination from WSN data [9]. This research presents a novel hybrid strategy called KF-BA for data collecting, classification, and noise reduction in WSNs. It is based on the Bat Algorithm (BA) [11] and Kalman Filter (KF) [10]. This approach increases network lifespan and efficiency.

This research looks at algorithms and techniques designed especially for WSN data, providing a range of possibilities for solutions and unique categorization, assessment, and discourse on different application fields.

The purpose of gathering data and aggregating it via a sophisticated transmission technology is to lengthen its lifespan. Discovers relationships between the data dynamically gathered from each sensor using the sample rate. This method operates in two-phase rounds: data aggregation, data collection, frequency adjustment, and selective transmission. Each round is further broken into four phases. Every sensor node measures distance using dynamic time warping to collect data samples. Data sampling makes use of both data collection and data transfer. To eliminate duplicate data from sensors while transmitting data to the base station, data sampling and data transmission are both utilized. The redundant data aggregates data first before transmission. Every sensor node has a defined sample rate throughout the data collection phase to get adoptive data. Secondly, the protocol employs a symbolic aggregate approximation method during the data aggregation phase to remove unnecessary data from temperature readings before sending them to the CH [6,12].

Scalability in WSNs is enhanced with energy efficiency through the preparation of a hierarchical architecture. The common architectural technique, known as clustering, improves scalability by substituting multiple-hop transmission for single-hop transmission. A periodic sensor network is meant to be divided into various clusters if the clustered-based architecture is considered. Every cluster has a CH that accepts data from its member nodes and handles additional transmission. Redundant data are those that are sensed by its member nodes when the environment remains constant [12]. The member node compiles and sends the data to the CH. The CH then receives various data sets from every member node. Due to proximity and fluctuating climatic conditions, redundant data is more likely to exist. This duplicate data causes high traffic, high workload, significant memory loss, and fast depletion of sensor batteries among cluster member nodes [12].

As hierarchical routing techniques for WSNs and provide a comparison of the benefits, drawbacks, and performance concerns of each protocol for routing in [13], [14]. The principles of using hierarchical energy-efficient routing protocols are provided by a detailed examination of the conventional and swarm intelligence methodologies. The improved design of hierarchical routing protocols considers many factors, including query-based routing, data aggregation, scalability, fault tolerance, location awareness, load balancing, QoS, and energy efficiency. A comprehensive inventory of unresolved problems for further studies is offered, encompassing difficulties related to objectives, approaches, categories, performance metrics, and additional subjects. Similarly, the lifetime energy consumption of WSNs is studied.

The following categories apply to the remaining sections: A summary of earlier research on the work proposal is provided. Similarly, the lifetime energy consumption of WSNs is studied. The relevant works' approach, performance, and outcomes are compiled in Table 1. The significance of research through data collecting, energy conservation, and smartphone distribution design is covered in the second section. The WSN data-gathering technique is explained in the third section, with a focus on the importance and need of data collection. The fourth part presents the suggested model for data categorization and collection with noise reduction to improve network efficiency. The fifth section presents a performance simulation of the suggested model and explains the key findings. The last section provides an overview of the research.

Table 1: Summarization table on the related works

Ref	Methodology	Performance/Results
[15]	• **SCDRE**	The outcomes confirm that SCDRE should be preferred over other methods. 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 to distance, transmission costs grow. Single hop is inappropriate for expansive regions. Multi-hop is therefore used in large areas.
[16]	• **Mobile Sink**	The authors suggested several movable troughs as effective ways to gather data in HWSN. As such, the energy consumption of the fixed basin is inefficient. By collecting data using a mobile sink, you may save energy and increase the longevity of the network
[17]	• **INBP**	The proposed study uses incremental naive Bayes prediction in conjunction with data mining in a wireless sensor network to eliminate duplicate data based on prediction. As a result, fewer data entities need to be sent to the sink. This is advantageous because it reduces the energy needed to transmit data to the sink. The INBP model is contrasted with two methods: the regular transmission model and the basic naive Bayes prediction model. Data from weather forecasts is utilized as input in this project. The suggested work significantly reduces energy consumption and extends the sensor network's lifespan.
[18]	• **IBOA**	IBOA on WSN life cycles and energy usage. In addition, a location- and direction-based low-power routing method has been created to enhance WSN's functionality and routing effectiveness. By precisely delivering messages to the designated nodes, this technique greatly increases data packet transmission speeds while using around 0.21% less energy than previous methods.
[19]	• **Open Mining**	Open mining is suggested as an economical and effective approach to data collection. This data mining method makes use of many. Each of them has a central node around which many virtual pits congregate and send data to the sink. The data-gathering strategy.
[20]	• **REDA**	A pattern-generating methodology serves as the foundation for the Redundancy Elimination Data collecting algorithm or REDA. The pattern uses differential data obtained from subsequent sensor node iterations and is specific to the sensed data. As such, redundant data transfer between sensor nodes in the same cluster and the corresponding cluster head (CH) is prevented in every iteration. The performance evaluation shows that when compared to protocols without data-gathering methods, the REDA algorithm can save up to 44% of energy consumption. Furthermore, ESPDA and SRDA are better than current data collection techniques.
[21]	• **Kalman Filter**	To reduce the environmental impact of the planned Kalman Filter, we have created and implemented a hardware accelerator for it. The method's hardware and software versions have been compared in terms of space requirements, time to completion, power consumption, and other parameters. The outcomes show that energy consumption and execution time were reduced by almost 97% while there was no appreciable increase in the area.
[22]	• **LBCMRP**	suggested a Clustering Multipath routing strategy for heterogeneous sensor networks that is based on load balancing. Following an explanation of the protocol's operation, the outcomes of the two protocols Distributed Energy efficient Adaptive Clustering Protocol and Weight-based clustering in wireless sensor networks were compared. As a result, we consider several measures for evaluating performance. Additionally, we observe that, while keeping the average data transmission time constant, the suggested protocol outperforms DEACP by 47% in terms of first node death and WBCHN by 35% in terms of last node death.
[23]	• **CNN**	• Computer vision applications have demonstrated the promise of Convolutional Neural Networks (CNNs), one of the AI learning models. Nevertheless, several variables, including the CNN architecture, learning rate, activation function, and pooling strategy, affect CNN performance. To identify DR in diabetic patients, the best model may be utilized to create IoT-based smart devices.
[29]	• **KF-SVM**	A multitude of information can be found in large datasets, some of it unnecessary and others useful. Performance is negatively impacted by this redundancy in terms of redundant transmission and computing costs. On the other side, redundant data in a network may be removed by data aggregation. This research introduces a novel technique called Kalman filter with a Support vector machine (KF-SVM) to categorize, aggregate data, and eliminate noise in wireless sensor networks (WSNs). This improves network lifetime and efficiency.

2 Importance of paper

The subject of WSNs research is highly motivating, since it aims to achieve both excellent technological performance and proportionate modernization. To be used with IoT, cloud computing, and smart networks, data redundancy reduction strategies and approaches should be as straightforward as feasible, requiring the least amount of calculation, processing, and transmission to save energy consumption and improve data accuracy. Limitations and difficulties exist. IoT applications improve the lifetime of wireless sensor networks and save energy by supporting scalability, dynamic environments, mobility, node location, and user satisfaction. An expressive literature survey and a categorization method for data redundancy reduction are briefly presented in this paper about energy efficiency in WSNs. As a result, existing schemes for reducing data redundancy and improving energy efficiency in WSN research remain contradictory. In recent years, there has been a growing interest in the possibilities for sensor collaboration in data collection, processing, and sending to the sink. On the other hand, the resource-constrained nature of sensor nodes creates many challenges in developing, running, and maintaining sensor networks in the real world. Energy consumption is a crucial factor to consider when designing WSNs. So, a lightweight system is favored to conserve the sensor node's energy and preserve data protected from various vulnerabilities. Additionally, multiple obstacles complicate energy-efficient communication in WSNs. Therefore, using energy-efficient routing protocols, and balancing energy consumption overseer all the nodes are the most significant challenges to be overcome to extend the network lifetime [24]. 1- Many services are hampered by the continuing difficulties in implementing networks. Many variables can be manipulated, and others cannot be manipulated. For this reason, properties affecting network deployment must be identified and controlled. 2- Data collection, communication quality, and data processing. Other things to think about are in terms of energy conservation, node state, and transmission method. 3- WSNs need a way to balance the power consumption on all nodes so that these nodes consume their entire energy and die at about the same time. 4- Next, it is required to maintain data confidentiality using the best lightweight block encryption with a routing protocol to achieve data confidentiality, integrity, and secure routing when transmitting sensor information from source nodes to the sink. While sensor networks are comparable in many ways to other distributed systems, they face a unique set of challenges and limitations. These constraints influence the design of a network, resulting in protocols and algorithms distinct from those of other distributed systems. The following subsections highlight some of the most significant WSN difficulties that have been resolved using some algorithms.

- **Energy consumption**

WSNs have some design challenges, including energy efficiency. Sensing, connectivity, and data processing are the three aspects of power usage that need improvement. The sensor node's lifetime is often influenced by battery capacity. When it comes to sensor networks, the most prevalent stumbling block is constrained energy budgets. Sensors are frequently powered by batteries, which must be replaced or recharged regularly. A rechargeable sensor node should be able to function until its job is completed or the battery must be replaced. The mission duration depends on the application.

- **Data aggregation**

Given that radio transmission is the primary energy consumption, one way to reduce this consumption is to reduce communication overhead via data aggregation. Rather than transmitting every individual node measurement to the sink, intermediary nodes consolidate the raw data into a manageable quantity of data packets containing meaningful information. By eliminating repetitive and unneeded data readings, data aggregation seeks to avoid redundant packet transfers and thereby reduce communication energy consumption. The ratio of data reduction to data aggregation determines data accuracy. Data correctness is a key element of information quality in WSNs. In WSNs, data redundancy arises from complicated data processes and poor transmission. A significant amount of data must be accessible to preserve data correctness, and the amount of data redundancy reduction must be fixed scientifically to be measured.

- **Design and deployment**

In some applications of WSNs, the distribution of sensor nodes over the area of interest is done according to a plan, contrary to the random deployment where sensors are scattered by throwing them on the area (e.g. dropped from an airplane). The design process in WSNs has the aim of specifying the type, quantity, and placement of sensor nodes to be deployed in an environment to have complete knowledge of their functional status. Computational Intelligence (CI) techniques can be useful to handle the designing and planning of WSNs deployment.

3 Data collection In WSNs

The challenges of minimizing data duplication in WSNs are covered in this section. A single item is considered redundant data if it is repeated more than twice. Another name for the similarity is the precise value. Redundancy is found when a physical object is detected by sensor nodes during the sensing process. WSNs have approximate redundancy issues, especially in harsh or extreme situations when sensors are not resupplied or recharged. On the other hand, as hundreds of sensors collect and aggregate data across a wide region, creating a significant amount of big data, big data is another challenge that comes with WSNs.

Since sensors gather a lot of data every minute and send it to the base station, WSNs are currently one of the main sources of big data on the Internet of Things. However, handling big data processing might be challenging. The cluster-based design for WSNs is described with the different levels of redundancy included. Among the most important data challenges in WSNs are clustering, processing, data analysis, and energy saving [23].

Examining the characteristics of unprocessed data and using correlations is known as data collecting. Sensor nodes process raw data using a data-collecting technique before sending it to the sink. Data collection reduces network congestion and transmission costs because the digest is smaller. We suggest that one of the most important tactics for reducing energy consumption in WSNs is data collection [24]. However, several obstacles need to be overcome before data collection performance can be improved.

Moreover, several collection procedures are limited in their versatility since they are tailored for specific data types (like temperature data) or network features (like grid networks). As such, we are driven to create aggregating algorithms that are neutral concerning both data and properties.

In a WSN, sensor nodes constantly send all raw data to the sink in the absence of data collection. Even though this information is typically relevant or duplicated, it has the following drawbacks: 1) redundant data is useless for the application; 2) there is a significant increase in the probability of network congestion; 3) network capacity is wasted; and 4) energy consumption increases proportionately. Previous studies [15], [23], show that raw data is often the basis for temporal and spatial correlations.

For a given sensor node, data acquired at different time intervals are correlated temporally, whereas data obtained from nearby sensor nodes is correlated spatially. As can be observed in Figure 1. (a), the results produced for temperature monitoring by a sensor node often stay constant for as long as thirty minutes, if not an hour. Furthermore, they occur when two sensor nodes are utilized to take a room's temperature. Data collected by one node is often like, if not identical to, data collected by another node (see Figure 1. (b)[25].

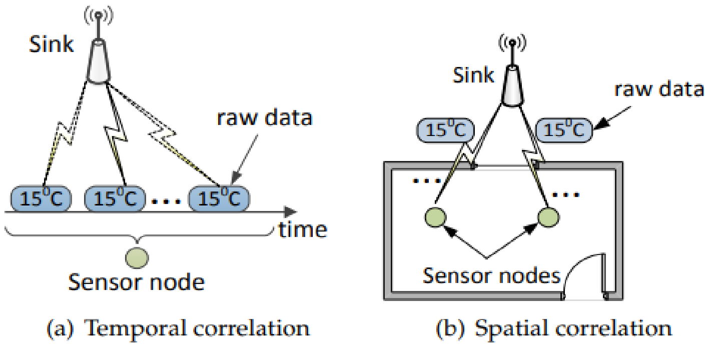

Figure 1. Data Correlation: (a) Temporal correlation. (b) Spatial correlation[26]

A collection protocol should achieve the following three fundamental objectives Energy-efficient: Data collection reduces a network's energy consumption by doing away with redundant connections and related communications. Energy conservation must be given top priority in the data-collecting structure since it is the main limitation of WSNs.

Data accuracy: Data precision is the degree of correctness between the recovered and original data. When raw data is combined by sensor nodes into a digest, some information might be lost. As a result, it is reasonable to anticipate some deviation from the raw data when recovering the sink-side data. The capacity of an application to precisely and satisfactorily reduce its energy usage is a universal requirement.

Saving network capacity: WSNs' network capacity has also been extensively investigated to conserve capacity due to sensor node bandwidth limitations. By aggregating data, network resources can be spared.

4 Proposed method

In this section, the proposed model which represents a new approach, namely Kalman Filter (KF) with Bat Algorithm (BA) called (KF-BA) is presented which is proposed for data mining and collection as well as noise removal in wireless sensor networks, which increases the network efficiency and lifetime.

a) Kalman filter

It is also referred to as a linear quadratic estimation (LQE), which starts with a collection of noisy and inaccurate data collected from WSNs. By adjusting certain system settings, it generates an estimate that is more accurate than the original data set by utilizing probability functions. It can be thought of as an ideal estimator with some noise and parameters derived from data that are not entirely certain. In general, it is optimal for the Kalman filter to minimize the mean square error of estimated parameters when the noise is Gaussian [10]. There are two phases in this method. Prediction, in which the subsequent actions are forecast based on the original datasets, and updating, in which the time step is used to measure the parameters of the current state [10].

- **Prediction**

$$\overline{x}_1 = a_{1-k}\ x_{1-k} + b_1 u_1 \quad (1)$$
$$\overline{p}_1 = a_{1-k}\ p_{1-k}\ a_{1-k}^t + q_{1-k} \quad (2)$$

At time step l, where before seeing the measurement, $\overline{x}_1$= the anticipated mean of the state and $\overline{p}_1$= Predicted covariance of the state, before seeing the measurement.
x_1 = estimated mean of the state and p_1= Estimated covariance of the state, after seeing the measurement.

- **Update**

$$v_1 = y_1 - h_1 \overline{x}_1 \quad (3)$$

y_1=. The measurement's meaning. v_1 = innovation or residue from measurement

$$s_1 = h_1 \overline{p}_1\ h_1^t + r_1 \quad (4)$$

s_1= measuring prediction covariance

$$k_1 = \overline{p}_1\ h_1^t\ s_1^{-1} \quad (5)$$

k_1= The filter gain indicates how much the forecasts need to be adjusted.

$$x_1 = \overline{x}_1 + k_1 v_1 \quad (6)$$

$$p_1 = p_1 - k_1 s_1 s_1^t \quad (7)$$

b) Bat algorithm (BA)

A wide variety of evolutionary or swarm-based artificial intelligence (AI) approaches have gained popularity for addressing challenging combinatorial optimization issues in the real world with the introduction of rapid computer resources.

The Bat Algorithm (BA) is a that mimics the echolocation behavior of bats. This bio-inspired search strategy has demonstrated the ability to reach potential regions. It is based on bats' social behavior and echolocation phenomenon, which allows bats to determine the distance between their current location and their prey. It is a technique for numerical optimization that is simple, straightforward to implement, substantially faster than existing algorithms, and robust. [27]. It has been used to handle a variety of optimization problems in diverse fields, including energy and power systems, distributed resource allocation, image processing, economic load dispatch, load frequency management, medical applications, engineering design, and distribution network reconfiguration (DNR), among others. It is ideally suited for highly nonlinear situations and can create optimal solutions for such issues with high precision. [27].

Bats vary in size and weight, yet they all behave similarly when it comes to foraging and navigation. In complete darkness, microbats primarily rely on echolocation to locate prey and avoid obstacles. Microbat behavior has been conceived as a new method of optimization. To create an algorithm capable of both local exploitation and global exploration, BA combines a population-based algorithm with a local search method. [27].

All bats in BA fly in random directions at a certain frequency and velocity at a certain place. The bats are made to travel randomly by use of two control parameters: loudness and Pulse Emission Rate (PER). Bats supply time-varying values to these parameters, so changing their location and velocity. Bats produce pulses and adjust PER in response to the presence of prey. A bat foraging behavior is characterized using an optimization model. There are two components to the description of bat foraging: random fly and local random walk (LRW), which are discussed in more depth below. [28].

- **Random fly**

In a D-dimensional issue search space, each bat is specified by its position. x_b, velocity v_b, frequency f_b, loudness A_b, and PER r_b. Using (8)-(10), the velocity and position of the *b*th bat at the *tenth* iteration are updated. These equations illustrate BA's global search procedure.

$$f_b = f_{min} + (f_{max} - f_{min})\beta \quad (8)$$

$$v_b^t = v_b^{t-1} + (x_b^t - x^*) f_b \quad (9)$$

$$x_b = x_b^{t-1} + v_b^t \quad (10)$$

Thus, β represents a uniformly distributed random number in the interval [0, 1], x^*Is the current best solution, f_{min} and f_{max} Re the minimum and maximum frequency ranges, respectively.

- **Local random walk (LRW)**

A new solution is created for each bat using LRW, represented as

$$x_{new,b} = x_{old,b} + \epsilon\langle A_b^t \rangle \quad (11)$$

Where $\epsilon \in [-1,1]$ is a uniformly distributed random number, $\langle A_b^t \rangle$Is the average loudness value for all bats, and r_b^tIt's the PER of the b^{th} bat.
The loudness and PER of each bat are modified iteratively using (12) and (13).

$$A_b^{t+1} = \alpha A_b^t \quad (12)$$

$$r_b^{t+1} = r_b^0 \, [1 - \exp(-\gamma t)] \quad (13)$$

where α and γRe constants having a value of 0.9. r_b^0How is the maximum possible PER? As the algorithm progresses, loudness approaches zero and PER approaches its maximum value. The pseudo-code of standard BA is presented in Algorithm. The pseudo-code of standard BA is described in Algorithm 1.

The research that is given examines a WSN with many sensors dispersed within a certain diffusion region. The participating sensor nodes collect and transmit environmental data to the CH upon detection of an event; the CH is selected using a heterogeneous network ordering algorithm [24]. CHs must wirelessly send the data they have collected to the sink.

Sensor nodes SN1, SN2, SN4, and SNn collect messages 1, 2, 3, and n from the environment. Every sensor node sends data to the collection header of the CH node instead of the sink. Based on its input, CH generates a single internal representation of the surroundings.

Large datasets therefore contain a variety of information, some of which are very useful and others which are superfluous. Data redundancy issues could arise if two sensors send four events at the same time or very close together. The most economical use of energy is the only way to address these problems. Consequently, before being delivered to the sink, the data are essentially aggregated in the cluster header. As we saw in the previous section, the wireless sensor network has a lot of redundancies in the data. To address this problem, we introduce a data collection method using (KF-BA), as shown in Figure 2, the proposed KF-BA strategy.

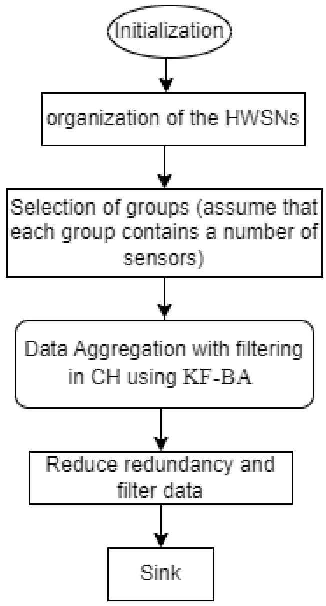

Figure 2: The proposed method of the KF-BA

Although there are numerous methods for combining data, KF is among the most used. It produces precise estimations of the state vector containing the relevant data while attenuating background noise. It has been widely used in several domains, including tracking, sensor fusion, estimate, and more. The system matrix of the initial state vector provides the prediction and sensor measurement updates that make up the KF framework.

The algorithm operates in two stages. The Kalman filter generates values of the state parameters and the related uncertainty during the prediction step. When the results of the subsequent measurement which are unavoidably contaminated with some inaccuracy, including random noise are observed, these estimations

Algorithm 1: Pseudo-code of the original Bat algorithm

```
Ascertain the goal function of (.)
Initialize the bat population x_b, v_b and f_{b,d}, ∀ b = 1,2,...,N_b, d = 1,2,3,....,D
Initialize pulse rate r_b and loudness A_b
Set t = 0
while t < itr_max do
 Create new solutions by varying the frequency and modifying the locations and velocities.
 using (8)–(10).
 if rand > r_b then
     Select the current best solution x*
     Generate a local solution OF(x_b) around x*
 end if
 Create a fresh solution by soaring at random.
 if rand < A_b and OF(x_b) < OF(x*) then        ▷ for minimization
     Accept the new solutions
     Increase r_b and reduce A_b
 end if
 Sort the bats and identify the best one right now.
     t = t + 1
end while
return the top bat and its corresponding fitness
```

are then updated using a weighted average, giving some estimates greater weight. As a result, this suggested collection can aid in achieving a better outcome in terms of power drain, decreasing latency, and extending WSN lifetime.

5 Performance evaluation

The lifespan of HWSNs may be increased by using the CHs for the KF-BA hybrid data clustering and categorization technique and comparing it with the BA. The number of securely tested nodes and, crucially, the remaining energy in the nodes when identical routing metrics and environments are employed in both cases are the metrics used to assess its effectiveness. This section is divided into two sections. The first one explains the heterogeneous network's simulation parameters and details. The second section displays the simulation's findings.

- **Simulation setup**

The simulations are conducted using MATLAB. The network needs to be set up in a few different ways to make it as realistic as possible. Table 2 shows a heterogeneous network consisting of 1000 normal sensors and 36 CHs arranged at random within a 300 m x 300 m square topographic area. Normal sensors are clustered near CHs using the clustering approach. Every radio-based system has used up all its broadcast cycles (2000). The packet length produced by each technique is 2 KB. On the other hand, starting from the same initial energies of (0.5 J) and (2.5 J), all Normal sensors and CHs sense transmissions of (20 m) and (50 m) (80 m). A random number between 0 and 10 should be used to create the typical sensor load. Each CH has a range of [0...50].

Table 2: 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 x 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 Heads	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

- **Simulation results**

This section provides a detailed description of the endeavor to provide an energy-efficient KF-BA data-gathering approach using cluster-based WSNs to quantify spatial and temporal correlation data. To identify redundant data and reduce transmission inside the cluster, the duplex technique is employed. Subsequently, sensor nodes are grouped into clusters to identify and share redundant data. Next, the inter-cluster transmission starts as CH uses KF-BA to collect data from every cluster member node. Lastly, the basin uses the method outlined to gather all the data it has received from the groups.

As such, we observe the algorithm's effect only on clusters, not on conventional sensors. The network longevity results produced by the two methods are compared by counting the number of active sensors following each data transmission cycle. As of right now, Figures 3 and 4 show the percentage of CHs and normal sensors that are still alive in the baseline technique (KF-BA) and the recommended strategy. Accordingly, when considering the total number of nodes in the network that are still active, the KF-BA method outperforms the BA algorithm. The network lifespan obtained with the proposed method is nearly (57%) more than that of BA in this case, and after sending 2,000 packets to two sensors throughout the network.

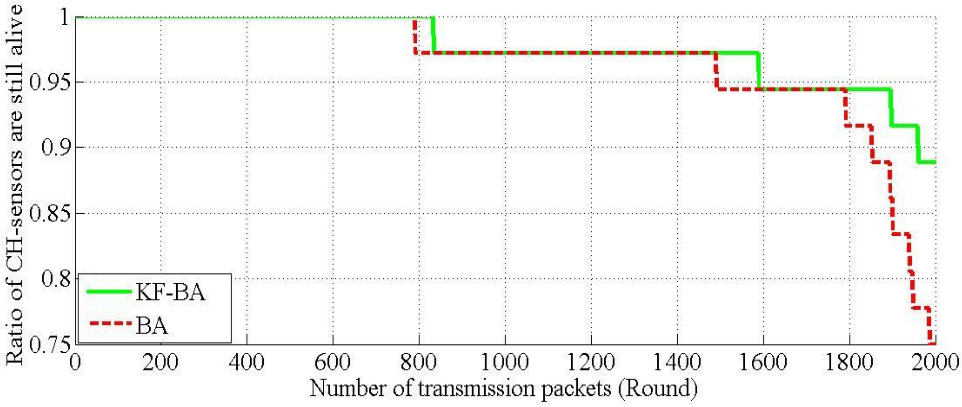

Figure 3: CHs ratio remains alive

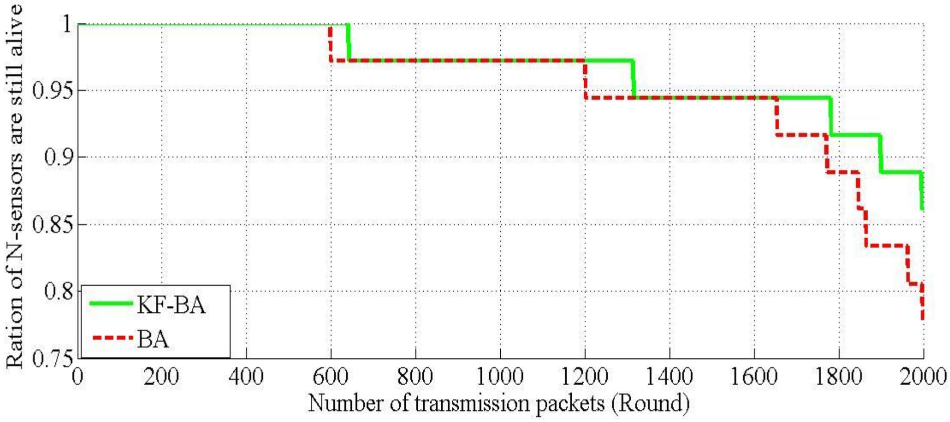

Figure 4: normal-sensors ratio remains alive

Depending on which of the two approaches is used, the percentage of residual power in CHs changes with the number of transmission rounds. In terms of efficiency and overall performance, the suggested system fared better than the BA. Periodically electing the cluster head improves energy drainage, evenly distributes traffic load among sensor nodes, and provides better scalability, especially for WSNs that are getting larger. Another benefit of clustering is that duplicated information can be prevented from being sent to the sink by the cluster head performing aggregation on the collected data, which improves the energy-saving efficiency of WSNs.

The variation in the ratio of leftover energy for CHs based on the transmission mode employed is illustrated in Figures 5 and 6. The recommended method outperforms the BA and keeps the network stable for as long as is practical.

Figure 7 illustrates how the suggested solution reduces end-to-end latency. Shorter wait times mean less energy wasted and more efficient data transmission. This means that to minimize network congestion and increase network lifetime, data packets are separated at the node level via multipath routing.

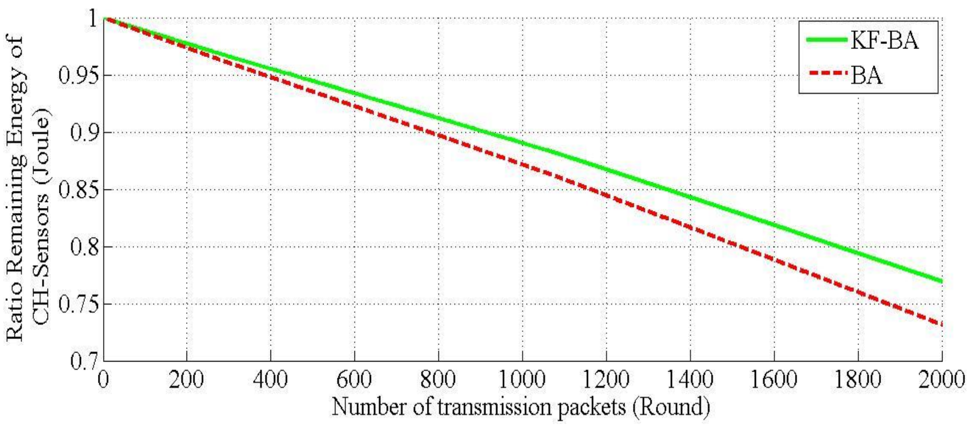

Figure 5: The energy ratio of the remaining CHs

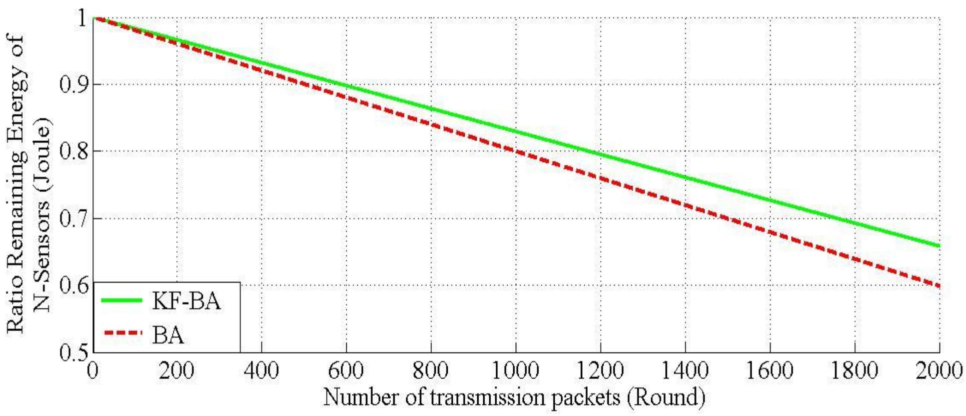

Figure 6: The energy ratio of the remaining normal sensors

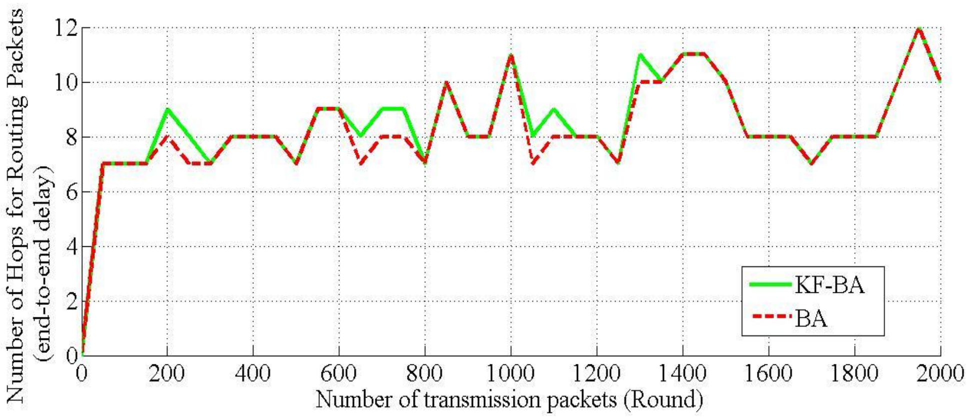

Figure 7: the number of hops end-to-end latency

The superiority of the hybrid algorithm (KF-BA) in the field of wireless sensor networks is due to several factors:
1. Energy efficiency: The Bat algorithm helps in optimizing energy consumption, which extends the life of the network.
2. Improvement in network distribution: It contributes to the optimal distribution of nodes for better coverage of the target area.
3. Adaptation to dynamic environments: It can quickly adapt to changes in the environment, which enhances the reliability of the network.
4. performance improvement: It achieves better results in reducing delay and increasing data transfer speed compared to other algorithms.

The hybrid algorithm has good features in its work as The Kalman filter is a powerful tool used for state estimation and data processing in wireless sensor networks. It enhances the accuracy of estimates by integrating multiple measurements with mathematical models of motion or state changes. By relying on previous information, it corrects current estimates, thereby improving the reliability of the collected data.

Additionally, the Kalman filter mitigates the effects of noise that measurements may encounter, such as interference or unstable environmental conditions. It uses a mathematical model to determine when and how to enhance the data, allowing for the filtering of incorrect measurements.

The filter operates in real-time, enabling it to continuously update estimates as data flows in. This capability makes it ideal for applications that require rapid responses, such as localization and motion tracking.

Regarding energy consumption, the Kalman filter helps reduce the need for frequent measurements, leading to lower energy usage. By improving estimates, sensors can operate less frequently, which enhances overall energy efficiency.

Moreover, the Kalman filter is adaptable to changes in the environment or system characteristics. It can adjust its models based on new data, making it effective in unstable conditions. It can also be easily integrated with other optimization algorithms, further enhancing overall system performance.

In summary, the Kalman filter offers multiple advantages in wireless sensor networks, including improved accuracy, noise reduction, energy efficiency, and real-time operation. These features make it an excellent choice for applications requiring high precision and reliability.

As for the Bat Algorithm can also handle multiple objectives simultaneously, making it useful for resource allocation in sensor networks. In the context of wireless sensor networks, it optimizes the placement and operation of sensors to minimize energy consumption, thus extending the network's lifespan. It enhances data collection efficiency by improving the arrangement of sensors, ensuring that data is gathered from the most relevant areas while reducing redundancy.

Furthermore, the algorithm is adaptable to changes in the environment, such as variations in sensor performance or the emergence of obstacles, making it suitable for dynamic scenarios. It can efficiently manage a large number of sensors, providing scalability for extensive networks. Additionally, the Bat Algorithm is resilient to noise and uncertainties in sensor measurements, which is crucial for maintaining performance in real-world conditions.

Applications of the Bat Algorithm in wireless sensor networks include optimizing sensor deployment for maximum coverage and minimal energy use, determining the best nodes for data aggregation, and enhancing routing protocols based on sensor locations and energy levels.

In summary, the Bat Algorithm offers an effective and innovative approach to solving optimization problems in wireless sensor networks, enhancing both performance and sustainability.

When comparing the Bat Algorithm with other algorithms used in wireless sensor networks, the following aspects can be focused on: The following Table 3 shows the most important features that were compared. These factors make the duck algorithm an attractive option for wireless network applications, especially in environments with energy and resource constraints.

Table 3: Comparison with the algorithms

Performance Improvement	Energy Consumption	Adaptability to Changes
KF-BA: It effectively balances exploration and exploitation, enhancing data transmission efficiency	**KF-BA**: Known for its energy efficiency, it helps prolong the network's lifespan	**KF-BA**: Quickly adapts to environmental changes, improving network stability.
Bee Algorithm: While it is flexible and adaptable, it may struggle with resource utilization in certain scenarios.	**Bee Algorithm:** It can be less energy efficient due to its iterative nature and the need for multiple evaluations.	**Bee Algorithm**: Suitable for problems of finding optimal solutions in dynamic environments, where criteria can change rapidly.
GA: It may be less efficient at times due to the need to evaluate multiple generations, but it has the potential to explore space on a large scale.	**GA**: Generally, consumes more energy due to complex operations like crossover and selection.	**GA**: The ability to adapt to changes is slower, as it is used in allocating frequency resources and reducing interference between users in WSN.
PSO: Effective in exploring the research space quickly and efficiently, helping to achieve a good balance between performance and power consumption.	**PSO**: It relies on a set of interacting particles, which enhances the efficiency of energy consumption during the search process.	**PSO**: May take longer to adjust, which can impact performance in a dynamic environment.

6 Conclusion

The exact extraction of patterns and application-oriented patterns from a continuously streaming, quickly changing set of data is known as data mining in a wireless sensor network. In this instance, all superfluous data should be deleted right away and should not be saved. As a result, data mining algorithms need to be fast at processing big volumes of data. Traditional data mining methods analyze data sets using multi-component mining algorithms and multi-step methodologies. WSN data is too big, too dimensional, and too sparse to be mined using standard data mining techniques. This research presents a novel approach to categorizing, aggregating, and removing noise in wireless sensor networks (WSNs). The method, known as the Kalman Filter (KF) with the Bat Algorithm (BA), increases network lifespan and efficiency. According to the suggested model's results, KF-BA performs significantly better than BA in terms of data reduction, network lifetime, and efficiency. In subsequent work, we suggest a novel method that employs an intelligent timer protocol to regulate the assembly procedure.

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,doi: 10.1109/ACCESS.2021.3128353.

[3] Chovanec, Michal, Michal Hodon, and Lukas Cechovic. "Tiny low-power WSN node for the vehicle detection." *Informatica* 38.3 ,2014.

[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,doi: 10.1145/3440249.

[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] Altmemi, Dhuha Kh, Abdulmalik Adil Abdulzahra, and Imad S. Alshawi. "A new approach based on an intelligent method to classify the quality of service." *Informatica* 46.9, 2022, doi: 10.31449/inf.v46i4.4323.

[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*, pp. 1–4, 2018,doi: 10.1109/ICSENS.2018.8589931.

[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,doi: 10.1109/ICCT46805.2019.8947129.

[11] Jagadeesh, K., et al. "Efficient Path Planning for Wireless Sensor Networks: Minimizing Exposure with a Modified Bat Algorithm." *2024 International Conference on Inventive Computation Technologies (ICICT)*. IEEE, 2024, doi: 10.1109/ICICT60155.2024.10544622.

[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] Hussein, Dheyaa Mezaal, et al. "Lifetime Maximization Using Grey Wolf Optimization Routing Protocol with Statistical Technique in WSNs." *Informatica* 47.5,2023, doi: 10.31449/inf.v47i5.4601.

[14] Alkenani, Jawad, and Khulood Ahmed Nassar. "Network Performance Analysis Using Packets Probe For Passive Monitoring." *Informatica* 46.7 ,2022,doi: 10.31449/inf.v46i7.4307.

[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] Sathish Kumar, br. "An Energy-Efficient Cluster Routing Protocol Based On Improved Butterfly Optimization Algorithm." *Adhoc & Sensor Wireless Networks* 59,2024, doi: 10.32908/ahswn.v59.11363.

[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,doi: 10.1007/s11276-020-02442-9.

[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, doi: 10.1109/HPCS48598.2019.9188156.

[22] Gherbi, Chirihane, and Doudou Roumaissa. "IoT-based traffic congestion control for environmental applications." *Informatica* 45.7 ,2021, doi: 10.31449/inf.v45i7.3671.

[23] Jena, Manaswini, et al. "Exploring the parametric impact on a deep learning model and proposal of a 2-branch CNN for diabetic retinopathy classification with case study in IoT-Blockchain based smart healthcare system." *Informatica* 46.2,2022, doi: 10.31449/inf.v46i2.3906.

[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.
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] Mohar, Satinder Singh, Sonia Goyal, and Ranjit Kaur. "Optimized sensor nodes deployment in wireless sensor network using bat algorithm." *Wireless Personal Communications* 116.4, pp 2835-2853, 2021,doi: 10.1007/s11277-020-07823-z.

[28] Tang, Jun, Gang Liu, and Qingtao Pan. "A review on representative swarm intelligence algorithms for solving optimization problems: Applications and trends." *IEEE/CAA Journal of Automatica Sinica* 8.10 (2021): 1627-1643, doi 10.1109/JAS.2021.1004129.

[29] Kheerallah, Yousif Abdulwahab, and Jawad Alkenani. "A new method based on machine learning to increase efficiency in wireless sensor networks." *Informatica* 46.9,2023, doi.org/10.31449/inf.v46i9.4396.