Received: February 7, 2024. Revised: March 19, 2024.


International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

# CCOA-DC: A Novel Optimization with NMF Data Compression in WSN Data Aggregation

**Mays Kareem Jabbar[1]*** **Thaar A. Kareem[1]**

[1]*Department of Faculty of Engineering, University of Misan, Al Amarah City, Misan Province, 62001, Iraq*
* Corresponding author's Email: m_mays85@uomisan.edu.iq

**Abstract:** Wireless Sensor Networks (WSNs) play a pivotal role in remote monitoring, surveillance, and Internet of Things (IoT) applications. The efficient utilization of battery-powered sensor nodes in WSNs, given their limited power capacity, is crucial for successful data transmission. Conventional clustering algorithms while efficient in clustering, often lacks the efficient management of data generated by sensor nodes, leading to redundant data in applications like IoT leading to reduced network lifetime. To overcome this issue, this paper introduces a novel approach, named CCOA-DC (Improved Coati Optimization with Cognitive Factor (CCOA) through Data Compression (DC)), in clustering heterogeneous aggregated WSN data. The research unfolds in two novel phases. Initially, a non-negative matrix factorization (NMF) model is introduced data compression for clustering, addressing the challenge of data transmission and energy efficiency. Subsequently, the performance is enhanced through load balancing, featuring dynamic cluster head selection via Improved Coati Optimization (COA) with cognitive factor (C), denoted as CCOA. A distinctive aspect is the incorporation of the NMF data compression technique in both clustering and cluster head selection processes, introducing an energy-efficient, load-balanced, and compressed data aggregation mechanism. The proposed CCOA-DC undergoes rigorous testing, comparing its performance against existing models to validate its superiority. Comparative analyses with renowned models such as TCBDGA, HEED, and FEEC-IIR underscore the distinct advantages of CCOA-DC. Notably, it achieves a reduction of 78.57% of packet loss ratio compared to FEEC-IIR model. The model achieves high packet delivery ratio which is 98.67%, and shows optimized energy consumption of 68.01% Joules. This novel compression-based metaheuristic data aggregation algorithm showcases its effectiveness in addressing the energy conservation challenge, affirming its prominence in the area of WSNs based IoT applications.

**Keywords:** Wireless sensor networks, Data aggregation, Clustering, Non-negative matrix factorization, Coati optimization algorithm, Data compression, Energy consumption, Load balancing.

## 1. Introduction

The deployment of Wireless Sensor Networks (WSNs) has witnessed remarkable growth across diverse domains, driven by advancements in sensors, processors, and wireless communication technologies [1]. These networks play a pivotal role in applications such as industrial monitoring, healthcare, environmental sensing, and military operations [2]. In WSN, the sensor nodes are highly distributed with limited energy sources and composed of various elements, including transceivers, memory, and microcontrollers. The main challenges of this system are data aggregation, energy efficiency, scalability, security, and load balancing among these heterogeneous nodes (nodes with varying states). Collecting and organizing data from various sensor nodes before transmitting it to the sink or next forwarding node is crucial in WSN and termed as Data Aggregation. However, sensor nodes require immense energy to deliver a large number of packets. If the load on a sensor node continuously increases, the sensor's energy depletes faster and it dies soon. In order to maintain the system's efficiency, the network's traffic load has to be balanced, termed as load balancing. In most of the studies, this issue of load balancing is addressed through clustering the

Received: February 7, 2024. Revised: March 19, 2024. 

Table 1 Overview of the Related works

| Reference | Optimization Algorithm | Residual Energy | Distance | Load balancing | Trust |
|---|---|---|---|---|---|
| [1] | RFPT Approx. Algorithm | ✓ | ✗ | ✗ | ✗ |
| [2] | GWO | ✓ | ✓ | ✓ | ✗ |
| [3] | Memetic-Based Approach | ✓ | ✗ | ✗ | ✗ |
| [4] | IIoT Model | ✗ | ✗ | ✗ | ✓ |
| [5] | Feature Extraction (NMF) | ✗ | ✗ | ✗ | ✓ |
| [8] | IDAF-FIT Algorithm | ✓ | ✓ | ✗ | ✗ |
| [9] | Fuzzy Logic for Routing | ✓ | ✓ | ✗ | ✗ |
| [10] | F-LEACH | ✓ | ✓ | ✗ | ✗ |
| [11] | Fog-Assisted EHDA Scheme | ✓ | ✓ | ✗ | ✗ |
| [14] | Hybrid EPO Algorithm | ✓ | ✗ | ✗ | ✗ |
| [15] | HFLSBC | ✓ | ✓ | ✗ | ✗ |
| [17] | Energy-Efficient CH Selection | ✓ | ✗ | ✗ | ✗ |
| [18] | FEWO Algorithm | ✓ | ✗ | ✗ | ✗ |
| [19] | DDCA-WSN | ✓ | ✗ | ✗ | ✗ |
| [20] | CRSH | ✓ | ✓ | ✗ | ✗ |
| [21] | EEC-MA-PSOGA | ✓ | ✓ | ✗ | ✗ |

nodes and efficient Cluster Head (CH) selection. The main goal of the proposed clustering algorithm is to carefully select the CH with sufficient energy such that the delay of the network is reduced with increased overhead, this prolongs the lifetime of network. This section focuses on reviewing variety of optimization algorithms and strategies that have been proposed to enhance load balancing through clustering approach to drive the motivation to contribute in this field. The related studies ensure prolonged network lifespan and improved performance based on the employed varying network parameters as objective functions.

Several energy-efficient routing techniques have been proposed for WSNs so far. The clustering and data aggregation strategy is the most efficient among these approaches. The presence of a cluster head in a WSN necessitates additional energy consumption and decreases the network's overall lifespan. This phenomenon is commonly referred to as the Load Balancing Clustering Issue (LBCI). In order to address this matter, an FPT-approximation technique called RFPT was proposed to determine the optimal route from CH to the sink [1]. However, Gateways located in close proximity to the sink node experience faster depletion of energy due to heavy traffic load. It is necessary to position gateways at a greater distance from the sink to address this issue. By employing two innovative fitness functions in the optimization algorithm, the lifespan of the network can be enhanced and reduced energy depletion [2]. While the clustering strategy is effective in minimizing the network's energy usage, it is plagued with energy gaps and uneven load distribution. In order to address the issue of clustering, researchers devised a memetic algorithm and a load-balancing clustering scheme [3]. The research has discovered security concerns in the Industrial Internet of Things (IIoT). To tackle these issues, a model based on WSN was designed to enhance the security performance of IIoT [4]. While these studies ensure less energy consumption to improve the network lifetime, the cost of the energy required to store and process the data among the network becomes high and affects the network lifetime. To overcome this issue, methods like in [5-7, 11, 13, 8, 19] introduce the use of data compression techniques, as noted in Table 1 as the overview for the related work reviewed here. For instance, in [13], the concept of data compression is used to compress the data sent to the sink by sensors. The primary objective of data compression is to transform unstructured or redundant data from various sensor nodes into a more organized and interpretable format suitable for network communication. The compression process helps alleviate the strain on communication resources, increases sensor storage capacity, and ultimately enhances the overall lifespan and performance of the WSN.

The Table 1, provides an overview of the related works in the field of WSN summarizing key aspects of each paper, including the utilization of data compression, optimization algorithms employed, and the objective functions employed for Cluster Head (CH) selection. The majority of papers (e.g., [1-3, 8-11, 14, 15, 17-21]) consider residual energy as a crucial factor for CH selection. This aligns with the common objective of prolonging network lifetime by selecting nodes with higher energy reserves. However, some papers do not prioritize residual energy and consider other factors like distance and trust.

Several works, such as [1-3, 8, 9, 22], have employed optimization algorithms like RFPT Approximation, Grey Wolf Optimization (GWO), Memetic-Based Meta-heuristic, IDAF-FIT Clustering, and Fuzzy Logic, respectively. These algorithms aim to enhance clustering efficiency, CH selection, and routing by considering parameters like efficient routing algorithms, residual energy, distance from the base station, and node degree for CH selection. However, these algorithms have their own advantages and disadvantages. For instance, EPO with ASO improves network life and reduces packet delay but does not prevent node capture attacks [14]. Additionally, algorithms like WOA-P are not universally applicable to all clustering techniques or to dense node scenarios [23]. To overcome this, recent studies have adopted the hybrid optimization approaches, as noted in Table 1. Hybrid techniques leverage the strengths of multiple algorithms to address the limitations of individual methods. For instance, the Hybrid Emperor Penguin Optimization (EPO) algorithm combines the benefits of both EPO and Ant Swarm Optimization (ASO) [14]. Similarly, a Hybrid Extended Multi-sink and Anycast Routing (EMPAR) integrates various strategies to improve network lifetime and load balancing simultaneously [13]. Hybrid optimization is favoured for its ability to offer a comprehensive solution, combining different algorithms' strengths to achieve superior load balancing and overall network optimization performance. Considering this, the present study is motivated to implement a hybrid approach for the dynamic selection of CH, capitalizing on the enhanced optimization capabilities of the coati optimization algorithm (COA) in diverse problem spaces.

Another challenging task is to identify the parameters for CH selection; from the literature, it is seen that common objective functions for CH selection involve factors such as residual energy, distance metrics, and node-specific characteristics. Residual energy, a fundamental parameter in WSNs, is consistently considered in works like [1, 2, 8, 9, 17, 21]. It directly influences the lifespan and performance of sensor nodes, making it a crucial metric for CH selection. Additionally, distance-related metrics, such as the distance to the base station or proximity to the CH, are commonly employed, as evident in [2, 9, 15, 22]. The consideration of node-specific attributes, like mobility, node degree, and energy levels, is also prominent in works like [3, 8, 15]. These factors collectively contribute to the overall efficiency and load balancing in WSNs. Hence, the study is motivated to propose an improved optimization-based clustering scheme where CH selection is based on multi-objectives based on the energy of the nodes as well as the consumption of energy by the sink.

It is also noted from the related works as in Table 1, the recent studies are limited as the nature of the nodes becomes different and distributed. The cost of the energy is still affected and ultimately affects the communication in Distributed data in WSN [19]. The motivation behind employing data compression in WSNs is to mitigate issues related to communication overload, energy consumption, and network lifetime [5-7]. Data compression techniques, particularly Non-Negative Matrix Factorization (NMF), are used to transform unstructured data into a more interpretable format suitable for communication. As seen in [5-7, 19], is chosen for its ability to handle the spectral representation of data and efficiently identify faulty nodes, reducing false measurements and improving overall network lifetime. Hence, considering this, the study is motivated to use the Matrix factorization-based technique to compress the data at the required position to improve the process of CH selection and reduce energy costs.

The key contributions to the energy consumption reduction in WSN data aggregation in this manuscript are:

- Employing Non-Negative Matrix Factorization to cluster nodes, emphasizing uniform node density for enhanced efficiency.
- Implementing a lossless matrix factorization approach, the aim is to compress aggregated data at the cluster head (CH), with the goal of reducing CH energy consumption and extending the overall system lifetime.
- Introduce a multi-objective problem formulation for the selection of an optimal CH, emphasizing load balance. This approach is designed to enhance the lifespan of active nodes and reduce energy consumption in both directions of communication between nodes and the CH.
- Propose a novel Cognitive Coati Optimization Algorithm (CCOA) to address the multi-objective optimization problem associated with optimal CH selection. The algorithm is intended to provide an effective solution for optimizing load balance and energy consumption in the network.

The paper is structured as follows: Section 2 presents the preliminaries of the methods used in this article. Section 3 proposes the CCOA for selecting the optimum cluster head, followed by data aggregation and load balancing. Section 4 presents the results and discussion, followed by the conclusion.

## 2. Preliminaries

This section discusses the basics of the methodology used in the proposed work to improve WSN networks' performance. A brief idea is provided in this section about the conventional COA optimization. The conventional COA method is presented here to discuss the possible future improvement to enhance the WSN networks' performance.

### 2.1 Conventional coati optimization algorithm (COA)

The conventional COA is used for finding the optimal cluster head, considering the fitness functions. The author in [23] introduce the Coati optimization algorithm (COA), which utilizes the population-based metaheuristic approach. The proposed COA method for global optimization has many advantages. Since this technique does not incorporate parameters, no regulation is needed. COA's has ability to solve complex, high-dimensional optimization issues in numerous disciplines is its second benefit. The third feature of the suggested approach is its ability to balance research and search processes, which speeds convergence and provides optimal choice variable values despite optimization dilemmas. Fourth, the suggested COA solves practical optimization issues well. The growing field of science and engineering involves multiple computational and non-linear problems that cannot be solved using a standard optimization algorithm. Although, this type of metaheuristic-based effective algorithm can easily find the best solution. The inherent behaviors of the coatis are elucidated in reference [23], which has a substantial impact on the proposed COA method.

#### 2.1.1 Initialization phase

The COA method is a metaheuristic algorithm that considers coatis as population members. The location of each coati within the search space directly corresponds to the values assigned to the decision variables. Therefore, within the context of the COA, the position of coatis suggests a potential solution to the problem. During the earliest stages of COA implementation, the coatis' positions in the search space are randomly initialized using Eq. (1).

$$X_i : x_{i,j} = lb_j + r(ub_j - lb_j), \mathrm{i} = 1,2,\ldots,\mathrm{N}, \mathrm{j} = 1,2,\ldots,\mathrm{m}, \tag{1}$$

The variable $x_{i,j}$ denotes the value of the $j^{th}$ decision variable, while $X_i$ represents the location of the $i^{th}$ coati in the search space. The quantity of coatis is represented by the symbol $N$, whilst the quantity of decision variables is represented by the symbol $m$. The symbol $r$ indicates a random real integer between 0 and 1. The lower bound and upper bound are expressed by the symbols $lb_j$ and $ub_j$, respectively. The population matrix $X$ is utilized to offer a numerical depiction of the coati population of the COA.

The following matrix X, referred to as the population matrix, is used to numerically depict the coati's population in the COA.

$$\mathrm{X} = \begin{bmatrix} \mathrm{X}_1 \\ \vdots \\ \vdots \\ \mathrm{X}_\mathrm{i} \\ \vdots \\ \mathrm{X}_\mathrm{N} \end{bmatrix} \mathrm{N} \times \mathrm{m} = \begin{bmatrix} \mathrm{x}_{1,1} & \cdots & \cdots & \mathrm{x}_{1,\mathrm{j}} & \cdots & \mathrm{x}_{1,\mathrm{m}} \\ \mathrm{x}_{2,1} & \cdots & \cdots & \mathrm{x}_{2,\mathrm{j}} & \cdots & \mathrm{x}_{2,\mathrm{n}} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathrm{x}_{\mathrm{i},1} & \cdots & \cdots & \mathrm{x}_{\mathrm{i},\mathrm{j}} & \cdots & \mathrm{x}_{\mathrm{i},\mathrm{m}} \\ \mathrm{x}_{\mathrm{N},1} & \cdots & \cdots & \mathrm{x}_{\mathrm{N},\mathrm{j}} & \cdots & \mathrm{x}_{\mathrm{N},\mathrm{m}} \end{bmatrix} \mathrm{N} \times \mathrm{m} \tag{2}$$

The evaluation of various values for the problem's objective function results from the positioning of potential solutions in choice variables. Eq. (3) is used to display these values.

$$F = \begin{bmatrix} F_1 \\ \vdots \\ F_i \\ \vdots \\ F_N \end{bmatrix} N \times 1 = \begin{bmatrix} F\ (X_1) \\ \vdots \\ F\ (X_i) \\ \vdots \\ F\ (X_N) \end{bmatrix} N \times 1 \tag{3}$$

The vector $F$ represents the objective function, while $F_i$ represents the objective function value acquired from the $i^{th}$ coati. The objective function value is used to evaluate candidate solutions in metaheuristic algorithms like the proposed COA. The population member who evaluates the best value for the objective function is considered the best member. As candidate solutions are updated during algorithm rounds, the best member of the population is likewise updated.

The COA algorithm updates the position of coatis (candidate solutions) by modeling two natural behaviors of these animals. The behaviors being studied are the strategy coatis use when attacking iguanas and their escape strategy from predators. As per the process, the COA population is updated in two distinct phases.

#### 2.1.2 Exploration phase

The initial coatis population update is modeled by replicating their tactics for attacking iguanas in the search space. One approach involves coatis climbing a tree to intimidate an iguana. Several coatis wait behind a tree till the iguana falls. Once the iguana is down, coatis attack and hunt it. By using this method, the COA can explore the problem-solving space globally by moving coatis to different spots. See Fig. 2 for the pattern diagram of this method.

$$X_i^{P1}: x_{i,j}^{P1} = x_{i,j} + r \,.\left(K_j - I \,. x_{i,j}\right),$$
$$\text{for i} = 1, 2, \ldots, \lfloor \text{N 2} \rfloor \text{ and j} = 1, 2, \ldots, \text{m} \tag{4}$$

In the construction of the COA, the assumption is made that the position of the most optimal member of the population corresponds to the position of the iguana. The vertical displacement of the coatis as they climb down from the tree is quantitatively calculated using Eq. (4). The iguana ($K$) is randomly put in the search space after falling to the ground. Eqs. (5) and (6) replicate the movement of ground-based coatis in the search space based on their random position.

$$K^G: K_j^G = lb_j + r\left(ub_j - lb_j\right), \quad , j = 1, 2, \ldots, \text{m} \tag{5}$$

$$X_i^{P1}: x_{i,j}^{P1} =$$
$$\begin{Bmatrix} x_{i,j} + r \,.\, \left(K_j^G - I \,. x_{i,j}\right), & F_{K^G} < F_i \\ x_{i,j} + r \,.\, \left(x_{i,j} - K_j^G\right), & else, \end{Bmatrix} \tag{6}$$
$$for\ i = \left\lfloor \frac{N}{2} \right\rfloor + 1, \left\lfloor \frac{N}{2} \right\rfloor + 2, \ldots \ldots, N\ and\ j = 1, 2, \ldots, \text{m}$$

The updated position for each coati is acceptable if it improves the objective function. Otherwise, the coati remains in the prior location. This update condition applies to simulated $i = 1, 2, \ldots, N$ using Eq. (7).

$$X_i = \begin{Bmatrix} X_i^{P1}, & F_i^{P1} < F_i, \\ X_i, & else \end{Bmatrix} \tag{7}$$

In this context, $X_i^{P1}$ represents the updated position of the $i^{th}$ coati, whereas $x_{i,j}^{P1}$ represents its $j^{th}$ dimension. The coati objective function value is denoted by $F_i^{P1}$. Random real number r ranges from 0 to 1. $K$ represents the iguana's best member position in the search space. $K_j$ represents its $j^{th}$ dimension. $I$ is a randomly chosen number from $\{1, 2\}$. In $K^G$, the iguana is randomly placed on the floor, and in $K_j^G$, its jth dimension. The objective function value for the iguana's position is $F_{K^G}$. Use the floor function $\lfloor\cdot\rfloor$ to find the largest integer less than or equal to the provided number.

#### 2.1.3 Exploitation phase

Based on how coatis naturally interact with and flee from predators, a mathematical model is used to simulate the second step of the process of updating coatis' locations in the search space.

$$lb_j^{local} = \frac{lb_j}{t}, ub_j^{local} = \frac{ub_j}{t},$$
$$\text{where t} = 1, 2, \ldots, \text{T} \tag{8}$$

$$X_i^{P2}: x_{i,j}^{P2} = x_{i,j} + (1 - 2r) \,.\Big( lb_j^{local} + r \,.\left(ub_j^{local} - lb_j^{local}\right)\Big), \tag{9}$$

$$i = 1, 2, \ldots, \text{N}, j = 1, 2, \ldots, \text{m},$$

To replicate the behavior of each coati, a random place is generated around their current location, based on Eqs. (8) and (9). If the objective function, represented by Eq. (10), exhibits a rise in value, then the newly estimated position is considered acceptable.

$$X_i = \begin{Bmatrix} X_i^{P2}, & F_i^{P2} < F_i, \\ X_i, & else \end{Bmatrix} \tag{10}$$

The equation $X_i^{P2}$ shows the updated position of the $i^{th}$ coati, determined by the second phase of COA. $x_{i,j}^{P2}$ represents the coati's $j^{th}$ dimension, whereas $F_i^{P2}$ represents its objective function. The variables $r$, $t$, $lb_j^{local}$, $ub_j^{local}$, $lb_j$, and $ub_j$ are random numbers between 0 and 1, the iteration counter, and the relative lower and higher bounds of the $j^{th}$ crucial parameter. Once the positions of all coatis in the search space have been updated according to the first and second stages, a COA's iteration is considered finished. The population is updated iteratively using Eqs. (4)-(10) until the final iteration of the procedure. Upon completion of the COA run, the result is the best answer achieved during all iterations of the algorithm. The COA is further enhanced cognitively to optimally selection of CH and discussed in section 3.2.2.

### 2.2 Non-negative matrix factorization (NMF)

NMF, or Non-negative Matrix Factorization, is a most widely used matrix decomposition method and applied in engineering problems. In this process the matrix $Y$ is decomposed into the product of two non-negative matrices $X$ and $I$, as represented by Eq. (11),

$$Y \approx XI = X_{nl} I_{lo} \tag{11}$$

In this equation, $Y \in S_{+}^{n\times o}$, $X \in S_{+}^{n\times l}$, $I \in S_{+}^{l\times o}$, and $l'' \frac{no}{n+o}$. The expression indicates that each column $y_j$ of the matrix $Y$ is approximated by a linear combination of the columns of $X$ weighted by the elements of . The choice of $l$ (the number of basis vectors) and the number of iterations ($u$) are critical in NMF, with the NMF error curve guiding their determination. The objective in NMF is to find matrices $X$ and $I$ that minimize the reconstruction error represented by matrix $F$ in Eq. (12):

$$F = Y - XI \tag{12}$$

To achieve this, the Maximum Likelihood Estimation (MLE) is employed, assuming a Gaussian distribution for the noise. The likelihood function $M(X, I)$ in Eq. (13) is maximized by minimizing the negative log-likelihood function $K(X, I)$ in Eq. (15). The gradient descent [26], iterations in Eqs. (16) and (17) are used to update $X$ and $I$, with multiplicative update rules (Eq. (21)) ensuring positivity.

$$M(X, I) = \prod_{(j,k)} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{F_{jk}^2}{2\sigma^2}} = \prod_{(j,k)} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\left(Y_{jk} - X_{jl} I_{lk}\right)^2}{2\sigma^2}} \tag{13}$$

$$\ln M(X, I) = \sum_{(j,k)} \left[ ln \frac{1}{\sqrt{2\pi}\sigma} - \frac{1}{2\sigma^2} \left(Y_{jk} - X_{jl} I_{lk}\right)^2 \right] \tag{14}$$

$$K(X, I) = \frac{1}{2} \sum_{j,k} \left[Y_{jk} - X_{jl} I_{lk}\right]^2 \tag{15}$$

$$\frac{\partial K(X,\iota)}{\partial X_{jl}} = (YI^T)_{jl} - (XII^T)_{jl} \tag{16}$$

$$\frac{\partial K(X,\iota)}{\partial I_{lk}} = (X^T Y)_{lk} - (X^T XI)_{lk} \tag{17}$$

$$X_{jl} = X_{JL} - \eta_1 \frac{\partial K(X,\iota)}{\partial X_{jl}} \tag{18}$$

$$I_{lk} = I_{lk} - \eta_2 \frac{\partial K(X,\iota)}{\partial I_{lk}} \tag{19}$$

The update rules for $X$ and $I$ are provided in Eqs. (18) and (19), with $\eta_1$ and $\eta_2$ being update coefficients. The final update rule (Eq. (21)) involves multiplying the current values by factors dependent on the accuracy of the estimate. This iterative, multiplicative approach ensures that each step maintains non-negativity and decreases the objective function, leading to a globally optimum matrix factorization.

$$\eta_1 = \frac{X_{jl}}{(X^T XI)_{lk}} \quad \eta_2 = \frac{X_{jl}}{(XII^T)_{lk}} \tag{20}$$

$$X_{jl} = X_{jl} \frac{(YI^T)_{j_l}}{(X\,I\,I^T)_{j_l}} \quad I_{lk} = I_{lk} \frac{(X^T Y)_{l_k}}{(X^T X\,I)_{l_k}} \tag{21}$$

In practical applications, NMF is employed in two stages: load balancing clustering and CH selection, as detailed in sections 4.1.1 and 4.1.2, respectively. This utilization of NMF enhances efficiency, particularly when dealing with high-dimensional matrices by effectively reducing the original matrix.

## 3. Methodology

### 3.1 Problem statement

The main aim of the study is to apply clustering and select an optimal CH while compressing the data aggregated in the heterogenous WSN. Maintaining and balancing the energy consumption and load of the nodes is the challenge to be addressed in WSN where data processing affects the battery life. This is achieved with efficient load balancing; it reduces the processing overhead, improves the energy consumption and focuses on improving the network lifetime.

#### 3.1.1 Load balanced clustering

In clusters of WSN, the huge amount of data produced causes overloading in respective CH and this effects the performance of the network causing delay in transmission of packets and huge energy consumption. To overcome this problem an efficient design of cluster for clustering process for load balancing is significant step [24]. Such overloading in CH will affect its capacity to hold the information and leads to delay in transmission. Here to address

Received: February 7, 2024. Revised: March 19, 2024. 

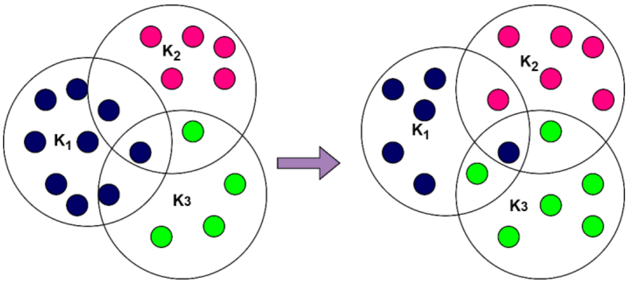

Figure. 1 Illustration of load balancing in WSN

the load balancing issue, the data is redistributed through a strategy. This concept is illustrated and explained through Fig. 1. In an uneven distributed WSN environment, with the varying sensor nodes in WSN given 9, 4 and 5 be the CHs expressed as K1, K3 and K2 respectively. Such uneven distributed led to overloading at the K1 risking its ability to transmit data and reduce the lifetime. In such cases, the overload can be mitigated by distributing it to the K2 and K3 extending the lifetime. In this study, it is formulated as, assuming the sensor nodes as $N$ distributed in random in WSN the data generated is given as $Y_{j\in\{1,2,..N\}} = (y_1, y_2, \cdots, y_N)$ , where $y_{j\in\{1,2..N\}}$ shows the data from the $jth$ sensor. The distances between nodes are formulated in the $Y$ matrix as seen in Eq. (22).

$$Y = \begin{bmatrix} y_{1,1} & \cdots & & \cdots & y_{1,k} & y_{1,o-1} & y_{1,o} \\ y_{2,1} & \cdots & & \cdots & y_{2,k} & \cdots & y_{2,o} \\ \cdots & \cdots & \cdots & \cdots & & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ y_{o-1,1} & \cdots & \cdots & y_{o-1,k} & & \cdots & y_{o-1,o} \\ y_{o,1} & \cdots & \cdots & y_{o,k} & & y_{o,o-1} & y_{o,o} \end{bmatrix} \tag{22}$$

To frame a clustering problem, the centroids for each data in $Y$ are discovered and a factorization model is achieved. The main objective of NMF is to decompose such input $(Y)$ as two matrices $X$ and $I$ as explained in Eq. (11). The noise and unnecessary data with same distance is removed by fine-tuning of the matrix $I$, addressing the features in the input data through NMF effectively. Such fine-tuning is obtained by the application of K-means algorithm.

Here through K-means the clustered index of the nodes in $N_{sensor}$ are calculated. Assuming the cluster centers $d$ as a set in $B$ given as, $\{b_1, \cdots, b_d\}$. Consider $Xl$ as a data collection in a d-dimensional Euclidean space, where the Euclidean norm is given by $e_{jl} = \|xi_j - b_l\|$ is applied. Let $V$ be a matrix with dimensions $(o \times d)$ where $\mu_{jl}$ is a binary variable $(\mu_{jl} \in \{0,1\})$ indicating whether the data point $xi_j$ corresponds to the k-th cluster (where $k = 1, \cdots, d$) and say $V = [\mu_{jl}]_{o\times d}$. Using the K-means clustering method, $k$ clusters are formed based on the Euclidean distance between values in the fine-tuned distance matrix, serving as the similarity measure for grouping into clusters.

### 3.1.2 CH selection

Although the proposed clustering in the above section ensures the even distribution of nodes and workload among the nodes in clusters, it is important to maintain the durability of CH. Such compromise in the durability of CH will cause disruption in network. To address this issue, in this proposed model, the objective function of optimisation algorithm is dependent on the density of nodes in the cluster $k$ such that the aim is to balance the clusters with distributed residual energy.
The proposed effective fitness function for load balancing of cluster heads [24], denoted as $f_1$, is calculated using the equation,

$$f_1 = \left(1 - \frac{\mu_{load}}{CH_{max}}\right) + \left(\frac{loaded\ cluster\ head}{total\ cluster\ head}\right) \tag{23}$$

Here, $\mu_{load}$ represents the mean of load calculated by Eq. (24) [24], and $CH_{max}$ is the CH with the maximum load.

$$\mu_{load} = \frac{\sum_{i=1}^{m} Load\ (CH_q)}{cluster\ heads} \tag{24}$$

The load on a cluster head $Load(CHq)$ is influenced by the number of sensor nodes $(N)$, initial energy $E_{initial}\ (CH_q)$ and remaining energy $E_{remain}\ (CH_q)$ as described in Eqs. (25) and (26).

$$Load\ (CH_q) = N \times \frac{E_{remain}\ (CH_q)}{E_{initial}\ (CH_q)} \tag{25}$$

The energy consumption of each node is modeled by the function $f_2 = E_T(N, d)$ where $d_0$ represents the distance between the sender and receiver, $N$ denotes the number of nodes, and $E_{elec}$ is the energy required by electronic circuitry.

$$f_2 = E_T(N, d) = \begin{cases} N \times E_{elec} + N \times \epsilon_{fs} \times d^2, & d < d_0 \\ N \times E_{elec} + N \times \epsilon_{mp} \times d^4, & d \geq d_0 \end{cases} \tag{26}$$

Here, $\epsilon_{mp}$ represents Multipath channel energy, and $\epsilon_{fs}$ represents free space energy. Additionally, data compression occurs at the CH, utilizing Non-negative Matrix Factorization (NMF) to achieve dimension reduction. The third fitness function, $f_3 = E_T(CH, d)$ captures the energy consumed by the cluster head after data compression.

$$f_3 = E_T(CH, d) = \begin{cases} CH \times E_{elec} + CH \times \epsilon_{fs} \times d^2, & d < d_0 \\ CH \times E_{elec} + CH \times \epsilon_{mp} \times d^4, & d \geq d_0 \end{cases} \quad (27)$$

The proposed CCOA utilizes three fitness functions, denoted as $f_1$, $f_2$, and $f_3$, as inputs to derive an optimal solution based on a lower-order rank matrix. The objective is to minimize these fitness functions, where a lower value indicates a superior solution. The computation of fitness function $F_1$ is explained using an example. Let's consider nodes $K1, K2, K3$, and $K4$ with loads 15, 9, 18, and 22, respectively. The load $\mu_{load}$ is calculated as the average of these loads, i.e., $\frac{15+9+18+22}{4}$. Given $CH_{min}$ and $CH_{max}$values as 9 and 22, respectively, $T_{max}$ and $T_{min}$ are determined according to the provided information [24].

Using $T_{max}$ and $T_{min}$values, a CH with a load of 22 is categorized as heavily loaded, while a load of 9 designates an underloaded CH. The fitness value $F_1$ is then computed using Eq. (23), which involves the ratio of underloaded to heavily loaded nodes. As an example, assuming uniform loads of 5 on $K1, K2, K3$, and $K4$, the calculated $F_1$ value is 0, indicating a perfectly balanced network. Achieving a fitness value close to zero is recommended in practical scenarios, as achieving a completely balanced network is often impractical. Hence, selecting a solution with a fitness value approaching zero is a pragmatic choice.

### 3.2 Proposed methodology

Considering the problem formulated in load balancing in section 4.1, the optimal CH selection is converted to a single objective problem by weighted sum of Eqs. (23), (24) and (26). The fitness function that consists of above objectives can be defined as $f(x) = w_1 f_1 + w_2 f_2 + w_2 f_2$, where $w_{\in\{1,2,3\}}$ are the weights assigned to the objectives $f_1, f_2, f_3$. The above three objectives are linked to a common node's attribute i.e. distance, so these don't present the pareto front optimization problem. Due to this, these objectives are converted into weighted single objective problem.

The optimization problem converted into identifying the optimal node to be eligible for CH with minimum value of the $f(x)$ into that cluster. The novel improved optimization algorithm CCOA has the responsibility to select the optimal node to elected as CH. Its an iterative process and a new set

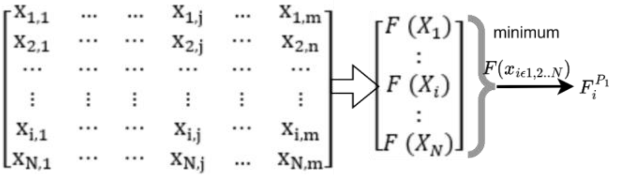


Figure. 2 CH selection by selecting the minimum objective value from possible search space in every iteration

of CHs in all clusters are selected by the CCOA. The dimension of tuning variable $m$ of CCOA is equal to the number of clusters i.e. $m = k$. In every iteration, the three objectives are calculated and a matrix $X$ of size $N \times M$ possible solutions is stored into $F$ as in Eqs. (2) and (3). The position with the minimum of $F_{N\times 1}$ is selected as the best set of CHs location so far. An example for this from the optimization values of CCOA is shown in Fig. 2.

The minimum objective value $F_i^{P_1}$ indexed solution $X_i^{P_1}$ is considered as the set of optimal CH parameters so far. $X_i^{P_1}$gets updated by the Eq. 6 in COA. The conventional COA algorithm traps into local minima in exploration step due to considering the best local position in the update equation.

In this article, we proposed a novel terminology addition in exploration phase of the COA to make it more able to exploit the search space and avoid the trap of local minima. This new term is named cognitive factor. Utilizing this factor can aid the COA in expediting its convergence towards the optimal solution by upholding the heterogeneity among its constituents, thereby preventing the occurrence of entrapment in local minima. The cognitive factor uses the subtraction of the local best position from the randomly selected current set of CH locations. Eq. (6) can be updated to incorporate the cognitive factor as in Eq. (28). The 3rd factor added is the cognitive factor and $local(j)$ is the local position selected randomly from the $X_i$ possible set of CH indexes in each cluster in the $j^{th}$ iteration. After this introduction in the exploration step, the new optimization algorithm is termed as Cognitive Coati Optimization Algorithm (CCOA). Fig. 3 depicts the addition of cognitive factor into the old coati optimization algorithm.

The novel CCOA can be used in many engineering applications. The usage of this algorithm in WSN CH selection needs the understanding of precise mapping of the CCOA terminologies in CH selection problem. Table 2 denotes the coati's structure. The coati's positions are represented as the index of the nodes which are picked as the CH in each cluster and of size $N \times m$ with $N$ number of possible set of selected CHs and $m$ is the number of CH as per the number of clusters in the network.

Received: February 7, 2024. Revised: March 19, 2024. 

$$X_i^{P1} : x_{i,j}^{P1} = \begin{Bmatrix} x_{i,j} + r\,.\,\left(K_j^G - I\,.\,x_{i,j}\right) + (\mathbf{1} - \boldsymbol{r}) \times \left(\boldsymbol{local}(\boldsymbol{j}) - \boldsymbol{I}\,.\,\boldsymbol{x}_{\boldsymbol{i},\boldsymbol{j}}\right),\ F_{K^G} < F_i \\ x_{i,j} + r\,.\,\left(x_{i,j} - K_j^G\right), \qquad else, \end{Bmatrix} \tag{28}$$

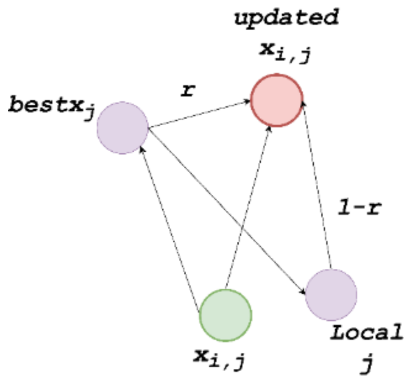

Figure. 3 Novel Cognitive Coati optimization algorithm (CCOA) Position Update

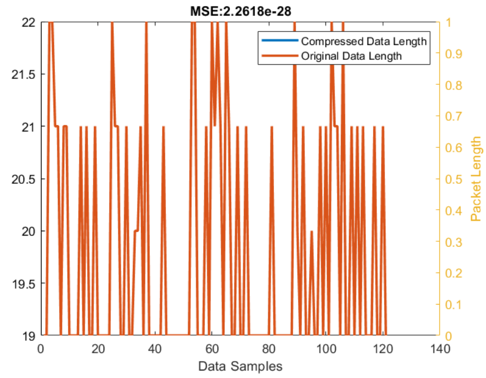

Figure. 6 Illustration of original data at the sink and compressed data with MSE value 3

Table 2. A generic layout of CCOA terminologies with CH selection problem

| **Structure of a Coati** |
|---|
| Matrix account for Coati's position => Matrix account for Resource Allocation |
| Fitness (Cost) => weighted sum of the three objectives of CH selection as in equation 23,26, and 27 |
| $\mathbf{F_i^{P1}}$ and $\mathbf{F_i^{P2}}$ => best fitness value in exploration and exploitation phase respectively |

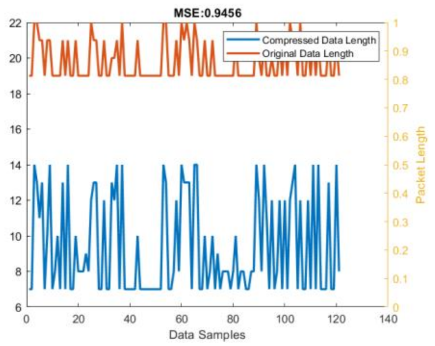

Figure. 4 Illustration of original data at the sink and compressed data with MSE value 1

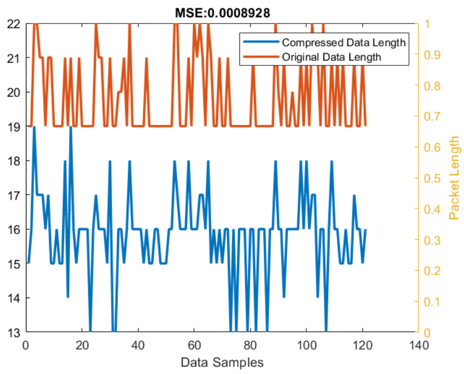

Figure. 5 Illustration of original data at the sink and compressed data with MSE value 2

The energy consumption by the nodes from the CH to the cluster member is considered one of the objective functions. The CH aggregates the data from clustered nodes and due to higher payload, the energy depletes faster, and CH node dies soon. The lossless data compression method proves to be a boon to enhance the CH lifetime. The NMF suggested in section 2.2 is used for that purpose. The received data matrix is factorized by NMF and the first orthogonal factor is transmitted to sink from CH which is lesser in payload but enrich in information. Eq. (12) is used to reconstruct the complete information if required using another broadcasted factor. The tolerable error in reconstruction is decided by the type of information broadcasted. The redundant bits in the data are removed by the NMF and an optimal compression ratio is selected that depends on the tolerance of error as shown in Fig. 4. Higher the compression ratio, higher is the reconstruction error, the simulated plots for three MSE values are provided in Figs. 4-6. The compression ratio from the Fig. 4 is $\frac{14}{22} = 0.36$ with MSE as 1.1178 whereas the compression ratio of 0.14 results in lesser reconstruction MSE of 0.0014. An optimal compression ratio of 0.14 can be selected for reconstruction MSE of 0.0014 in the CH aggregated data compression.

| **Algorithm 1: Pseudo-code of the Hybrid Improved CCOA Algorithm** |
|---|
| **Input:** Number of clusters $(m)$, Number of iterations $(max_{iter})$, Number of nodes $(N)$, Initial CH positions $(X_i)$, Objective weights $(w1, w2, w3)$.<br>**Procedure:**<br>**#Clustering Phase** |

Received: February 7, 2024. Revised: March 19, 2024. 

1. Generating a matrix $Y_j$ where $j \in \{1,2..N\}$ for $N_{sensor}$ sensor nodes in WSN
2. Applying the NMF through Eq. (11), where $Y = (X, I)$
3. Using the matrix information from $I$ after data reduction apply K-means
4. For data in $I$ obtained from previous stage the distance is calculated through Euclidean space $e_{jl} = \| xi_j - b_l \|$ forming the clusters $k$ based on it.
5. Calculate cluster centroids of the clusters.

**#CH selection Phase, improved CCOA**

6. Calculate the fitness functions F1, F2 and F3 using the Eqs. (23), (26) and (27)
7. Calculate the objective function $f(x)$, $f(x) = w_1 f_1 + w_2 f_2 + w_2 f_2$.
8. Initialize the CCOA model with $max_{iter}$
9. For $iter$ in $max_{iter}$:
   (1) A matrix $X$ is evaluated for where $X = (N, M)$ # applying NMF
   (2) Generate the initial population and evaluate the objective function $f(x)$ for values stored in
   (3) Update the coati's position in updated exploration phase of COA through Eq. (28)
   (4) Train the exploitation phase through Eq. (8)-(10).
   (5) Save the best solution so far $F_{N\times1} = X$
   (6) Update the new position of coati such that $min(F_{N\times1}) = best\ CH$
10. End For

**Output:** Return the final CH positions and compressed data parameters

## 4. Result & discussion

### 4.1 Simulation environment

The whole work is simulated in MATLAB on the Intel core i7-2600 CPU with 3.40 GH and 12 GB RAM. The WSN environment simulation parameters are listed in table 3.

Table 3. WSN parameters

| Simulation Parameters | Values |
|---|---|
| Topology Pattern | Random |
| Nodes Quantity | 500 |
| Antenna Direction | Omnidirectional |
| Initial Energy of nodes | 0.5 J |
| Transmission Distance | 20 m |
| Packet Size | 64 bytes |
| Data Rate | 14 packets/sec |

The improvement in the COA by the novel contribution of cognitive factor is evaluated on the benchmark functions of the optimization which is discussed in section 4.2 and final results evaluation on the basis of network parameters like number of alive nodes, packet delivery and residual energy is done in section 4.3, followed by the state-of-the-art comparison in section 4.4.

### 4.2 Performance evaluation of CCOA (Benchmark testing)

Several test functions are used to evaluate the CCOA's convergence performance in comparison to the classic COA. The CCOA method proves its dual-dimensional problem-solving capabilities when tested on these benchmark test functions. The test results are displayed in Table 4 together with the standard fitness functions [29]. Figs. 5-8 show the best-performing convergence curves that use the CCOA optimization benchmark functions. As can be seen from the figures, the CCOA demonstrates an improved degree of convergence on the given test functions. On top of that, CCOA outperforms conventional COA in terms of functional performance for non-linear functions and convergence speed for other test functions.

### 4.3 Results evaluation

In this Study, initially the nodes are distributed at random with node count as 500 showing the large-scale application like IoT. The clustering process is carried out using NMF and 5 clusters are generated to challenge the uniform number of nodes in clusters; this is illustrated in Fig. 11. The number of nodes in 5 clusters are 101, 97, 96, 107, 99. This shows the load balanced clustering by suggested NMF. With the proposed model the convergence speed of CCOA is improved compared to conventional COA shown in Figs. 7-10, run for test functions 7, 8, 13, and 21. With such improvement, the optimal selection of CH is achieved by CCOA.

The concept of data compression through NMF is applied and the proposed model is evaluated over a range of metrics and showing that it has achieved a better network lifetime. Each network metric is compared into four test cases. The legends in the plots are indicting to those test case and the used acronym are CCOA-DC (Cognitive Coati Optimization Algorithm with Data Compression) as the final proposed methodology, CCOA as the test case without the data compression and similar are COA-DC and COA.

The node lifetime is simulated and evaluated over the simulation time as seen in Fig. 12. The number of

Received: February 7, 2024. Revised: March 19, 2024. 

Table 4. Test Results for fitness functions

| Function | Function | COA | CCOA |
|---|---|---|---|
| $F_6$ | $f(x)=\sum_{i=1}^{n}([x_i+0.5])^2$ | 0.03 | 0.01 |
| $F_7$ | $f(x)=\sum_{i=0}^{n} ix_i^4+random\ (0,1)$ | $10^{-3}$ | $10^{-4}$ |
| $F_8$ | $f(x)=\sum_{i=1}^{n}(-x_i\sin(\sqrt{\lvert x_i\rvert}))$ | -2800 | -3200 |
| $F_9$ | $f(x)=\sum_{i=1}^{n}[x_i^2-10\cos(2\pi x_i)+10]$ | $10^{-11}$ | $10^{-10}$ |
| $F_{10}$ | $f(x)=-20exp\left(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2}\right)-exp\left(\frac{1}{n}\sum_{i=1}^{n}\cos(2\pi x_i)\right)+20+e$ | $10^{-15}$ | $10^{-15}$ |
| $F_{11}$ | $f(x)=1+\frac{1}{4000}\sum_{i=1}^{n}x_i^2-\prod_{i=1}^{n}\cos\left(\frac{x_i}{\sqrt{i}}\right)$ | $10^{-15}$ | $10^{-13}$ |
| $F_{12}$ | $f(x)=\frac{\pi}{n}\{10\sin(\pi y_1)\}+\sum_{i=1}^{n}(y_i-1)^2\left[1+10sin^2(\pi y_i+1)+\sum_{i=1}^{n}u(x_i,10,100,4)\right],\ where\ y_i=1+\frac{x_i+1}{4}, u(x_i,a\,k,m)\begin{cases}K(x_i-a)^m & if\ \ x_i>a\\ 0 & -a\le x_i\ge a\\ K(-x_i-a)^m & -a\le x_i\end{cases}$ | 1 | 1 |
| $F_{13}$ | $f(x)=0.1\left(sin^2(3\pi x_1)+\sum_{i=1}^{n}(x_i-1)^2[1+sin^2(3\pi x_i+1)]+(x_n-1)^2+sin^2(2\pi x_n)\right)+\sum_{i=1}^{n}u(x_i,5,100,4)$ | 1.2 | 1 |
| $F_{14}$ | $f(x)=\left(\frac{1}{500}+\sum_{j=1}^{25}\frac{1}{j+\sum_{i=1}^{2}}(x_i-a_{ij})\right)^{-1}$ | - | 2.984 |
| $F_{15}$ | $f(x)=\sum_{i=1}^{11}\left[a_i-\frac{x_1(b_i^2+b_ix_2)}{b_i^2+b_ix_3+x_4}\right]^2$ | $10^{-4}$ | $10^{-4}$ |
| $F_{16}$ | $f(x)=4x_1^2-2.1x_1^2+\frac{1}{3}x_1^6+x_1x_2-4x_2^2+4x_2^4$ | -1 | -0.9 |
| $F_{17}$ | $f(x)=\left(x_2-\frac{5.1}{4\pi^2}x_1^2+\frac{5}{\pi}x_1-6\right)^2+10\left(1-\frac{1}{8\pi}\right)\cos x_1+10$ | 0.45 | 0.4 |
| $F_{18}$ | $f(x)=[1+(x_1+x_2+1)^2(19-14x_1+3x_1^2-14x_2+6x_1x_2+3x_2^2)]\times[30+(2x_1-3x_2)^2\times(18-32x_i+12x_1^2+48x_2-36x_1x_2+27x_2^2)]$ | 30 | 30 |
| $F_{19}$ | $f(x)=-\sum_{i=1}^{4}c_iexp\left(-\sum_{i=1}^{3}a_{ij}(x_j-p_{ij})^2\right)$ | -3.8 | -3.8 |
| $F_{20}$ | $f(x)=-\sum_{i=1}^{4}c_iexp\left(-\sum_{i=1}^{6}a_{ij}(x_j-p_{ij})^2\right)$ | -3.1 | -3.2 |
| $F_{21}$ | $f(x)=-\sum_{i=1}^{5}[(X-a_i)(X-a_i)^T+c_i]^{-1}$ | -2 | -3 |

Received: February 7, 2024. Revised: March 19, 2024. 

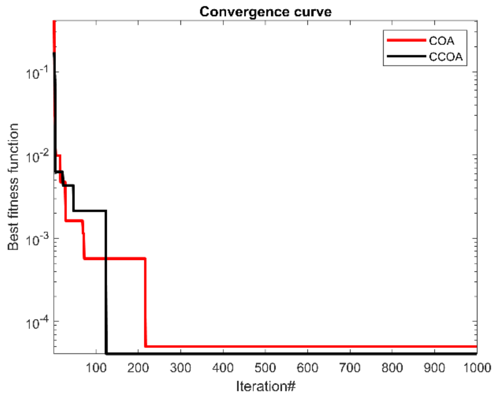


Figure. 7 Result for test function 7

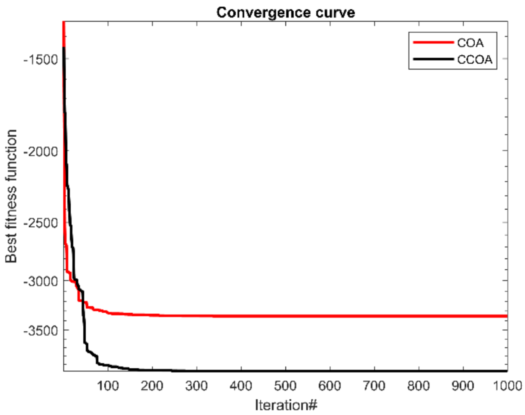


Figure. 8 Result for test function 8

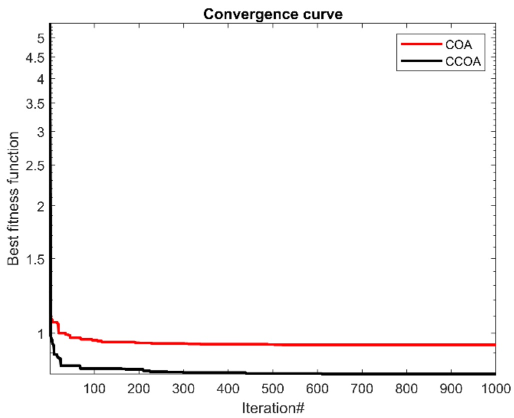


Figure. 9 Result for test function 13

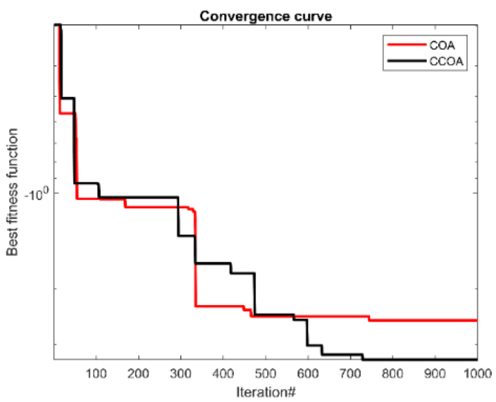


Figure. 10 Result for test function 21

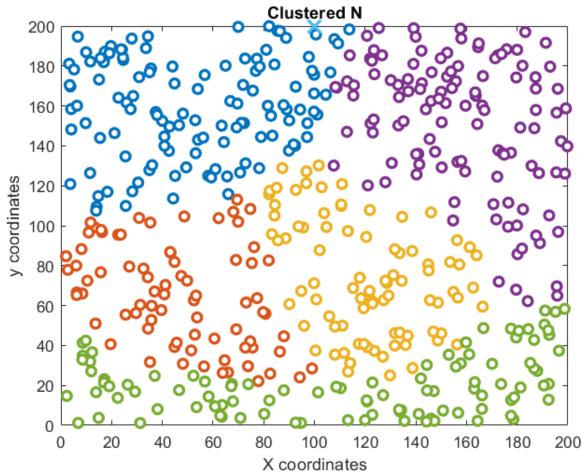


Figure. 11 Illustration of clustering for a node count of 500

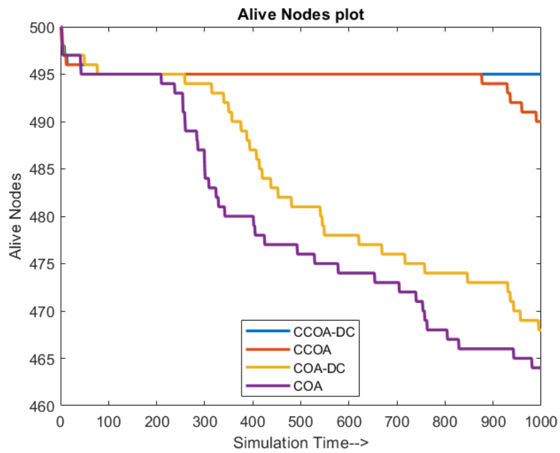


Figure. 12 Network lifetime analysis in terms of node lifetime

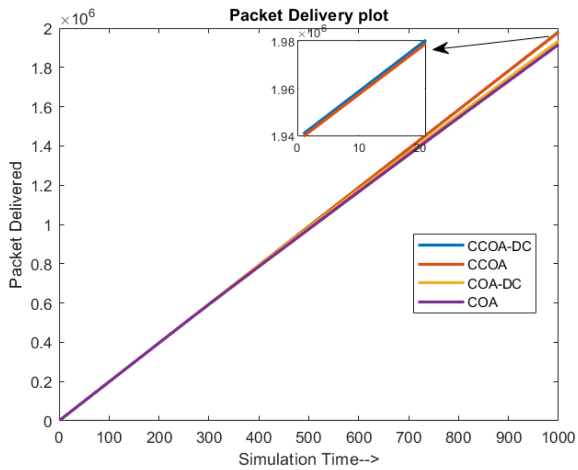


Figure. 13 PDR analysis of proposed CCOA-DC data aggregation model

alive nodes is used as a critical parameter to quantify the network's operational efficiency and resilience. Nodes typically have limited energy resources, and as they deplete their energy, they may become unable to

Received: February 7, 2024. Revised: March 19, 2024. 

communicate or perform sensing tasks. As illustrated in Fig. 12, the proposed Cognitive Coati Optimization Algorithm with Compression (CCOA-DC) demonstrates superior performance in terms of node lifetime compared to other state-of-art models. It depicts the total number of alive nodes across multiple simulation rounds (1 to 500). Notably, CCOA exhibits a higher count of alive nodes compared to CCOA without compression. Without data compression, the CCOA is able to have a higher lifetime than COA with data compression even. Though the CCOA without DC is able to compete with proposed methodology yet for higher number of simulation rounds after 876, the nodes start to die.
CCOA not only preserves a greater number of alive nodes but also strategically avoids selecting cluster heads with lower residual energy, thereby further improving the network's longevity. The comparison between compression-based and non-compression models reveals consistent improvements in network lifetime for each tested scenario. Particularly, CCOA-DC achieves a maximum improvement of 1.02% in terms of alive nodes compared to CCOA without compression.

Furthermore, the findings extend to Fig. 13, where a similar positive trend is observed in the number of delivered packets. This curve represents the performance of CCOA-DC in delivering a higher number of packets to the base station compared to other optimization methods. CCOA outperforms other hybrid optimization methods, showcasing its effectiveness in packet delivery. The improvement achieved by CCOA-DC, using the specified objective, is 0.08% and 2.34% compared to CCOA and COA-DC, respectively. It's worth noting that, even without data compression, CCOA exhibits improved packet delivery due to its enhanced convergence capabilities in the exploration phase. When CCOA is coupled with NMF based data compression, an additional improvement of 0.08% over CCOA without compression is acheived. This improvement is attributed to NMF achieving a favorable compression ratio and reducing the reconstruction error, contributing to enhanced packet delivery in the proposed scheme.

Analyzing the effectiveness of CCOA with DC in the network operation is centred around the assessment of residual energy in each node. Residual energy denotes the remaining energy in both CHs and Cluster Members (CMs) following a successful information transfer from source to destination [20]. The graphical representation highlights that, in comparison to CCOA, CCOA-DC, CA without CCOA, CCOA without compression, CA with CCOA, and CCOA, the residual energy is maximized

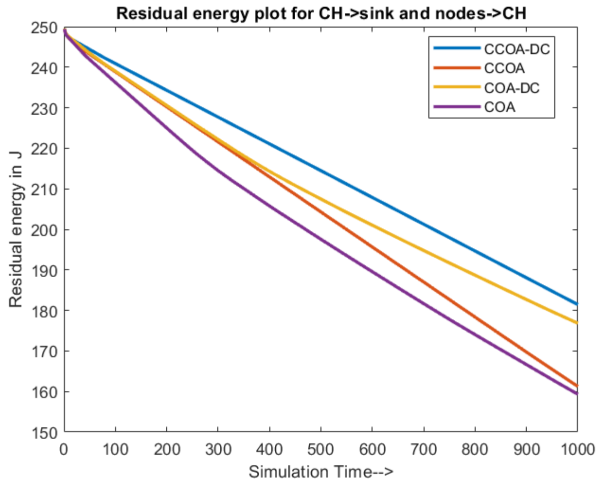


Figure. 14 Residual energy analysis of proposed CCOA-DC data aggregation model

in the case of CCOA with compression. Specifically, CCOA-DC exhibits a 12.53% increase in residual energy compared to CCOA without compression.
In terms of residual energy improvement, CCOA with compression, utilizing the proposed objective function, demonstrates a 12.53% and 2.61% enhancement compared to CCOA and COA-DC, respectively. Notably, when contrasted with non-compression-based approaches, CCOA with compression yields significantly higher residual energy, as depicted in Fig. 14. This phenomenon can be attributed to the fact that processing uncompressed data demands substantial power from the nodes, resulting in considerable energy wastage.

### 4.4 State-of-the-art comparison (SOTA)

This comparison involves 500 nodes, and the outcomes are juxtaposed with the performance of the TCBDGA [31], HEED [30], and FEEC-IIR [32], [33] algorithms. Firstly, in [30], the HEED algorithm focuses on prolonging network lifetime and enhancing energy efficiency by introducing a protocol tailored for clustered non-uniform sensor networks. However, the proposed Hybrid Improved CCOA Algorithm goes beyond traditional clustering methods by integrating NMF for data compression CCOA for cluster head selection. The authors in [31] introduces the TCBDGA (Tree-Cluster-Based Data-Gathering Algorithm) algorithm to mitigate the hotspot problem in WSNs with a mobile sink. While TCBDGA aims to balance network load and prolong network lifetime by introducing a novel tree-cluster-based data-gathering approach. n [32], the FEEC-IIR (Adaptive Fuzzy Rule-Based Energy Efficient Clustering and Immune-Inspired Routing) protocol aims to improve energy efficiency and network lifetime through adaptive fuzzy rule-based clustering

Table 5. The state-of-the-art comparison with the proposed Data aggregation scheme

| **Parameters** | **HEED [30]** | **TCBDGA [31]** | **FEEC-IIR [32]** | **POHDA-WSN [33]** | **FTOPSIS-HJBO [34]** | **CCOA-DC** |
|---|---|---|---|---|---|---|
| Packet Loss Ratio % | 17 | 8 | 5 | - | - | 2.8 |
| Packet Delivery Ratio % | 85 | 88 | 96 | 85 | 96 | 98.67 |
| Energy Consumption (joule) | 180 | 145 | 65 | - | - | 68.01 |

and immune-inspired routing. Although FEEC-IIR addresses cluster head selection and routing challenges

Our proposed algorithm, CCOA with compression, exhibits enhanced results compared to the Hybrid LEACH and Jellyfish Algorithm (HEED), Tree-Cluster-Based-Data Gathering Algorithm (TCBDGA), and Fuzzy Rule-Based Energy-Efficient Clustering and Immune-Inspired Routing (FEECIIR) in terms of energy consumption, packet delivery ratio, and packet loss ratio. The packet delivery ratio signifies the proportion of successfully transmitted data to the total information sent and is inversely related to node density. This is why the CCOA-DC (CCOA with compression) achieves the highest packet delivery ratio among the tested algorithms with 500 nodes with improvement of 13.85%, 10.81%, 2.7% compared to HEED, TCBDGA and FEEC-IIR respectively. In terms of energy consumption, CCOA-DC demonstrates minimal energy usage for data transmission. The formation of clusters mitigates communication costs and prevents nodes from overlapping, leading to restrained energy utilization in CCOA-DC. A comprehensive comparison is provided in Table 5 showing 62%, 53% reduction in energy consumption compared to [30] [31] respectively.

POHDA-WSN [33], FTOPSIS-HJBO [34] are the recent state-of-art models proposed for data aggregation in WSN and promote energy efficiency. POHDA-WSN (Power Optimization and Hybrid Data Aggregation) employs a careful selection process for CH to optimize energy consumption and packet delivery. This selection process ensures that CHs are strategically placed to manage long-distance communication effectively. The model in [33] proposes a weighted approach to identify the CH considering residual energy, Distance, and degree of mobility as the CH selection parameters. However, the proposed model CCOA-DC shows a 13.85% improved packet delivery ratio compared to POHDA-WSN at 500 sensor nodes over 300 seconds. This is because NMF-derived clusters provide valuable insights into the spatial distribution and connectivity patterns of the sensor nodes. By strategically placing CHs in central locations within their respective clusters, NMF-based clustering ensures efficient data aggregation and forwarding, further enhancing the packet delivery ratio.

Similarly, [34] employs a Fuzzy TOPSIS-based Hybrid Jarratt Butterfly Optimization (FTOPSIS-HJBO) for optimal routing from CH to BS. This model aims to find the shortest path for data transmission by optimally selecting cluster heads through ensemble clustering [34]. The proposed method returns compressed data parameters along with the final CH positions. This enables efficient data transmission and reduces the overhead associated with transmitting data in the network, leading to improved network performance and energy efficiency. This is why the proposed model shows a 2.7% improvement in PDR compared to the model in [34].

## 5. Conclusion

This study addresses the challenges of energy efficiency and excessive data generation in large-scale applications, particularly in IoT based heterogeneous WSN network with 500 sensor nodes. Non-Negative Matrix Factorization (NMF) is applied for the clustering process, resulting in the generation of 5 clusters. The proposed novel Cognitive Coati Optimization Algorithm with Data Compression (CCOA-DC) demonstrates improved capability of handling nonlinear functions than conventional COA. CCOA-DC achieves improvement of 1.02% in terms of alive nodes compared to CCOA without compression. This shows that CCOA-DC strategically avoids selecting CH with lower residual energy, further improving the network's longevity. Further, the PDR improvement achieved by CCOA-DC is 0.08% and 2.34% compared to CCOA and COA-DC, respectively. Thus, the objective of improving the network performance along with reducing the redundant data is achieved. CCOA-DC achieves a 12.53% increase in residual energy compared to CCOA without compression. In terms of residual energy improvement, CCOA-DC

Received: February 7, 2024. Revised: March 19, 2024. 

outperforms CCOA and COA-DC by 12.53% and 2.61%, respectively. Thus, the proposed novel optimization with compression shows that energy consumption is optimized with optimal selection of CH. Moreover, comparative analysis with state-of-the-art models, including TCBDGA, HEED, and FEEC-IIR, demonstrates the superiority of CCOA-DC in terms of packet loss ratio (2.8%), packet delivery ratio (98.67%), and energy consumption (68.01%). In conclusion, CCOA-DC is successful in tackling the issues of load balancing, energy economy, and data redundancy in large-scale WSNs, particularly in situations that resemble IoT applications. The results emphasize the accelerated rate at which it converges, the extended lifespan of the network, and the greater performance it exhibits in comparison to current optimization methods. The latest comparison confirms that CCOA-DC is highly effective in attaining energy efficiency and reliability in data aggregation in WSNs compared.

**Notations:**

| | |
|---|---|
| $N$ | Number of coatis |
| $m$ | Number of decision variables |
| $X_i$ | Location of the $i^{th}$ coati in the search space |
| $lb_j$ and $ub_j$ | Lower bound and upper bound for the $j^{th}$ decision variable |
| $r$ | Random real integer between 0 and 1 |
| $Y$ | Matrix representing sensor nodes in the WSN |
| $X$ | Population matrix of coatis |
| $F$ | Objective function values |
| $X_{nl}$ | Non-negative matrix factor of X |
| $I_{lo}$ | Non-negative matrix factor of I |
| $F_i$ | Objective function value of the $i^{th}$ coati |
| $K$ | Position of the iguana in the search space |
| $K_j$ | Position of the iguana in the $j^{th}$ dimension |
| $F_i^{P1}$ and $F_i^{P2}$ | Objective function value of the $i^{th}$ coati in phase 1 and 2 respectively |
| $t$ | Iteration counter |
| $lb_j^{local}$ and $ub_j^{local}$ | Relative lower bound and upper bound for the $j^{th}$ decision variable and |
| $X_i^{P1}$ and $X_i^{P2}$ | Updated position of the $i^{th}$ coati in phase 1 and phase 2 |
| $\eta_1$ and $\eta_2$ | Update coefficient for $X$ and $I$ |
| $N_{sensor}$ | Number of sensor nodes |
| $K$ | Cluster index |
| $\mu_{load}$ | Mean load on cluster heads |
| $CH_{max}$ | Cluster Head with maximum load |
| $Load(CHq)$ | Load on a cluster head |
| $E_{initial}$ | Initial energy of a cluster head |
| $E_{remain}$ | Remaining energy of a cluster head |
| $E_T$ | Energy consumption function |
| $\in_{fs}$ | Free space energy parameter |
| $\in_{mp}$ | Multipath channel energy parameter |
| $CCOA$ | Cognitive Coati Optimization Algorithm |
| $f_1$ | Fitness function for load balancing |
| $f_2$ | Fitness function for energy consumption |
| $f_3$ | Fitness function for energy consumption at CH |

## Conflicts of Interest

The authors declare no conflict of interest

## Author Contributions

Mays Kareem Jabbar and Thaar A. Kareem contributed to this research article as follows: Conceptualization, methodology, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, supervision, project administration, and funding acquisition.

## References

[1] R. Yarinezhad and S. N. Hashemi, "Solving the load balanced clustering and routing problems in WSNs with an fpt-approximation algorithm and a grid structure", *Pervasive and Mobile Computing*, Vol. 58, p. 101033, 2019.

[2] A. Lipare, D. R. Edla, and V. Kuppili, "Energy efficient load balancing approach for avoiding energy hole problem in WSN using Grey Wolf Optimizer with novel fitness function", *Applied Soft Computing*, Vol. 84, p. 105706, 2019.

[3] V. K. Chawra and G. P. Gupta, "Load balanced node clustering scheme using improved memetic algorithm based meta-heuristic technique for wireless sensor network", *Procedia Computer Science*, Vol. 167, pp. 468-476, 2020.

[4] X. Liu, J. Yu, K. Yu, G. Wang, and X. Feng, "Trust secure data aggregation in WSN-based IIoT with single mobile sink", *Ad Hoc Networks*, Vol. 136, p. 102956, 2022.

[5] J. Ludeña-Choez, J. J. Choquehuanca-Zevallos, and E. Mayhua-López, "Sensor nodes fault detection for agricultural wireless sensor networks based on NMF", *Computers and Electronics in Agriculture*, Vol. 161, pp. 214-224, 2019.

[6] Z. Zhao et al., "A novel optimization method for WSN based on mixed matrix decomposition of

NMF and 2-SVD-QR", *Ad Hoc Networks*, Vol. 115, p. 102454, 2021.
[7] P. Favati, G. Lotti, O. Menchi, and F. Romani, "Adaptive Clustering via Symmetric Nonnegative Matrix Factorization of the Similarity Matrix", *Algorithms*, Vol. 12, No. 10, p. 216, 2019.
[8] M. V. Babu et al., "An improved IDAF-FIT clustering based ASLPP-RR routing with secure data aggregation in wireless sensor network", *Mobile Networks and Applications*, Vol. 26, No. 3, pp. 1059-1067, 2021.
[9] Y. H. Robinson, E. G. Julie, R. Kumar, and L. H. Son, "Probability-based cluster head selection and fuzzy multipath routing for prolonging lifetime of wireless sensor networks", *Peer-to-Peer Networking and Applications*, Vol. 12, No. 5, pp. 1061-1075, 2019.
[10] S. N. Sajedi, M. Maadani, and M. F. Nesari Moghadam, "F-LEACH: a fuzzy-based data aggregation scheme for healthcare IoT systems", *The Journal of Supercomputing*, Vol. 78, No. 1, pp. 1030-1047, 2022.
[11] A. Ullah et al., "Fog-assisted secure healthcare data aggregation scheme in IoT-enabled WSN", *Peer-to-Peer Networking and Applications*, Vol. 13, No. 1, pp. 163-174, 2022.
[12] J. S. Manoharan, "Double attribute-based node deployment in wireless sensor networks using novel weight-based clustering approach", *Sādhanā*, Vol. 47, No. 3, pp. 1-11, 2022.
[13] A. Jari and A. Avokh, "PSO-based sink placement and load-balanced anycast routing in multi-sink WSNs considering compressive sensing theory", *Engineering Applications of Artificial Intelligence*, Vol. 100, p. 104164, 2021.
[14] M. A. Angel and T. Jaya, "An Enhanced Emperor Penguin Optimization Algorithm for Secure Energy Efficient Load Balancing in Wireless Sensor Networks", *Wireless Personal Communications*, pp. 1-27, 2022.
[15] P. Maratha and K. Gupta, "HFLBSC: Heuristic and Fuzzy Based Load Balanced, Scalable Clustering Algorithm for Wireless Sensor Network", *Wireless Personal Communications*, pp. 1-24, 2022.
[16] E. Yousefpoor, H. Barati, and A. Barati, "A hierarchical secure data aggregation method using the dragonfly algorithm in wireless sensor networks", *Peer-to-Peer Networking and Applications*, Vol. 14, No. 4, pp. 1917-1942, 2021.
[17] L. Rajesh and H. S. Mohan, "Adaptive Group Teaching Based Clustering and Data Aggregation with Routing in Wireless Sensor Network", *Wireless Personal Communications*, Vol. 122, No. 2, pp. 1839-1866, 2022.
[18] V. R. Pasupuleti and C. Balaswamy, "Optimised routing and compressive sensing-based data communication in wireless sensor network", *IET Communications*, Vol. 14, No. 6, pp. 982-993, 2022.
[19] E. Tagne Fute, H. M. Kamdjou, A. El Amraoui, and A. Nzeukou, "DDCA-WSN: A Distributed Data Compression and Aggregation Approach for Low Resources Wireless Sensors Networks", *International Journal of Wireless Information Networks*, Vol. 29, No. 1, pp. 80-92, 2022.
[20] P. Rawat and S. Chauhan, "A novel cluster head selection and data aggregation protocol for heterogeneous wireless sensor network", *Arabian Journal for Science and Engineering*, Vol. 47, No. 2, pp. 1971-1986, 2022.
[21] P. Sreedevi and S. Venkateswarlu, "An Efficient Intra-Cluster Data Aggregation and finding the Best Sink location in WSN using EEC-MA-PSOGA approach", *International Journal of Communication Systems*, Vol. 35, No. 8, article e5110, 2022.
[22] L. Abualigah, A. Diabat, S. Mirjalili, M. Abd Elaziz, and A. H. Gandomi, "The arithmetic optimization algorithm", *Computer Methods in Applied Mechanics and Engineering*, Vol. 376, article 113609, 2021.
[23] M. Dehghani, Z. Montazeri, E. Trojovská, and P. Trojovský, "Coati Optimization Algorithm: A new bio-inspired metaheuristic algorithm for solving optimization problems", *Knowledge-Based Systems*, Vol. 259, article 110011, 2021.
[24] D. R. Edla, A. Lipare, R. Cheruku, and V. Kuppili, "An efficient load balancing of gateways using improved shuffled frog leaping algorithm and novel fitness function for WSNs", *IEEE Sensors Journal*, Vol. 17, No. 20, pp. 6724-6733, 2017.
[25] A. Urbańczyk, B. Nowak, P. Orzechowski, J. H. Moore, M. Kisiel-Dorohinicki, and A. Byrski, "Socio-cognitive evolution strategies", In: *Proc. of International Conference on Computational Science*, Cham: Springer International Publishing, pp. 329-342, 2021.
[26] Y. Shi, H. Liu, L. Gao, and G. Zhang, "Cellular particle swarm optimization", *Information Sciences*, Vol. 181, pp. 4460-449, 2011.
[27] S. K. Chaurasiya, A. Biswas, P. K. Bandyopadhyay, A. Banerjee, and R. Banerjee, "Metaheuristic load-balancing-based clustering

technique in wireless sensor networks", *Wireless Communications and Mobile Computing*, pp. 1-21, 2022.

[28] P. Kathiroli and K. Selvadurai, "Energy efficient cluster head selection using improved Sparrow Search Algorithm in Wireless Sensor Networks", *Journal of King Saud University – Computer and Information Sciences*, Vol. 34, pp. 8564–8575, 2022.

[29] L. Abualigah, A. Diabat, S. Mirjalili, M. Abd Elaziz, and A. H. Gandomi, "The arithmetic optimization algorithm", *Computer methods in applied mechanics and engineering*, Vol. 376, article 113609, 2021.

[30] R. Priyadarshi, L. Singh, and A. Singh, "A novel HEED protocol for wireless sensor networks", In: *Proc. of 2018 5th international conference on signal processing and integrated networks (SPIN), IEEE*, pp. 296-300, 2018.

[31] C. Zhu, S. Wu, G. Han, L. Shu, and H. Wu, "A tree-cluster-based data-gathering algorithm for industrial WSNs with a mobile sink", *IEEE Access*, Vol. 3, pp. 381-396, 2015.

[32] S. S. L. Preeth, R. Dhanalakshmi, R. Kumar, and P. M. Shakeel, "An adaptive fuzzy rule based energy efficient clustering and immune-inspired routing protocol for WSN-assisted IoT system", *Journal of Ambient Intelligence and Humanized Computing*, pp. 1-13, 2018.

[33] M. M. Al-Heeti, J. A. Hammad, and A. S. Mustafa, "Design and implementation of energy-efficient hybrid data aggregation in heterogeneous wireless sensor network", *Bulletin of Electrical Engineering and Informatics*, Vol. 13, No. 2, pp. 1392-1399, 2024.

[34] S. Ambareesh, H. C. Kantharaju, and M. Sakthivel, "A novel Fuzzy TOPSIS based hybrid jarratt butterfly optimization for optimal routing and cluster head selection in WSN", *Peer-to-Peer Networking and Applications*, Vol. 16, No. 5, pp. 2512-2524, 2023.