International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6399
***Article Received**: 02 February 2023 **Revised**: 09 March 2023 **Accepted**: 14 March 2023*

Cat and Mouse Based Task Optimization Model for Optimized Data Collection in Smart Agriculture

[*1]Vuppala Sukanya, [2]S. Ramachandram
[*1]Department of Computer Science and Engineering,
University College of Engineering, Osmania University,
Hyderabad, India.
[*1]Email: Sukanya.v@uceou.edu
[2]Department of Computer Science and Engineering,
University College of Engineering, Osmania University,
Hyderabad, India.
[2]Email: schandram@gmail.com

Abstract: Data collection from agricultural fields is tiring and requires novel methodologies to produce reliable outcomes. The combination of edge and wireless sensor networks (WSN) for smart farming enabled the efficient collection of data from remote fields to a vast extent. Adopting an optimization algorithm to achieve the data collection task is prioritized in the proposed work, and a new and effective data collection framework is proposed. The proposed framework initially collects the data from the agricultural fields via sensors and then transmits it to the edge server. The path between the sensors and the edge server is optimally obtained using the cat and mouse based task optimization (CMTO) model. The sensed data are transmitted through the optimal route, and then the edge server obtains and evaluates the data based on the data quality metrics such as precision, correctness, completeness and reliability. The valid data are then identified and transferred to the cloud servers for storage. The simulation of the work is done in Python platform and evaluated using the crop recommender dataset. The evaluations proved the method's efficacy compared to the existing state-of-the-art algorithms. The proposed work also provided upto 12.5% of improvement in terms of energy consumption, 7.14% of improvement in terms of communication latency, 4% of improvement in terms of execution cost, 2.27% of improvement in terms of completeness, 1.12% of improvement in terms of precision, 9.52% of improvement in terms of correctness, and 3.37% of improvement in terms of reliability.

Keywords: Data collection, edge computing, wireless sensor network, smart agriculture, optimization algorithm, optimal route, data quality measurement.

I. Introduction

Nowadays, crop cultivation is a tedious task in agriculture due to unpredictable climatic changes. Several techniques have been applied in this domain to predict climatic changes and its impacts on crop growth [1]. The ground water level is the major source of agriculture that is degraded due to the changing climatic conditions. Nowadays, information technology is combined with the agriculture to meet production's spatial and temporal variability. The integration of information technology with the agriculture field minimizes environmental issues and increases production [2]. The application of a wireless sensor network (WSN) in agriculture production is a promising technology to solve several issues in the production process [3, 4]. In WSN, several sensors are deployed in various fields to gather information from their surroundings [5]. Sensors in the network will sense and gather information from its surroundings regarding the water level, humidity, PH level, nutrition in soil and pH level [6].

All these objectives are included in precision agriculture (PA), improving field management from different perspectives [7]. The PA minimizes the excessive use of pesticides while providing adequate nutrients to increase productivity. In addition, the PA covers different agriculture issues, thereby improving field crop production [8]. The data management through various temporal and spatial dimensions is a tedious task due to the number of sensors. Big data and data analytics effectively offer predictive insights and real-time decision making [9]. In advance, the artificial intelligence based methods are integrated with big data analytics in WSN [10]. Cloud computing offers cost-efficient source sharing in the agriculture sector [11], in which the data is collected by several nodes to take certain control actions [12]. Cloud computing is a centralized data gathering and processing system with four layers: infrastructure, platform, hardware and application [13]. Conventional cloud computing technology is insufficient to process the massive data being generated from the agricultural domain [14].

As the rural areas lack internet connection, so the farming data cannot be transferred to cloud in such situations. Large amounts of data gathered by the sensors result in huge data traffic in the system, thus high powered processing devices are needed to take timely actions in this scenario [19]. In order to mitigate the issues such as poor bandwidth and high latency, auxiliary computing models are adopted along with cloud

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6399
***Article Received**: 02 February 2023 **Revised**: 09 March 2023 **Accepted**: 14 March 2023*

computing. This model reduces the data volume in the cloud and reduces latency [20]. Edge computing acts as an interface between cloud services and control processes, which process the data before it forwarding to the cloud [17]. Edge computing facilitates efficient access and utilization of agricultural data with minimized storage issues [18]. Thus, an edge computing technology is developed that maintains computing devices from the edge of the network to the end device. Furthermore, pure edge computing is not much effective due to the small size of the nodes. However, the combined use of edge and cloud computing is an encouraging technology for data collection in smart agriculture.

A. Motivation

Data collection from the agricultural fields is a time consuming process which hinders rate of productivity. The lack of appropriate equipment and the complexity involved in reaching the remote areas has presented a challenge in accurately collecting and determining data. The integration of WSN and edge computing with smart agricultural applications recently promoted promising data acquisition results. Though this method is advantageous, not many techniques are introduced in the literature to completely use the benefit of these technologies to acquire valuable data. The use of optimization algorithms is one of the promising research directions to find the optimal path in minimizing fitness. This has been proven in several works in different domains where the optimal path selection is needed. Currently, minimal work is based on optimization algorithms to collect the data via optimal paths in precision agriculture. The research lags identified in this domain remained a motivation to propose a new and effective strategy to accomplish optimized data collection.

B. Contribution

The major contributions of the proposed work are as follows:

- An effective framework for optimal data collection is proposed to enhance smart farming applications in the real world.
- The routes to transmit the data in the agricultural fields are identified using an algorithm called the cat and mouse based task optimization (CMTO) algorithm that finds the route iteratively.
- The data quality is assured before storing it in the cloud server based on certain data quality metrics. The space expended in cloud servers for storing invalid data can be reduced with quality assurance.
- Finally, extensive evaluations are conducted to demonstrate the efficacy of the proposed method in comparison with the other existing metaheuristic optimization algorithms.

C. Paper organization

The rest of the paper is planned as follows: Section 2 presents the literature review of the proposed work, Section 3 covers the proposed methodology with problem formulation and mathematical descriptions, Section 4 presents the results and discussion, and Section 5 concludes the paper.

II. Related work

Several data acquisition schemes and other irrigation schemes have been presented in the literature that combines WSN with smart farming. Some of the most recent and effective ones are reviewed in this section:

The integration of WSN with the agricultural sector resulted in significant advantages, and effective monitoring systems are being developed to monitor the fields spanned over acres. Since the nodes incorporated into WSNs are resource-constrained, it is important to design and develop power aware algorithms that can perform efficiently in data acquisition. Lakshmi et al. [21] presented an energy efficient routing protocol (EERP) to address the problems arising in smart farming. Energy efficiency was attained in the route setup phase through a reduced transmission count and distributed data acquisition. The nodes were deployed in a distributed manner in the agricultural lands to sense the temperature, moisture and pH parameters. The cluster head (CH) selection and data transmission were performed based on the retained energy. The aim of the algorithm was to elapse the network lifetime with increased coverage. The experimental analysis proved the efficacy of the algorithm compared to existing works.

The sensors are used directly deployed in the farm area to collect crucial information regarding the farm fields for crop growth. Since the resources of the sensors are limited, it is difficult to achieve long-distance transmissions and the successful collection of agricultural data. At the same time, it is also important to increase the lifetime of the network to regulate the periodic data gathering from the farm. To address these issues, Mahajan and Badarla [22] presented a nature-inspired algorithm based cross-layer clustering protocol (NICC) based on bacterial foraging optimization (BFO). The algorithm managed to model a trade-off between optimal data transmission and energy efficiency. For fitness evaluations, the cross-layer parameters from the medium access control (MAC), physical and network layers were formulated. Experimental analysis proved that the protocol was more efficient than most existing state-of-the-art clustering protocols.

With the network dynamicity, the traffic arriving on the routes may result in congestion, thereby affecting the performance of the network. This is one of the major disruptive cases seen in precision agriculture (PA) applications

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6399
***Article Received**: 02 February 2023 **Revised**: 09 March 2023 **Accepted**: 14 March 2023*

that raise the immediate requirement of possible counter measures. This scenario was studied, and an effective solution was put forth by Agarkhed et al. [23] with the cluster based routing technique. The technique was named PA with cluster based optimal routing in WSN (PAwCOR). Initially, the CHs were selected using non-probability method and then the rule-based deployment strategy was adopted for scheduling of the nodes. Every node deployed in the network supported a stochastic scheduling algorithm to enable accurate data collection. The paths were then constructed to ensure congestion-less and reliable data transmission. Then a decision support system (DSS) was utilized to select the optimal path among the constructed paths. The evaluation of the method proved optimal and more effective than the other methodologies in the literature.

While routing the collected data to the destination, a routing protocol must be energy efficient to enable long-term monitoring. Accordingly, a non-uniform clustering routing protocol based on effective energy consumption (UCEEC) was developed by Miao et al. [24]. The nodes were distributed in the wheat farmland to acquire the data from the fields. Clustering was performed to enable effective routing of the data to the destination based on the similarity between the nodes. A multi-hop route selection technique was framed between the CHs based on predicting the two-hop effective energy consumption. A cost factor was computed to maintain a balance in energy consumption in the entire network. Further, simulations were performed, and the results proved the method's efficiency in terms of power compared to other energy efficient routing protocols.

The adoption of edge and cloud computing infrastructures in smart agricultural scenarios raised the significance of smart farming. Efficient data collection is achieved by integrating edge computing in WSN based smart farming. Li et al. [25] developed a methodology for multiple data collection using edge computing in agricultural WSN. Initially, a data collection framework was constructed, and the data collection process was modelled with different tasks and sensors. Then, the best node among the available sensor nodes was determined using a double selection strategy. The best node was selected on the basis of fulfilling the latency and quality of data (QoD) requirements. Finally, the data collection algorithm was formulated to find the best nodes to complete the execution process. The evaluations of the model proved that the coupling of edge computing with the WSN yielded better results than the traditional WSN based data collection models.

The sensors located on the agricultural fields are responsible for sensing and gathering the required data from the fields. One of the main problems with these sensors is that these sensor nodes are limited in battery power and drain off easily during the data transmission process. Therefore, an energy-efficient transmission framework was designed by Lerdsuwan and Phunchongharn [31]. The framework enabled the nodes in the field to effectively collect the data by adapting to the environmental fluctuations. A data-driven algorithm was introduced using the greedy strategy to achieve a higher data rate and minimize energy during data transmissions. The algorithm also offered low complexity in implementations and succeeded in providing better energy efficiency than other traditional evaluation algorithms.

Alejandrino et al. [32] presented a protocol-independent data acquisition scheme to achieve effective precision farming. The proposed method supported inter-organizational communication by identifying optimal paths for mesh networks with diversity. The method also focused on decentralization so that the protocols run independently. Moreover, the optimal paths for data transmission were chosen without changes or modifications in the considered technology and architecture. The mobility of the systems were maintained with the integration of definite source configuration. Congestion was prevented, and the data load was balanced with the individual processing of every sensor. Tests were conducted by circulating messages of increasing size to identify the performance of the method. Also, the data transmission approach proved effective in terms of various metrics against several existing methodologies.

The use of metaheuristic optimization algorithms for the task of route optimization has been a higher focus in the research community to attain better performance. Cui et al. [34] introduced a route optimization mechanism with the integration of ant colony optimization (ACO) and genetic algorithm (GA). Initially, the ant that achieve the destination was considered as an individual and the crossover operation was applied to replace weaker parents. Then, the mutation operation was carried out to determine the next nodes. Another approach for route optimization was presented by Zamar et al. [35] where the K-means and nearest neighbour approach were combined to identify the best route. The bale collection problem (BCP) was rectified using the combination of k-means and nearest neighbour where the nodes were initially clustered and then route was optimized. Srivastava et al. [36] suggested a route optimization approach to suggest optimal routes for unmanned aerial vehicles (UAVs) in crop fields. The optimal routes of spray points were identified in the stressed regions with the help of travelling salesman problem (TSP) based routing algorithm and Voronoi diagram. The comparative analysis of the existing literature works is presented in Table 1.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6399
***Article Received**: 02 February 2023 **Revised**: 09 March 2023 **Accepted**: 14 March 2023*

Table 1: Comparative analysis of the existing literature works

Authors	Methods	Objective	Advantages	Disadvantages
Lakshmi et al. [21]	EERP	Energy efficient data transmission	Achieved higher coverage rate and resulted in elapsed network lifetime	This protocol results in overhead as clustering is carried out on each round
Mahajan and Badarla [22]	NICC	Energy efficient and optimal data transmission	Maintained better trade-off between energy efficiency and optimal data transmission	The BFO algorithm used in the work consists of fixed step size which makes it difficult to find a balance between exploration and exploitation and results in trapping into local optima
Agarkhed et al. [23]	PAwCOR	Optimal congestion-less data transmission	The methodology achieved reliable and effective data transmission	Since the method is based on decision support system, it results in information overload that is heavy to process
Miao et al. [24]	UCEEC	Energy efficient data transmission	Achieved optimal energy balance in the network via multi-hop route selection	This methodology is unable to identify the errors in the data records and is unable to recover the corruption issues
Li et al. [25]	Double selection strategy	Optimal data collection	Achieved optimal data collection with the coupling of edge computing with WSN	This method takes extra memory space for storage and may involve storage of invalid data
Lerdsuwan and Phunchongharn [31]	Greedy strategy	Energy efficient data transmission	Effectively adapted to environmental fluctuations and resulted in increased data rate	The outcome is not always optimal for larger and complex network environments
Alejandrino et al. [32]	Protocol independent data acquisition scheme	Optimal data acquisition	Congestion and load balancing issues are effectively addressed in this method	Several other network parameters are required to be optimized to achieve performance improvement
Cui et al. [34]	ACO, GA	Route optimization	Optimally identified the routes with the integration of two metaheuristics	The adoption of ACO created stagnation issues which resulted in congestion in the network
Zamar et al. [35]	K-means and nearest neighbor	Route optimization	Resulted in optimized routing and achieved effective transmission rate	The communication overhead of the method is high
Srivastava et al. [36]	TSP based routing algorithm and Voronoi diagram	Route optimization	Automated the entire process of identification and application of fertilizers to the field and reduced the need of manual labor	This method is capable of spraying only a single fertilizer at a time whereas, multiple fertilizers are needed based on different crop types

Upon reviewing the existing literary works, it has been identified that the existing works focused on enhancing the smart agriculture field by various means. Only a few works particularly focus on the data collection task in smart agriculture. Also, it is seen that data collection is a hard process, and an effective strategy for optimal data collection is an urgent need. One problem with the data collection process is the chosen area, as finding appropriate arrival paths for remote areas is difficult. Moreover, the quality of data collected from the fields is needed to be focused on, as invalid data may result in wrong observations and practices. The time complexity associated with the data collection is another major problem, as the sensors should search and identify an optimal path to transmit data that can quickly reach the sink. To overcome all the above problems, this work introduces a new and effective optimization based strategy for optimal data collection.

III. Proposed methodology

This work focuses on introducing a new and optimized data collection model to discover the optimal path for transmitting the data in edge and cloud based WSN. In the case of agricultural applications, the sensors are mounted on agricultural fields to collect the required information for crop production. The field data, including details from the soil, helps to take the necessary measures to acquire higher production rates. The edge and cloud assisted WSN framework helps to accurately acquire the data from the fields and offers efficient and feasible services. The proposed task optimization framework is intentionally designed to enhance the data collection task for smart agricultural applications. The QoD is assessed in the edge server based on certain quality parameters. The general architectures use a BS to transmit the data ultimately. This is replaced with the use of edge computing in the proposed work as the edge server is capable of temporarily storing the data to check the valid data. This procedure enhances the performance of the proposed method by allowing only the valid data to be stored in cloud servers. While transmitting the data from the sensors in the fields to the server, the optimal path between the sensors and the destination is identified using a nature inspired optimization strategy. The cloud and edge assisted WSN framework for smart agriculture considered in the proposed work is displayed in Figure 1.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6399
***Article Received**: 02 February 2023 **Revised**: 09 March 2023 **Accepted**: 14 March 2023*

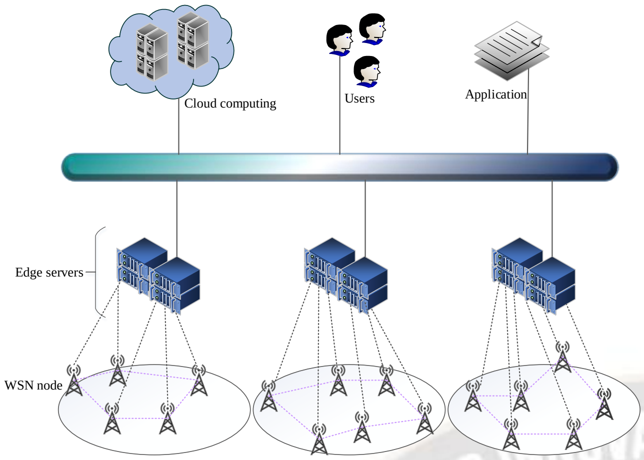

Figure 1: Cloud and edge assisted WSN for smart agriculture

The considered framework has three major layers: the sensing layer, the edge layer and the cloud layer. The sensing layer consists of sensors to collect information from the fields. The edge layer comprises servers close to the WSN nodes to efficiently provide on-time computing services. The edge server obtains all the sensed data from the sensors and checks for data validity. The valid data are then forwarded to the cloud servers for storage, and the invalid data are discarded.

A. System model

The system model for the proposed data collection system is presented in this section. The following assumptions are made to simplify the mathematical modelling of the work. A constant mesh topology is maintained, and the considered area is deployed with multiple sensors to sense multiple soil parameters, including nitrogen, potassium, phosphorous, etc. The data rate arriving at the sensors is constant, and every sensor in the field is assumed to have the same sensing capability. The agricultural area considered for data collection can be indicated as Ψ. The different nodes deployed in the field can be represented as $S = \{s_1, s_2, \ldots s_M\}$ where, M specifies the total number of nodes. The sensors that are present within the nodes can be indicated as $N = \{n_1, n_2, \ldots n_K\}$ where, K specifies the total number of sensors. The area considered comprises a different number of sensor nodes, and every node includes a different number of sensors to sense multiple soil parameters. This means that all the nodes within the area Ψ are capable of sensing all the soil parameters under study. The edge servers are assumed to be close to the sensors to ensure the required functionalities are on time. The function to describe that the sensor belongs to the WSN node can be indicated as follows:

$$f_{ij}(s_i, n_j) = \begin{cases} 1, & if\ s_i \mapsto n_j \\ 0, & otherwise \end{cases} \tag{1}$$

where, $s_i \mapsto n_j$ indicates the connection between node s_i and sensor n_j. The working principle of the proposed methodology is given in the flow diagram in Figure 2.

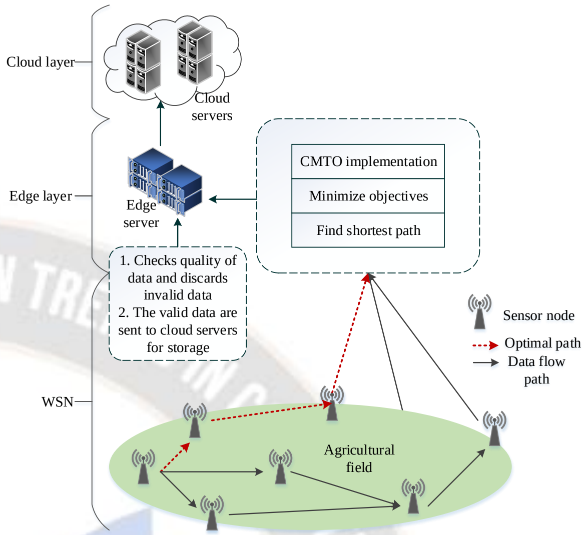

Figure 2: Flow diagram of the proposed CMTO task optimization framework

B. Problem formulation

The sensed data is transmitted to the edge server for data quality evaluation. The system evaluates the random path based on a fitness function to transmit the data to the edge server. For optimal path identification, this work introduces a new meta-heuristic based task optimization model based on the CMTO algorithm. The path is selected based on resource utilization, communication latency, execution cost, and distance. The fitness function is formulated as follows:

1) Resource utilization: Generally, in WSN, resource utilization indicates the quantity of energy expended by the network to complete a process or task. Here, the quantity of energy expended by the network to complete the sensing process is known as energy consumption [25]. The mathematical formulation for energy consumption can be given as follows:

$$E_n = \rho_1 t_s + \rho_2 t_c + \rho_3 t_w \tag{2}$$

where, ρ_1, ρ_2 *and* ρ_3 are the power consumed by the sensors for sensing, communication and waiting for the data and t_s, t_c *and* t_w are the time taken to sense, communicate and wait for the data.

2) Communication latency: Communication latency specifies the delay between the initialization and completion of the routing process. The mathematical formulation for communication latency can be given as follows:

$$L_t = F_t - I_t \tag{3}$$

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6399
***Article Received**: 02 February 2023 **Revised**: 09 March 2023 **Accepted**: 14 March 2023*

where, I_t indicates the initial time at which the data is sensed and F_t indicates the final time at which the data reaches the destination.

3) Execution cost: Execution cost implies the cost incurred in data transmission. In the proposed work, the quality of the link established for data transmission is considered to evaluate the execution cost. When there is a link failure, the execution cost of that transmission is increased. The mathematical formulation [33] for execution cost can be given as follows:

$$E_C = \rho_i + \left(1 / D_i(\beta) \right) + \left(1 / SNR_i \right) \tag{4}$$

where, ρ_i indicates the node energy, $D_i(\beta)$ is the distance between the node and edge server, and SNR_i is the signal-to-noise ratio of the node i mathematically indicated as follows:

$$SNR = \frac{received\ signal\ strength}{background\ noise} \tag{5}$$

4) Distance: Distance measure indicates the distance between two neighbouring sensor nodes in the network. One of the main aim of the proposed routing algorithm is to discover the best path with minimum distance between the sensor and edge server. The Euclidean distance metric is adopted to compute the distance between the two sensor nodes, and the mathematical formulation can be given as follows:

$$D_t(s_i, s_j) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \tag{6}$$

where, s_i *and* s_j indicates the node transmitting the sensed data and the destination node and x_i *and* y_j are the node coordinates. The minimum distance between the node and edge server indicates that the considered route is optimal.

5) Fitness function: The fitness function is formulated based on the minimization of the above objectives. The mathematical formulation for the fitness function is as follows:

$$F = \alpha * \frac{1}{E_n} + \beta * \frac{1}{L_t} + \gamma * \frac{1}{E_C} + \delta * \frac{1}{D_t(s_i, s_j)} \tag{7}$$

where, $\alpha + \beta + \gamma + \delta = 1$.

C. Cat and mouse based task optimization model

The proposed CMTO model works by optimizing the above formulated fitness function. The fitness function is evaluated for every route to find the optimal path and initiate the transmission.

1) Concept of CMTO algorithm: The CMTO is adopted for choosing the optimal path to engage in data transmissions. The CMTO is a population-based algorithm inspired by the attacking behaviour of cats on mice and the escaping behaviour of mice from the cats towards haven. The search space consists of two different populations such as the cat population and the mice population, and these populations are responsible for scanning the search space randomly. While scanning the search space, the fitness function is evaluated, and the position update for the members is performed in two stages. In the first stage, the cats' movement towards mice is exhibited, and in the second stage, the escaping scenario of mice towards haven is exhibited. Every member in the population is a solution to the problem, and where they are in the problem space determines the fitness solution. The main reason for choosing this algorithm is that this algorithm is more effective in identifying quasi-optimal solutions than the other algorithms. Moreover, the proposed problem is to identify the optimal path, and it is important to obtain solutions closer to global optima to attain higher performance.

A population matrix is initially constructed with the specification of objective function values for every problem variable. Then, the population matrix is sorted based on the objective function values and split into two to represent cat and mouse populations. For both the cat and mice population, respective population matrices are constructed. Then, the positions of the cat and mice are determined based on their natural movements and the update functions are formulated separately for both populations. For every iteration, the above-said procedures are repeated. The algorithm iteratively updates the position to solve the problem until a stopping criterion is met. The final obtained solution is known to be the quasi-optimal solution.

2) CMTO for optimal path identification; The proposed CMTO model's available routes to the edge server are initialized as the initial population. The problem is modelled in the search space, where the objective is to discover the optimal path for transmitting the sensed data to the edge server. Initially, the population in the problem space are represented using the matrix form as follows:

$$R = \begin{bmatrix} R_1 \\ . \\ . \\ R_i \\ . \\ . \\ R_N \end{bmatrix}_{N \times l} = \begin{bmatrix} r_{1,1} \ldots r_{1,d} \ldots r_{1,l} \\ \ldots \quad \ldots \quad \ldots \\ \ldots \quad \ldots \quad \ldots \\ r_{i,1} \ldots r_{i,d} \ldots r_{i,l} \\ \ldots \quad \ldots \quad \ldots \\ \ldots \quad \ldots \quad \ldots \\ r_{N,1} \ldots r_{N,d} \ldots r_{N,l} \end{bmatrix}_{N \times l} \tag{8}$$

where, R indicates the population matrix representing the possible routes to the edge server, R_i indicates the search

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6399
***Article Received**: 02 February 2023 **Revised**: 09 March 2023 **Accepted**: 14 March 2023*

agent i. $r_{i,d}$ indicates the objective value attained by the d^{th} problem variable by the search agent i, N indicates the total routes available and l is the total problem variable.

The routes represented in the matrix are evaluated using the fitness function formulated in equation (7). Based on the evaluation, ranking of the routes is done to rank them from best route to worst route. Based on the ranking, the routes in the matrix are sorted along with their objective function values. The sorted matrix can be represented as follows:

$$R^{\varsigma} = \begin{bmatrix} R_1^{\varsigma} \\ . \\ . \\ R_i^{\varsigma} \\ . \\ . \\ R_N^{\varsigma} \end{bmatrix}_{N\times l} = \begin{bmatrix} r_{1,1}^{\varsigma} \ldots r_{1,d}^{\varsigma} \ldots r_{1,l}^{\varsigma} \\ \ldots \quad \ldots \quad \ldots \\ \ldots \quad \ldots \quad \ldots \\ r_{i,1}^{\varsigma} \ldots r_{i,d}^{\varsigma} \ldots r_{i,l}^{\varsigma} \\ \ldots \quad \ldots \quad \ldots \\ \ldots \quad \ldots \quad \ldots \\ r_{N,1}^{\varsigma} \ldots r_{N,d}^{\varsigma} \ldots r_{N,l}^{\varsigma} \end{bmatrix}_{N\times l} \tag{9}$$

where, R^{ς} indicates the sorted matrix of routes, R_i^{ς} indicates the member i of the sorted matrix and $r_{i,d}^{\varsigma}$ indicates the value for d^{th} problem variable by search agent i. Now, the sorted matrix is split into two population matrices per the algorithm to represent cat and mouse populations. Here, the two populations are determined based on the obtained fitness function values. The population matrix with better fitness values is indicated as H, and the population with low fitness is indicated using P. The mathematical representation for both the population matrices can be represented as follows:

$$H = \begin{bmatrix} H_1 = R_1^{\varsigma} \\ . \\ . \\ H_i = R_i^{\varsigma} \\ . \\ . \\ H_{Nl} = R_{Nl}^{\varsigma} \end{bmatrix}_{Nl\times l} = \begin{bmatrix} r_{1,1}^{\varsigma} \ldots r_{1,d}^{\varsigma} \ldots r_{1,l}^{\varsigma} \\ \ldots \quad \ldots \quad \ldots \\ \ldots \quad \ldots \quad \ldots \\ r_{i,1}^{\varsigma} \ldots r_{i,d}^{\varsigma} \ldots r_{i,l}^{\varsigma} \\ \ldots \quad \ldots \quad \ldots \\ \ldots \quad \ldots \quad \ldots \\ r_{Nl,1}^{\varsigma} \ldots r_{Nl,d}^{\varsigma} \ldots r_{Nl,l}^{\varsigma} \end{bmatrix}_{N_l\times l} \tag{10}$$

$$\mathrm{P} = \begin{bmatrix} \mathrm{P}_1 = R_{Nl+1}^{\varsigma} \\ . \\ . \\ \mathrm{P}_j = R_{Nl+j}^{\varsigma} \\ . \\ . \\ \mathrm{P}_{Np} = R_{Nl+Np}^{\varsigma} \end{bmatrix}_{N_p\times l} = \begin{bmatrix} r_{Nl+1,1}^{\varsigma} \ldots r_{Nl+1,d}^{\varsigma} \ldots r_{Nl+1,l}^{\varsigma} \\ \ldots \quad \ldots\ldots \quad \ldots\ldots \\ \ldots \quad \ldots\ldots \quad \ldots\ldots \\ r_{Nl+j,1}^{\varsigma} \ldots r_{Nl+j,d}^{\varsigma} \ldots r_{Nl+j,l}^{\varsigma} \\ \ldots \quad \ldots\ldots \quad \ldots\ldots \\ \ldots \quad \ldots\ldots \quad \ldots\ldots \\ r_{Nl+N_p,1}^{\varsigma} \ldots r_{Nl+N_p,d}^{\varsigma} \ldots r_{Nl+N_p,l}^{\varsigma} \end{bmatrix}_{N_p\times l} \tag{11}$$

where, H indicates the population matrix with higher fitness, P indicates the population matrix with lower fitness. N_l is the total number of routes with better fitness, H_i indicates the i^{th} route with high fitness. N_p is the number of routes with lower fitness and P_j is the j^{th} route with low fitness.

Hereafter, the routes with better fitness can be termed as the mice population, and the routes with low fitness can be termed as the cat population according to the considered optimization algorithm. The position update formulation based on the natural behaviour of cats can be modelled in the initial phase as follows:

$$\mathrm{P}_j^{new} : p_{j,d}^{new} = p_{j,d} + rnd \times \left(l_{k,d} - M \times p_{j,d}\right) \& \; j = 1:N_p, d = 1:l, k \in 1:N_l \tag{12}$$

$$M = round\left(1 + rnd_1\right) \tag{13}$$

$$\mathrm{P}_j = \begin{cases} \mathrm{P}_j^{new}, \left| F_j^{p,new} < F_j^{p} \right. \\ \mathrm{P}_j, \left| else \right. \end{cases} \tag{14}$$

where, P_j^{new} indicates the new grade of j^{th} cat, $p_{j,d}^{new}$ indicates the new value of d^{th} problem variable attained by j^{th} cat, rnd and rnd_1 are the random variables within the range [0, 1], $l_{k,d}$ indicates the k^{th} mouse's dimension, $F_j^{p,new}$ indicates the new objective function value attained by j^{th} cat.

In the next phase, the mice's escape to the havens is modelled, and the position formulation is introduced for the mice population. Based on the positions of the mice in the search space, the haven's position is randomly created in the search space. According to CMTO, there is a random haven for every mouse in the population. The mathematical formulation to update the position of the mouse can be represented as follows:

$$\mathrm{T}_i : t_{i,d} = r_{l,d} \;\&\; i = 1:N_l, d = 1:l \in 1:N \tag{15}$$

$$H_i^{new} : l_{i,d}^{new} = l_{i,d} + rnd \times \left(t_{i,d} - M \times l_{i,d}\right) \times sign\left(F_i^{l} - F_i^{\mathrm{T}}\right) \& \quad i = 1:N_l, d = 1:l \tag{16}$$

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6399
***Article Received**: 02 February 2023 **Revised**: 09 March 2023 **Accepted**: 14 March 2023*

$$H_i = \begin{cases} H_i^{new}, & \left| F_i^{l,new} < F_i^l \right. \\ H_i, & \left| else \right. \end{cases} \tag{17}$$

where, T_i *and* F_i^{T} indicates the haven of i^{th} mouse and its fitness value, H_i^{new} *and* $F_i^{l,new}$ indicates the new grade of the mouse and its fitness value. The pseudocode of the proposed CMTO-based data collection model is presented in Algorithm 1.

Algorithm 1: CMTO based data collection model
Input: Random paths for data transmission, algorithmic parameters Output: Optimal path for data transmission 1. The population matrix is initialized with each random path indicating a solution for the considered problem. 2. The objective values for all the random paths are defined, and the matrix is modified with the fitness specification. 3. The ranking is carried out for the paths based on the objective function value from the best path to the worst path. 4. The routes defined in the matrix are then sorted with respect to the ranks provided. 5. From the sorted matrix obtained, two population metrics are constructed to represent the cat and mouse population. 6. The population matrix with higher fitness is indicated as H (mouse), and the population matrix with lower fitness is indicated as P (cat). 7. Based on the natural behaviour of cats, the position of routes in the matrix P is updated using equation (12). 8. The escape behaviour of mice is modelled, and the position of routes in the matrix H is updated using equation (15). 9. The new grades of each route in the matrix H are identified, and the fitness values are updated. 10. The best optimal path obtained is returned for data transmission.

The entire steps are reiterated to find the optimal route for the sensor to reach the edge server. A new route is obtained for every iteration, and the route fitness is evaluated based on the defined objectives. After a certain number of iterations, the algorithm exhibits an error, and the fitness values of the solutions get degraded. A stopping condition is defined at this point, and the algorithm is terminated. The final obtained solution is termed the quasi-optimal solution, and this solution is selected as the best route, and the transmission of data is initiated. In this way, the proposed algorithm finds the optimal route to achieve an optimal data collection task.

D. Data quality evaluation

The transmitted data by the sensors in the field finally reaches a destination known as the edge server, which is responsible for evaluating the data quality and forwarding it to the cloud. The main aim of this process is to eliminate invalid data and to transmit only valid data so that the storage space in the cloud will not be wasted. The quality of the data entering the edge server is ensured based on the data quality parameters. This work ensures the four main quality parameters: completeness, correctness, precision, and data reliability. The edge server maintains a constant threshold for all of these parameters in percentage to ensure data validity for propagation to cloud servers. Initially, the completeness [37] of the data is measured based on the following formulation:

$$c1 = \left(1 - \frac{count\ of\ \ incomplete\ cells}{count\ of\ \ cells}\right) * 100 \tag{18}$$

$$c2 = \left(1 - \frac{count\ of\ \ incomplete\ rows}{count\ of\ \ rows}\right) * 100 \tag{19}$$

$$Q_1 = c1 + c2 \tag{20}$$

The completeness of the data can be indicated as Q_1. The data unavailable in the database are identified accurately in terms of percentage using the completeness measure. When the percentage value goes below the defined threshold, the piece of data sent by the sensor is discarded. It is not forwarded to the cloud to avoid unnecessary storage usage.

The correctness of the data is then measured using the following mathematical formulation:

$$Q_2 = \frac{(total\ \ records\ in\ \ dataset - count\ \ of\ \ errors\ \det ected) * 100}{total\ \ records\ in\ \ dataset} \tag{21}$$

Correctness of the data is another important quality measure capable of contributing to the reliability of the data. As more errors are detected in the arrived data, the correctness of the data is reduced.

The next parameter used to evaluate data quality is precision, measured by the correctness of storage. The data stored in an incorrect format are considered data with insufficient precision. The mathematical formulation to measure the precision of data can be represented as follows:

$$Q_3 = \frac{count\ of\ \ records\ without\ enough\ precision * 100}{count\ of\ records\ present in\ database * count of\ attributes\ present in\ database} \tag{22}$$

The above formulation gives the data precision, contributing to the calculation of data reliability. When the precision percentage of the data is high, the percentage of data to be forwarded to the cloud for storage is high. The reliability

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6399
***Article Received**: 02 February 2023 **Revised**: 09 March 2023 **Accepted**: 14 March 2023*

of the data is then measured from the metrics calculated before. The mathematical formulation for reliability can be given as follows:

$$Reliability = 100 - \left(100 - Q_1 + \max\{100 - Q_2, 100 - Q_3\}\right) \quad (23)$$

The value obtained for reliability is then compared with the pre-defined threshold value (Th) to determine the quality and validity of the data. The threshold value is manually set, which is half of the entire percentage of reliability value provided by the data. In our work, the value of the threshold is set to 0.5. The data that surpasses the quality test is then forwarded to the cloud for storage. The condition to ensure the reliability of the data can be given as follows:

$$DQ = \begin{cases} 1, & if\ Reliability \geq Th \\ 0, & otherwise \end{cases} \quad (24)$$

In the above condition, 1 indicates that the data quality is high and can be moved to the cloud for storage and 0 indicates the data is to be discarded.

IV. Results and discussion

The performance of the proposed algorithm is proved through different experiments and evaluations. Different algorithms are implemented under the same simulation environment to ensure the fairness of analysis. The simulation environment adopted, performance metrics used, and performance analysis conducted is explained in the upcoming sections.

A. Simulation scenario

The proposed algorithm has been implemented sequentially by initially building a WSN framework in smart agriculture. A total of 100 sensors are present in the simulated agricultural environment. The area chosen for implementation can be given as 1250×1250m. Each of these sensors can sense the environment with the same sensing capability and acquire details of 11 different soil nutrient components. For evaluation purposes, a crop recommender dataset downloaded from Kaggle is utilized. The dataset can be found on the Kaggle website and is publicly available [26]. The dataset includes 11 columns with 100 rows, and the total number of records present in the dataset is 1100. This dataset comprises 11 components: nitrogen, phosphorous, potassium, soil pH, electrical conductivity, sulphur, copper, iron, manganese, zinc, and boron. An edge server module is created as the destination for the sensors to transmit the data. The complete simulation of the work is done in the Python platform. The system configuration considered for implementation is as follows: a system comprised an Intel core processor of 3.3 GHz with an installed RAM of 8GB running on a 64-bit windows 10 operating system.

B. Performance metrics

The major performance metrics considered in the proposed work for analysis include energy consumption, communication latency, execution cost, completeness, precision, correctness and reliability. The mathematical formulations for these metrics can be seen in sections 3.2 and 3.4.

C. Performance analysis

The performance of the proposed method has been analyzed based on the comparison with the existing optimization algorithms such as the COOT optimization algorithm [27], genetic algorithm [28], wild horse optimization [29] and flow direction algorithm [30]. These algorithms are the latest ones and proved to provide better results for different problems with better adaptability. Moreover, the genetic algorithm is also chosen for comparison as it is one of the most reliable algorithms is solving different engineering problems. All these chosen algorithms are implemented in the same scenario to ensure the fairness of analysis. The performance values attained by these implemented algorithms are compared with the proposed algorithm's performance values, and discussions are provided. A deep analysis of the proposed model is explained below.

1) Analysis of data collection task: In the proposed model, two main tasks are achieved: the data collection task and the data quality measurement task. Therefore, separate analyzes are provided for these two tasks to explain the performance improvement of the proposed technique and its significance. This section provides the performance analysis of the proposed data collection task and the CMTO model is evaluated, and the results are presented.

Table 2: Energy consumption analysis of the proposed and existing methods

No. of sensors	COOT	GA	WHO	FDA	Proposed
20	14.712	7.846	6.865	18.636	3.923
40	13.731	7.846	6.865	19.616	4.904
60	12.750	8.827	6.865	18.636	4.904
80	13.731	8.827	6.865	19.616	5.885
100	14.712	9.808	7.846	19.616	6.865

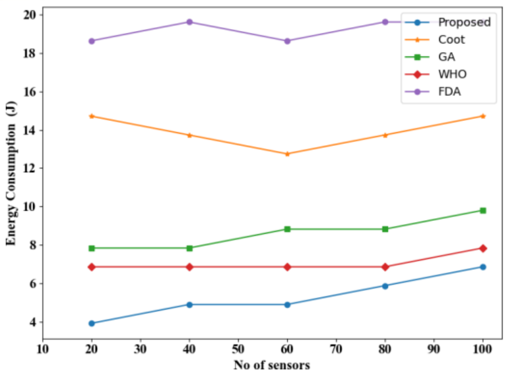

Figure 3: Graphical representation for energy consumption analysis

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6399
***Article Received:** 02 February 2023 **Revised:** 09 March 2023 **Accepted:** 14 March 2023*

The energy consumed by the proposed and existing algorithms to complete the data collection process is presented in Table 2. The values obtained clearly show that the proposed approach significantly reduced the need for energy to be consumed compared to the other algorithms. A graphical representation of energy consumption is presented in Figure 3. The figure shows that the proposed model is more optimal in energy consumption than the existing models. The edge server maintains a constant threshold for all of these parameters in percentage to ensure data validity for propagation to cloud servers.

Moreover, the algorithm's efficiency is high, which quickens the route selection process resulting in reduced energy consumption. The WHO algorithm consumed less energy among the compared algorithms, whereas the FDA algorithm provided poor performance. The energy values are obtained by varying the range of sensors from 20 to 100. The overall energy consumed by the proposed approach for 100 sensors is 6.865J, whereas the compared COOT, GA, WHO and FDA algorithms resulted in 14.712J, 9.808J, 7.846J and 19.616J, respectively.

Table 3: Communication latency analysis of the proposed and existing methods

No. of sensors	COOT	GA	WHO	FDA	Proposed
20	11.77	6.86	16.67	19.61	1.96
40	12.75	7.84	19.61	24.52	2.94
60	13.73	8.82	29.42	34.32	3.92
80	14.22	9.80	39.23	48.06	4.90
100	14.71	11.77	49.04	58.85	6.86

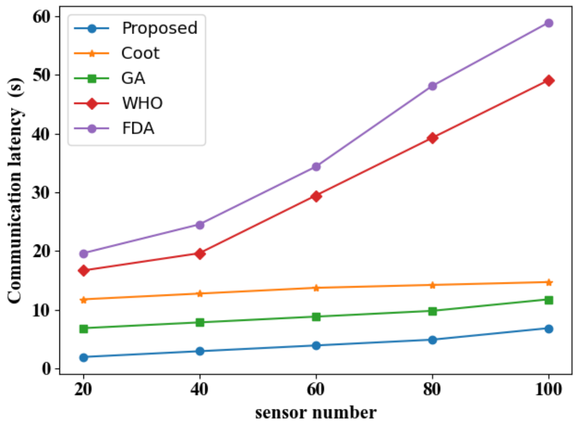

Figure 4: Graphical representation for communication latency analysis

The performance of the proposed and existing algorithms in communication latency is analyzed, and the results are presented in Table 3. The values obtained clearly show that the proposed algorithm resulted in very low communication latency compared to the existing algorithms. This can also be seen in the graphical representation depicted in Figure 4. From the figure, it is clear that the proposed algorithm provided better results for a varying number of sensors. Among the compared algorithms, optimal performance is seen in the GA algorithm, and FDA resulted in poor performance in terms of communication latency. The overall communication latency attained by the proposed algorithm for 100 sensors is 6.86s, whereas the existing algorithms, such as COOT, GA, WHO and FDA, resulted in 14.71s, 11.77s, 49.04s and 58.85s, respectively.

Table 4: Execution cost analysis of the proposed and existing methods

No. of sensors	COOT	GA	WHO	FDA	Proposed
20	4.904	7.846	9.808	11.770	1.961
40	5.394	8.337	10.298	13.241	2.452
60	5.885	8.827	10.789	14.712	2.942
80	6.865	9.808	11.770	16.183	3.923
100	7.356	10.298	12.260	16.674	4.413

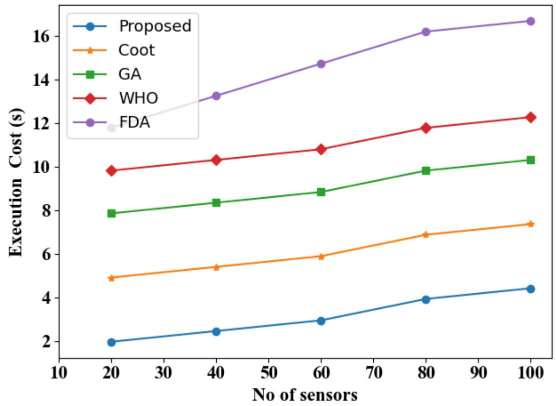

Figure 5: Graphical representation for execution cost analysis

The execution cost of the proposed and existing algorithms are analyzed, and the results are displayed in Table 4. The values in the table clearly show that the proposed method is optimal in reducing the overall execution cost needed compared to the existing algorithms. The graphical depiction in Figure 5 proves the algorithm's effectiveness in reducing the overall execution cost required. Among the compared algorithms, COOT provided optimal performance, and FDA resulted in poor performance. The overall execution cost of the proposed algorithm for 100 sensors is 4.413s, whereas the execution costs of the compared algorithms, such as COOT, GA, WHO and FDA, are 7.356s, 10.298s, 12.260s and 16.674s, respectively.

Table 5: Network overhead analysis of the proposed and existing methods

No. of sensors	COOT	GA	WHO	FDA	Proposed
20	34.32	39.23	44.13	49.04	29.42
40	39.23	44.13	49.04	53.94	34.32
60	44.13	49.04	53.94	58.85	39.23
80	49.04	53.94	58.85	63.75	44.13
100	58.85	60.81	66.69	68.65	49.04

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6399
***Article Received**: 02 February 2023 **Revised**: 09 March 2023 **Accepted**: 14 March 2023*

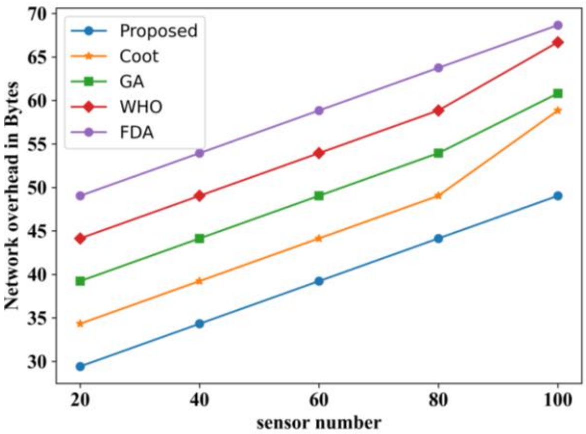

Figure 6: Graphical representation for network overhead analysis

The network overhead of the proposed and existing algorithms are analyzed and the results are presented in Table 5. From the table, it is clearly evident that the proposed method resulted in minimum overhead compared to the existing algorithms. Thus, it can be suggested that the proposed algorithm is highly suitable for the data collection time than the existing algorithms. This is also evident from the graphical representation displayed in Figure 6. The overhead of the method is analyzed by varying the number of sensors from 20 to 100 and for different sensor count, the proposed method resulted in optimal values. The overall network overhead of the proposed method for 100 sensors is 49.04bytes whereas, the overhead of the compared algorithms such as COOT, GA, WHO and FDA are 58.85 bytes, 60.81 bytes, 66.69 bytes and 68.65 bytes respectively.

2) Analysis of data quality: This section analyses the data quality after optimally finding the edge server. The need for data quality measurement is to ensure the data validity to be stored in the cloud so that the wastage of cloud memory can be reduced. The data that passes the quality analysis test alone is transferred to the cloud servers for storage. The analysis of the proposed method for data quality is done in terms of correctness, precision, completeness and reliability. The performance analysis and descriptions are presented below.

Table 6: Completeness analysis of the proposed and existing methods

No. of sensors	COOT	GA	WHO	FDA	Proposed
20	91	89	83	78	92
40	91	86	81	76	92
60	91	86	81	79	92
80	88	79	79	75	92
100	88	87	83	77	90

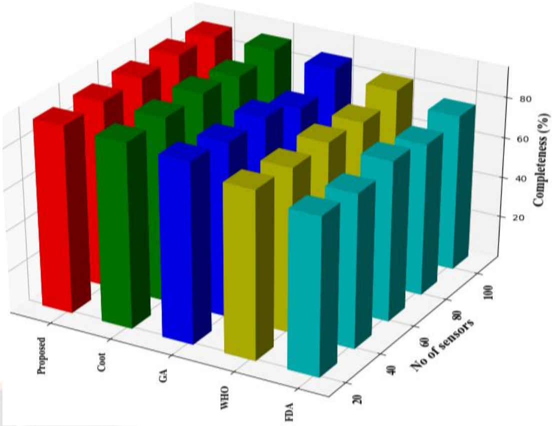

Figure 7: Graphical representation for completeness analysis

The completeness of data provided by the proposed method is compared with the existing methods, and the results are presented in Table 6. The values obtained clearly show that the proposed method resulted in optimal performance compared to the existing methods. This is also highlighted in the graphical representation provided in Figure 7. Compared to the existing methods, the percentage of completeness of the data provided by the proposed method is high. This proves that the proposed method accurately measures data quality, and only the valid data can be passed to the cloud, meanwhile avoiding the loss of relevant data. The values are calculated by varying the sensor number from 20 to 100. The overall completeness achieved by the proposed algorithm for 100 sensors is 90%, whereas the compared algorithms, such as COOT, GA, WHO and FDA, resulted in 88%, 87%, 83% and 77% of completeness, respectively.

Table 7: Precision analysis of the proposed and existing methods

No. of sensors	COOT	GA	WHO	FDA	Proposed
20	87	86	77	76	88
40	80	76	75	75	88
60	88	83	83	82	87
80	88	87.6	84	84	91
100	89	86	83	77	90

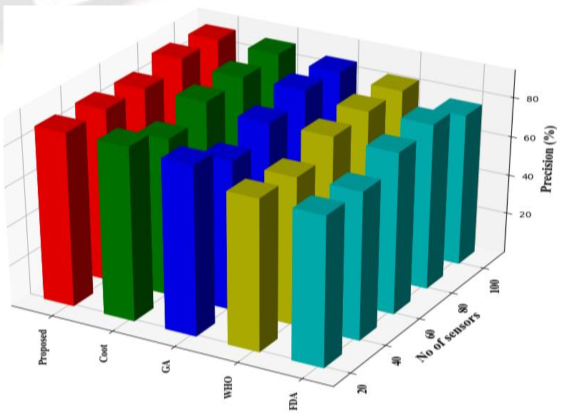

Figure 8: Graphical representation for precision analysis

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6399
***Article Received**: 02 February 2023 **Revised**: 09 March 2023 **Accepted**: 14 March 2023*

The precision of the proposed method is compared with the existing methods, and the results are presented in Table 7. From the values, it is clear that the performance of the proposed method is more optimal than the existing algorithms. The proposed algorithm can retrieve precise data and maintain data quality within the defined threshold. The optimality of the data is ensured well by the proposed method. The values are noted by varying the sensor count from 20 to 100. This is shown in the graphical depiction in Figure 8. Among the compared algorithms, COOT resulted in better performance, followed by GA and FDA resulted in low performance. The overall precision value attained by the proposed method for 100 sensors is 90%, whereas the existing algorithms, such as COOT, GA, WHO and FDA, resulted in 89%, 86%, 83% and 77% of precision, respectively.

Table 8: Correctness analysis of the proposed and existing methods

No. of sensors	COOT	GA	WHO	FDA	Proposed
20	88	80	78	75	90
40	90	86	84	81	90
60	87	82	80	79	87
80	84	84	85	77	90
100	84	83	79	76	92

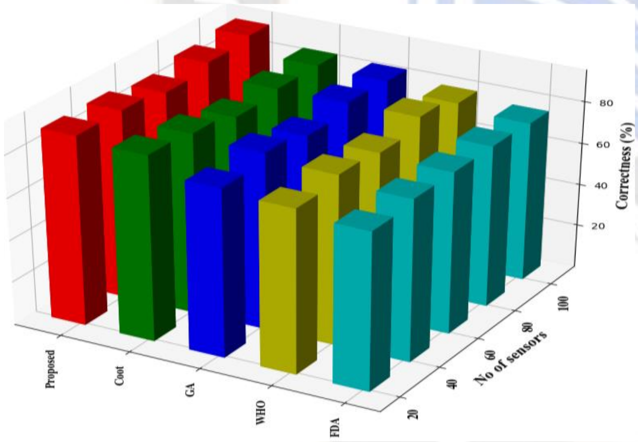

Figure 9: Graphical representation for correctness analysis

The proposed method's performance in terms of data correctness is shown in Table 8. From the values obtained, it is significant that the proposed method resulted in higher correctness values than the existing algorithms. The correctness of the data is more accurately measured by the proposed model than by the other methods. The data's correctness values are obtained by varying the sensor count from 20 to 100. This is shown in the graphical depiction displayed in Figure 9. COOT and GA resulted in optimal performance among the compared algorithms, and FDA produced poor results in all sensor counts. The overall correctness of data attained by the proposed method is 92%. In contrast, the correctness of data produced by the existing methods such as COOT, GA, WHO and FDA is84%, 83%, 79% and 76%, respectively.

Table 9: Reliability analysis of the proposed and existing methods

No. of sensors	COOT	GA	WHO	FDA	Proposed
20	90	87	80	75	91
40	87	85	83	75	88
60	86	84	82	76	92
80	85	80	75	75	86
100	89	88	85	80	92

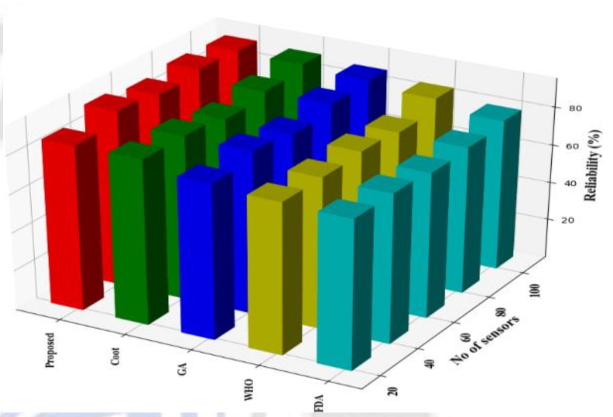

Figure 10: Graphical representation for reliability analysis

Reliability is the major metric used to actually measure the data quality and to which the threshold value is compared to determine the data quality. The reliability of the proposed method is compared to that of existing methods, and the results are displayed in Table 9. From the values in the table, it is clear that the proposed method produced higher values of reliability for all the sensor counts. Also, the existing algorithms resulted in lower values, as shown in the graphical representation in Figure 10. COOT resulted in better values among the compared algorithms, followed by GA and FDA algorithms attaining the least performance. The overall reliability attained by the proposed model for 100 sensors is 92%, whereas the reliability values secured by the existing algorithms such as COOT, GA, WHO and FDA are 89%, 88%, 85% and 80%, respectively.

D. Time complexity analysis

The time complexity of the proposed work is one of the major focuses and is analyzed in this section. The data collected from the sensors must be transmitted through optimal routes to reach the destination. The data collected by the sensors is one of the prime factors that result in more time, thereby leading to time complexity. Thus, this time is required to be reduced to enable efficient transmissions. The graphical representation for time complexity analysis is presented in Figure 11.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6399
***Article Received**: 02 February 2023 **Revised**: 09 March 2023 **Accepted**: 14 March 2023*

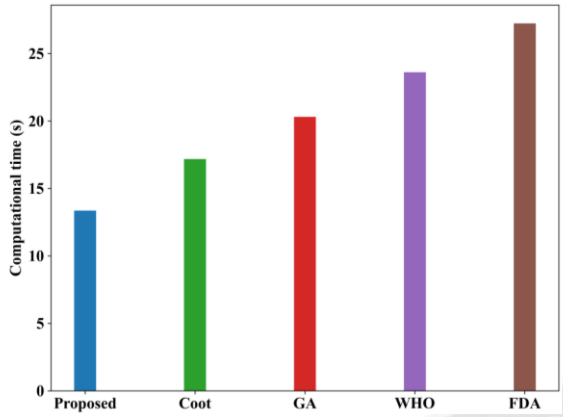

Figure 11: Graphical representation for time complexity analysis

From the figure, the computational time taken by the proposed method is very much low compared to the other algorithms. The graph has been plotted for the total time taken by 100 sensors in the field. The proposed work's data collection time is 13.36s for 100 sensors whereas, the data collection time taken by COOT, GA, WHO and FDA algorithms are 17.18s, 20.31s, 23.62s and 27.23s respectively. This value is found to be more optimal than many existing literature methods meant for data collection. The data collection time is calculated for the proposed work for 100 sensors. The value might reduce when the sensor count is decreased. This proves that the proposed work is more optimal and efficient than the other existing methodologies.

The overall analysis proves that the performance of the proposed method is more optimal than the other methods and is capable of efficiently completing the data collection and transmission processes. Moreover, the data quality after collection is also measured, and the analysis proved its effectiveness. Generally, the sensors in the field take more time to collect the data, and the transmission process is delayed. Therefore, it is advisable to optimize the data collection time and the whole data collection and transmission process. In our work, the data collection task's efficacy is enhanced by applying an optimization algorithm. The placement of edge servers before cloud storage enhanced the reliability of data and the framework, thereby dealing with the wastage in memory of cloud servers. Thus, it is suggested that the proposed method can be implemented to attain optimal results in data collection and transmission than the existing methodologies.

V. Conclusion

In this work, a new data collection strategy in smart farming with the integration of WSN and edge computing is proposed. The proposed method initially places the sensors in the agricultural field to collect the required data, including soil properties, rain factors, leaf and plant properties, etc., to increase productivity. The sensed data are then transmitted to the edge server by finding the optimal routes. This is done with the help of the CMTO algorithm that finds the optimal routes based on energy consumption, execution cost, distance and communication latency. The optimal route is chosen based on an iterative process, and the resultant route is provided as the output by the algorithm. Through the optimal path, the data is transmitted and reaches the edge server. The edge server then evaluates the validity of the data to transmit it to the cloud servers. The data quality measures such as precision, correctness, completeness and reliability are utilized to assure validity. A threshold condition is pre-defined, and the data that satisfy the thresholding condition is moved to the cloud servers for storage. The entire work implementations are carried out in the Python environment and are evaluated using the crop recommender dataset obtained from the Kaggle website. The simulation results proved that the proposed algorithm is much more effective than the other methods chosen for comparison. In the future, it is planned to use other hybrid strategies to attain more optimal outcomes and see improvement in efficiency.

References

[1] G. Sahitya, N. Balaji, and C.D. Naidu. "Wireless sensor network for smart agriculture." In 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), IEEE, pp. 488-493, 2016.

[2] M.H. Anisi, G. Abdul-Salaam, and A.H. Abdullah. "A survey of wireless sensor network approaches and their energy consumption for monitoring farm fields in precision agriculture." Precision Agriculture, vol. 16, no. 2, pp. 216-238, 2015.

[3] S.A. Kumar, and P. Ilango. "The impact of wireless sensor network in the field of precision agriculture: A review." Wireless Personal Communications, vol. 98, no. 1, pp. 685-698, 2018.

[4] Y. Kim, P. Bae, J. Han, and Y.B. Ko. "Data aggregation in precision agriculture for low-power and lossy networks." In 2015 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM), IEEE, pp. 438-443, 2015.

[5] C. Jamroen, P. Komkum, C. Fongkerd, and W. Krongpha. "An intelligent irrigation scheduling system using low-cost wireless sensor network toward sustainable and precision agriculture." IEEE Access, vol. 8, pp. 172756-172769, 2020.

[6] H.M. Jawad, R. Nordin, S.K. Gharghan, A.M. Jawad, and M. Ismail. "Energy-efficient wireless sensor networks for precision agriculture: A review." Sensors, vol. 17, no. 8, pp. 1781, 2017.

[7] H. Agrawal, R. Dhall, K.S.S. Iyer, and V. Chetlapalli. "An improved energy efficient system for IoT enabled precision agriculture." Journal of ambient intelligence and humanized computing, vol. 11, no. 6, pp. 2337-2348, 2020.

[8] V. Pandiyaraju, R. Logambigai, S. Ganapathy, and A. Kannan. "An energy efficient routing algorithm for WSNs

using intelligent fuzzy rules in precision agriculture." Wireless Personal Communications, vol. 112, no. 1, pp. 243-259, 2020.

[9] R. Morais, N. Silva, J. Mendes, T. Adão, L. Pádua, J.A. López-Riquelme, N. Pavón-Pulido, J.J. Sousa, and E. Peres. "Mysense: A comprehensive data management environment to improve precision agriculture practices." Computers and Electronics in Agriculture, vol. 162, pp. 882-894, 2019.

[10] M.N. Akhtar, A.J. Shaikh, A. Khan, H. Awais, E.A. Bakar, and A.R. Othman. "Smart sensing with edge computing in precision agriculture for soil assessment and heavy metal monitoring: A review." Agriculture, vol. 11, no. 6, pp. 475, 2021.

[11] S. Rajeswari, K. Suthendran, and K. Rajakumar. "A smart agricultural model by integrating IoT, mobile and cloud-based big data analytics." In 2017 international conference on intelligent computing and control (I2C2), IEEE, pp. 1-5, 2017.

[12] A.I. Montoya-Munoz, and O.M.C. Rendon. "An approach based on fog computing for providing reliability in iot data collection: A case study in a colombian coffee smart farm." Applied Sciences, vol. 10, no. 24, pp.8904, 2020.

[13] V. Mihai, C.E. Hanganu, G. Stamatescu, and D. Popescu. "WSN and fog computing integration for intelligent data processing." In 2018 10th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), IEEE, pp. 1-4, 2018.

[14] C. Zhang, X. Liu, X. Zheng, R. Li, and H. Liu. "Fenghuolun: A federated learning based edge computing platform for cyber-physical systems." In 2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), IEEE, pp. 1-4, 2020.

[15] H. Liazid, and M. Lehsaini. "A brief review on integration between wireless sensor networks and Cloud." Concurrency and Computation: Practice and Experience, vol. 33, no. 20, pp. e6328, 2021.

[16] P. Hu, S. Dhelim, H. Ning, and T. Qiu. "Survey on fog computing: architecture, key technologies, applications and open issues." Journal of network and computer applications, vol. 98, pp. 27-42, 2017.

[17] F.J. Ferrández-Pastor, J.M. García-Chamizo, M. Nieto-Hidalgo, J. Mora-Pascual, and J. Mora-Martínez. "Developing ubiquitous sensor network platform using internet of things: Application in precision agriculture." Sensors, vol. 16, no. 7, pp. 1141, 2016.

[18] M.J. OGrady, D. Langton, and G.M.P. OHare. "Edge computing: A tractable model for smart agriculture?." Artificial Intelligence in Agriculture, vol. 3, pp. 42-51, 2019.

[19] T. Wang, Y. Lu, Z. Cao, L. Shu, X. Zheng, A. Liu, and M. Xie. "When sensor-cloud meets mobile edge computing." Sensors, vol. 19, no. 23, pp. 5324, 2019.

[20] S. Premkumar, and A.N. Sigappi. "A survey of architecture, framework and algorithms for resource management in edge computing." EAI Endorsed Transactions on Energy Web, vol. 8, no. 33, pp.e15-e15, 2021.

[21] L.T. Aishwarya, B. Hariharan, and P. Rekha. "An Energy Efficient Routing Protocol for Internet of Things Based Precision Agriculture." In International Conference on Inventive Computation Technologies, Springer, Cham, pp. 684-691, 2019.

[22] H.B. Mahajan, and A. Badarla. "Cross-layer protocol for WSN-assisted IoT smart farming applications using nature inspired algorithm." Wireless Personal Communications, vol. 121, no. 4, pp. 3125-3149, 2021.

[23] J. Agarkhed, P.Y. Dattatraya, and S. Patil. "Precision agriculture with cluster-based optimal routing in wireless sensor network." International Journal of Communication Systems, vol. 34, no. 10, pp. e4800, 2021.

[24] Y. Miao, C. Zhao, and H. Wu. "Non-uniform clustering routing protocol of wheat farmland based on effective energy consumption." International journal of agricultural and biological engineering, vol. 14, no. 3, pp. 163-170, 2021.

[25] X. Li, L. Zhu, X. Chu, and H. Fu. "Edge computing-enabled wireless sensor networks for multiple data collection tasks in smart agriculture." Journal of Sensors, vol. 2020, 2020.

[26] https://www.kaggle.com/datasets/manikantasanjayv/crop-recommender-dataset-with-soil-nutrients

[27] I. Naruei and F. Keynia. "A new optimization method based on COOT bird natural life model." Expert Systems with Applications, vol. 183, pp. 115352, 2021.

[28] S. Katoch, S.S. Chauhan, and V. Kumar. "A review on genetic algorithm: past, present, and future." Multimedia Tools and Applications, vol. 80, no. 5, pp. 8091-8126, 2021.

[29] I. Naruei, and F. Keynia. "Wild horse optimizer: A new meta-heuristic algorithm for solving engineering optimization problems." Engineering with Computers, pp. 1-32, 2021.

[30] H. Karami, M.V. Anaraki, S. Farzin, and S. Mirjalili. "Flow Direction Algorithm (FDA): a novel optimization approach for solving optimization problems." Computers & Industrial Engineering, vol. 156, pp. 107224, 2021.

[31] P. Lerdsuwan, and P. Phunchongharn. "An energy-efficient transmission framework for IoT monitoring systems in precision agriculture." In International Conference on Information Science and Applications, Springer, Singapore, pp. 714-721, 2017.

[32] J.D. Alejandrino, R.S. Concepcion II, V.J.D. Almero, M.G. Palconit, R.R.P. Vicerra, A. Bandala, E. Sybingco, and E.P. Dadios. "Protocol-independent data acquisition for precision farming." Journal of Advanced Computational Intelligence and Intelligent Informatics, vol. 25, no. 4, pp. 397-403, 2021.

[33] K. Haseeb, I.U. Din, A. Almogren, and N. Islam. "An energy efficient and secure IoT-based WSN framework: An application to smart agriculture." Sensors, vol. 20, no. 7, pp. 2081, 2020.

[34] Y. Cui, W. Liu, Z. Zhao, Wireless Sensor Network Route Optimization Based on Improved Ant Colony-Genetic Algorithm, International Journal of Online Engineering, vol. 11, no. 9, 2015.

ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6399
***Article Received**: 02 February 2023 **Revised**: 09 March 2023 **Accepted**: 14 March 2023*

[35] D.S. Zamar, B. Gopaluni, S. Sokhansanj, A constrained k-means and nearest neighbor approach for route optimization: With an application to the bale collection problem, InThe 20th World Congress of the International Federation of Automatic Control 2017.

[36] K. Srivastava, P.C. Pandey, J.K. Sharma, An approach for route optimization in applications of precision agriculture using UAVs, Drones, vol. 4, no. 3, pp. 58, 2020.

[37] M.B. Yankelevich, A Homogeneous Framework to Measure Data Quality.