Data Mining–Assisted Collision-Aware Routing in Wireless Sensor Networks Using Multi-Objective Dolphin Optimization

Ms. S. Sangeetha[1], Dr. B. Srinivasan[2], Dr. P. Prabhusundhar[3]
[1]Assistant Professor, Department of Computer Science, VET Institute of Arts and Science (co-education) College, Thindal, Erode, Tamilnadu, India
[2]Associate Professor, Department of Computer Science, Gobi Arts and Science College (Autonomous), Gobichettipalayam, Tamilnadu, India
[3]Assistant Professor, Department of Computer Science, Gobi Arts and Science College (Autonomous), Gobichettipalayam, Tamilnadu, India

Abstract: Efficient routing in dense Wireless Sensor Networks (WSNs) is severely affected by packet collisions, repeated retransmissions, and unbalanced energy depletion. To address these shortcomings, the paper proposes a Data Mining-Assisted Collision-Aware Routing System using Multi-Objective Dolphin Optimization (MODO-CAR), in which routing choices are made from acquired knowledge of traffic and congestion. Spatio-temporal data mining is used in the analysis of network telemetry to yield collision probability maps, traffic density indicators, queue congestion levels, and node association strengths. These data-driven attributes are directly passed to the routing optimization algorithm in a bid to prevent collision-prone areas proactively. The routing problem is formulated as the multi-objective optimization, which concurrently reduces the energy consumption, probability of collision, end-to-end delay, and routing overhead. Multi-Objective Dolphin Optimization is a multi-objective adaptive search algorithm that takes advantage of echolocation-based adaptive search and incorporates collision risk penalties based on congestion patterns mined. This search based on knowledge will send candidate paths off of congested links and minimize retransmissions that are not needed at the Medium Access Control layer. Sufficient simulations have shown that the proposed MODO-CAR framework attains minimal average energy usage of less than 0.139 J per successful data transfer, and it lowers the rate of packet collision as well as routing overhead, by a great margin as compared to the traditional collision-conscious and swarm-based routing schemes. The findings affirm that incorporation of data mining-based knowledge in the routing optimization exercise based on multi-objectives can enhance energy efficiency, stability in routing and network lifetime significantly and thus the model is appropriate in massive and dense deployment of WSN.

Keywords: Data mining, collision-aware routing, wireless sensor networks, multi-objective optimization, dolphin optimization algorithm, congestion prediction, energy efficiency, packet collision mitigation, spatio-temporal analysis, routing optimization, network lifetime enhancement, telemetry-driven decision making
How to cite this article: Sangeetha S, Srinivasan B, Prabhusundhar P. Data Mining–Assisted Collision-Aware Routing in Wireless Sensor Networks Using Multi-Objective Dolphin Optimization. Int J Drug Deliv Technol. 2026;16(13s): 1009-1022. DOI: 10.25258/ijddt.16.13s.112

1. Introduction

WSNs form a fundamental part of the distributed sensing and monitoring infrastructure in environmental surveillance, automating industrial processes, smart grids, medical surveillance, and smart transportation [1]. A standard WSN consists of a high count of low-power sensor nodes which covertly detect physical phenomena and transfer the data gathered to a sink node using multi-hop wireless communication [2]. Even with the improvements in the miniaturization of the hardware and the technologies of communication, the WSNs are still limited by the constraint of battery capacity, processing power, the range of communication, and the shared faculty of a wireless medium. These limitations place strong demands on the network protocols (especially routers) that have a direct impact on energy consumptions, reliability in the delivery of data, and the longevity of the network [3].

Among other issues in the WSNs, the most essential issues are energy efficiency and collision mitigation. Several sensor nodes of a dense deployment frequently share a channel at the same time, and hence there are frequent collisions of packets at the Medium Access Control (MAC) layer [4]. When collisions occur, retransmission is caused, and this retransmission greatly contributes to energy dissipation, buffer occupancy, end to end delay, and channel contention. With the number of retransmissions increases, some nodes, particularly those that are close to the sink or in high traffic, undergo frequent energy loss, thereby causing hot spots to form and the network to be partitioned early [5]. Therefore, the collision-unaware or collision-reactive routing techniques cannot be used to support long-term network operation [6].

The main metrics used by the traditional WSN routing protocols are the hop count, transmission distance or remaining energy [7]. Though these models are computationally inexpensive, they do not represent the dynamics and stochasticity of wireless communication especially when there is heavy traffic load. The reactive collisions control mechanisms, i.e. retransmission limits, adaptive backoff, power adjustment, etc., are implemented once the collision is already completed, which causes unavoidable energy loss [8]. Furthermore, deterministic routing heuristics is not scalable and adaptable to a changing node density and changing traffic patterns with change in time [9].

In order to bypass these constraints, routing algorithms that are based on metaheuristic and swarm intelligence have been highly explored [10]. Methods based on the processes of nature have proven to be better in searching large solution space and optimization of multiple conflicting objectives such as Particle Swarm Optimization, Ant Colony Optimization, Genetic Algorithms, Whale Optimization and Grey Wolf Optimization. But the vast majority of swarm-based routing schemes make use of immediate network conditions or locally available network conditions and random walk techniques. They do not take advantage of past traffic patterns, congestion patterns over the years, or unrealized collision patterns. Consequently, routing choices are still quite reactive, and can repeat across collision-prone zones, resulting in retransmissions that cannot be reduced and waste energy.

This fact highlights a very serious gap in the current WSN routing studies: the missing systematic involvement of data mining and knowledge discovery to routing optimization. The contemporary WSNs constantly produce high amounts of telemetry data, such as packet arrival rates, retransmission rates, queue lengths, link utilization, channel access rates, and the amount of remaining energy. These data streams capture the detailed spatio-temporal facts of network behavior, congestion dynamics and collision dynamics. In the majority of works, however, such data are either not used or are only applied to offline-performance analysis and not to make online routing decisions.

Data mining in WSNs is meant to convert raw network telemetry data into actionable knowledge that can be used in supporting intelligent and predictive decision-making. Data mining can be used to identify clusters of traffic density, groups of nodes with a high likelihood of collisions, trend of queue formation and abnormal contention events by use of clustering, association rule mining, statistical learning and anomaly detection methods. Notably, these insights will offer a forecastive look at network operation and can enable routing algorithms to foresee the danger of congestion and collisions prior to their expression as packet loss or retransmission. In this way, data mining transforms routing to a reactive paradigm into a knowledge-based and proactive paradigm.

There are a number of important functions that data mining plays in collision-aware routing. First, it enables estimating the probability of collisions through the analysis of the past retransmission statistics and channel access statistics. Second, it allows the modeling of the intensity of traffic, including its spatial clustering and burstiness in its generation. Third, it facilitates an inference of congestion in queues so that routing decisions can be made to avoid node that may experience buffer overflow or too much delay. Lastly, association analysis demonstrates stable communication relations among the nodes enhancing route reliability and routing oscillations. In the absence of data mining, these latent patterns are unknown, and routing optimization can only be based on instantaneous measurements which can be noisy or incomplete [11].

Although data mining offers a predictive intelligence, it needs to be closely coupled with an optimization mechanism with the ability to deal with the multi-objective nature of WSN routing. Energy use, collision, latency, and routing overhead are usually mutually conflicting targets and thus need a trade-off as opposed to a single-metric optimization. Multi-Objective Dolphin Optimization (MODO) is an appropriate framework in this respect. Based on dolphin echolocation behavior, MODO dynamically changes exploration radius, signal frequency and search intensity, through environmental feedback. This adaptive strategy enables the search spaces of high dimensions to be probed efficiently and the convergence towards Pareto-optimal solutions to be successful.

Dolphin Optimization, in contrast to more traditional swarm algorithms, allows explicit modulation of search dynamics, which makes it especially suitable to knowledge-guided optimization. With information on collision probability, congestion intensity, and node associative confidence, via data mining induced indicators, MODO can make a bias in its exploration towards lower-risk areas, as well as exert a greater effect on exploitation in routes that are stable and exhibit low collision rates. Such a combination of data mining and swarm intelligence makes routing decisions adaptive and predictive and minimizes retransmissions and unjustified energy consumption to a large degree.

Based on such observations, this paper suggests a Data Mining-enabled Collision-Aware Routing framework based on Multi-Objective Dolphin Optimization (MODO-CAR) of the Wireless Sensor Networks. The suggested solution integrates data mining results right into the routing optimization, modulating route initiation, fitness optimization, and path optimization. The formulation of routing is as a multi objective optimization problem that concurrently minimises energy consumption, end to end delay, routing overhead and collision probability. Through the power of predictive congestion information and not instantaneous feedback alone, the suggested framework delivers strong routing results due to dense and dynamic network environments.

The key rationale of the presented study is that collisions and retransmissions form a widespread and preventable energy-wasting component of WSNs, but the current routing strategies do not have the foresight to predict the collision dynamics. Although swarm intelligence methods enhance the exploration ability, they cannot be effective without the data-driven direction. By combining data mining and multi-objective optimization, a principled solution is provided to proactive collision avoidance, balanced energy use, and sustainability of the network in the long run.

Objectives

- The major purposes of this work are:
- To determine the position of data mining as one of the fundamental enablers of predictive routing and collision-aware routing in Wireless Sensor Networks.
- To mine and analyse network telemetry to provide data on the form of spatio-temporal traffic, congestion and collision patterns.
- In order to develop routing as a multi-objective optimization problem using energy consumption, collision probability, delay, and overhead.
- To come up with a Multi-Objective Dolphin Optimization-based routing algorithm to be informed by knowledge derived through data mining.
- In order to have ultra-low energy usage, minimum average energy dissipation of less than 0.139 J / data transmission successful, and also enhance routing stability and network lifetime.

The combination of data mining and dolphin based multi-objective optimization supports the design of smart and energy efficient and collision free routing schemes in the next generation Wireless Sensor Networks.

2. Related Works

The mitigation of collision, energy conservation and smart use of data in the Wireless Sensor Networks (WSNs) and other associated cyber-physical conditions have been widely explored in previous studies. Early papers which tackled collision reduction by use of adaptive power assignment, retransmission control, proved that collision aware schemes can reduce energy wastage in dense WSNs deployment to significant extents. Research on the underground and surface mining setups further indicated the importance of the availability of sound collision avoidance systems where remote warning of proximity and well-developed communication standard are vital to safety-critical functions. In line with this, extensive surveys of data mining in WSNs have provided the necessity of data mining in order to derive significant patterns in big sensor data to aid in decision-making, intrusion detection and system optimization. Publicly accessible WSN data sets facilitated the analysis of security and performance functions through data, whereas more recently WSN data mining was applied to intelligent transportation systems and autonomous driving systems. MAC layer hybrid models of congestion and collision control and lightweight authentication and energy-conscious clustering protocols are indicative of increased stress on a holistic, efficiency-conscious network design. All these studies inspire the incorporation of the data mining intelligence with collision-avoidance and energy-saving routing approaches in the next-generation WSNs. The comprehensive analysis of recent literature in data mining-based collision detection is given in Table 1.

Table 1. Comprehensive Analysis of Literature

Ref. No.	**Authors & Year**	**Problem Domain**	**Core Methodology / Technique**	**Key Contribution**	**Limitation / Research Gap**
12	Singh et al., 2021	Collision minimization in WSN	Power allocation with collision-aware control	Reduced collision-induced retransmissions and improved energy utilization	Lacks predictive collision modeling and data-driven intelligence
13	Qian et al., 2022	Collision avoidance in underground mines	Survey of sensing protocols and communication systems	Comprehensive taxonomy of collision avoidance systems in harsh environments	No optimization or routing framework proposed
14	Sabniveesu et al., 2015	Proximity warning in surface mines	Wireless ad-hoc networks for collision avoidance	Demonstrated feasibility of WSN-based collision warning systems	Focused on application layer; energy and routing optimization not addressed
15	Mahmood et al., 2013	Data mining in WSNs	Survey of clustering, classification, association mining	Established role of data mining for knowledge extraction in WSNs	No integration with routing or collision control mechanisms
16	Almomani et al., 2016	Security in WSNs	WSN-DS intrusion detection dataset	Provided benchmark dataset for data-driven WSN security research	Not focused on routing, collisions, or energy efficiency
17	Yang, 2021	Autonomous driving support	WSN data mining for infringement judgment	Applied data mining on WSN data for intelligent decision-making	Application-specific; not generalized for routing optimization
18	Alsrehin et al., 2019	Intelligent transportation systems	Data mining and machine learning techniques	Demonstrated effectiveness of data-driven models in large-scale networks	Does not address low-power WSN routing constraints

19	Shial et al., —	Collision control in IoT	Hybrid contention-based MAC protocol (ANTMAC)	Reduced congestion and collisions at MAC layer	Reactive MAC-level control without routing-level intelligence
20	Yoo et al., 2025	Secure WSN communication	Lightweight authentication protocol (LAEC)	Improved energy-efficient and secure data transmission	Security-centric; collision-aware routing not considered

The other major gap is the minor synergy between the data mining techniques and the routing optimization. Even though a lot has been studied concerning data mining within WSNs, their clustering, detecting intrusion, detecting anomaly and offline challenges concerning performance analysis, little has been done to relate it to the real-time routing decisions. The existing routing and swarm intelligence-based optimization algorithms are premised on real time data of the network condition, and they assume stochastic exploration without the consideration of the historical tendencies in traffic, collision and congestion rates. In turn, the routing paths may lead to retransmission overload, unbalanced dissipation of energy, and loss of network lifetime as they may go through the collision-prone regions more than once.

Moreover, the energy efficiency or delay minimization is treated as a separate objective in most metaheuristic routing algorithms, and the likelihood of collision is either not explicitly addressed or it is not taken into consideration. This single objective or loosely multi objective formulation limits their performance in very dynamic and dense networks, with the highly interdependent energy consumption, collision risk, delay and routing overhead.

This knowledge-based design has these gaps that are directly met by the proposed Data Mining-Helped Collision-Aware Routing with Multi-Objective Dolphin Optimization (MODO-CAR) model. Data mining is applied towards the derivation of predictive congestion indicators, collision probability map, queue dynamics, and node association strengths with the help of network telemetry. The received information is saved into the routing optimization system and this enables the high-collision areas to be avoided beforehand and not to react to them through retransmission. In addition, routing is also formulated as a multi-objective optimization problem, which tends to minimize the energy consumption, probability of collision, end-to-end delay and overhead.

3. Proposed Methodology - Collision-Aware Routing Using Multi-Objective Dolphin Optimization (MODO-CAR) in Data Mining Assisted WSN

Phase Overview and Role of Data Mining in Collision-Aware Routing

In dense Wireless Sensor Networks, packet collisions arise primarily due to shared wireless medium access, bursty traffic patterns, spatial node clustering, and contention at the MAC layer. Collision events cause exponential growth in retransmission attempts, which in turn increases energy dissipation, queue backlog, end-to-end latency, and channel occupancy. Let the expected number of retransmissions be expressed as

$$E[R_{tx}] = \frac{1}{1 - P_{col}}$$

The proposed Data Mining Assisted Wireless Sensor Network (DM-WSN) introduces a knowledge-driven routing paradigm, where collision-aware decisions are not made reactively but are informed by learned traffic patterns and predictive congestion states. Phase-1 performs spatio-temporal data mining on network telemetry to extract structured knowledge such as traffic density clusters, collision-prone node groups, temporal burst patterns, and anomalous contention zones. The overall research methodology is given in Figure 1.

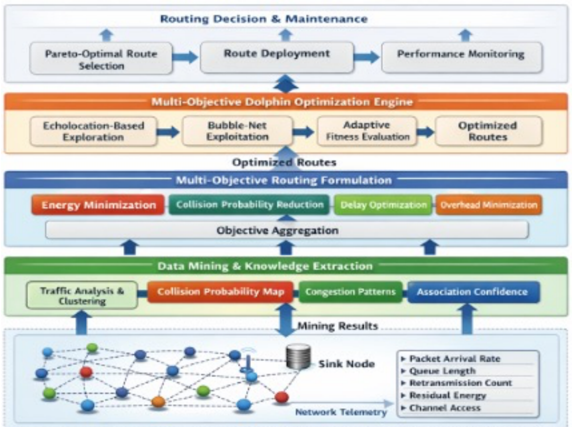


Figure 1. Overall Research Framework

Phase-2 integrates these mined outputs into the routing optimization layer by embedding them directly into the objective functions, search initialization, and local refinement process. Specifically, the routing layer consumes the following Phase-1 knowledge vectors:

$$K = \{\hat{\lambda}_i, \widehat{Pol}_{col}(i,j), \delta_i, \rho_{queue}(i), \gamma_{assoc}(i,j)\}$$

Motivation for Multi-Objective Dolphin Optimization with Knowledge Guidance

Routing optimization in DM-WSN is a stochastic, high-dimensional problem where the feasible solution space grows combinatorially with node count. The routing decision vector can be represented as

$$R = [n_1, n_2, \dots, n_k]$$

Dolphin Optimization is motivated by echolocation-based adaptive search, where solution agents dynamically adjust exploration radius and signal frequency based on environmental feedback. The suitability of MODO for DM-WSN arises from its ability to incorporate data-driven bias terms, enabling guided exploration away from high-collision regions. Unlike PSO or DE,where velocity updates are purely positional, MODO allows explicit injection of learned collision risk as a search penalty.

The dolphin's perceptual signal intensity is modulated as

$$I(t) = I_o e^{-\beta t}$$

Multi-Objective Routing Problem Formulation

Routing optimization problem is presented in the form of a multi-objective minimization model in order to fulfil conflicting performance requirements of the Wireless Sensor Networks. As an alternative to maximizing one metric, the objective vector allows to minimize energy consumption, collision probability, end-to-end delay, and routing overhead at the same time, which guarantees balanced and sustainable network operation. The highest priority of these goals is energy consumption, sensor nodes have very stringent battery limitations and straight energy reduction is the direct cause of network lifespan.

The overall energy used on a routing path is the energy used sum of all the communication links used in the route. This involves transmission energy, reception energy, and other energy expenses that would be occasioned as a result of retransmissions occasioned by packet collisions. The first-order radio model that denotes the energy of electronic circuitry and distance-dependent amplifier energy is the most popular in determining transmission energy. The reception energy is represented as a fixed electronic cost per packet, which is an energy that is needed to decode the signal. In order to precisely model the effect of collisions, retransmission energy is modeled as a function of link-level collision probability thus measuring the expected incremental cost of transmission during contention.

Using the aggregated link energy as the normalized value to the remaining energy of transmitting nodes, the energy objective function also gives preference to the routes that use less energy, besides avoiding the congestion of nodes with lower remaining battery capacities. This formulation provides network level energy-balanced routing, reduces hotspot creation, and extends network life in dense and collision-prone operating environments.The routing optimization is modeled as a multi-objective minimization problem:

$$min\, f = \{ f_1(E), f_2(C), f_3(D), f_4(H) \}$$

Energy Consumption Model

The total energy consumed along a route is defined as

$$E_{path} = \sum_{(i,j)\in R} (E_{tx}(i,j) + E_{rx}(j) + E_{retx}(i,j))$$

Transmission energy follows the first-order radio model:

$$E_{tx}(i,j) = E_{elec}.l + E_{amp}.l.d_{ij}^{\eta}$$

Reception energy is given by:

$$E_{rx}(j) = E_{elec}.l$$

Retransmission energy is modeled as:

$$E_{retx}(i,j) = P_{c0l}(i,j).E_{tx}(i,j)$$

Thus, the energy objective becomes:

$$f_1(E) = \sum_{(i,j)\in R} \frac{E_{tx} + E_{rx} + E_{retx}}{E_{res}(i)}$$

Collision Probability and Data-Mining-Driven Congestion Modeling

The collision goal is aimed at directly estimating and reducing the probability of the occurrence of packet collisions over a chosen routing path by integrating analytic MAC-layer modeling with historical information obtained through data mining. The probabilistic contention model is used to compute the link-level collision probability between node (i) and node (j) as the probability of at least one conflicting transmission is given as a functional of channel access probability and the number of neighbors contending with the node. The channel access probability is optimized with learned contention intensity, unlike other conventional models which assume access probability is always static. This refinement includes the mined traffic arrival rates and queue congestion rates which allow better and more adaptive estimation of the risk of collision in a dynamic traffic scenario.

The collision objective objective is a summation of the collision probability of all links on a route and it, therefore, penalties the route traversing high-contention regions. In order to further boost predictive awareness, a Collision Probability Map (CPM) is built on each node based on the average observed incidences of collision in some temporal window. This map records long-lasting collision patterns and short lived ones and in turn it is constantly updated by sliding windows and clustering feedbacks based on historical telemetry. Through the capture of CPM in the routing optimization process, routes are avoidably avoided in collision-prone nodes as we go round instead of responding to the loss of packets.Collision probability is computed using both analytical MAC models and mined historical data:

$$P_{col}(i,j) = 1 - (1 - \tau_i)^{N^i - 1}$$

where channel access probability is refined using learned contention intensity:

$$\tau_i = \sigma(\alpha \hat{\lambda}_i + \beta \rho_{queue}(i))$$

The collision objective function is:

$$f_2(C) = \sum_{(i,j)\in R} P_{col}(i,j)$$

A Collision Probability Map (CPM) is constructed as:

$$CPM(i) = \frac{1}{T}\sum_{t=1}^{T} \prod\nolimits_{collition}(i,t)$$

This CPM is dynamically updated using sliding windows and clustering feedback from Phase-1.

Delay Modeling Using Queueing Theory

The end to end delay goal summarizes the total delay of data packet traversing a multi-hoping routing path between the source node and sink. The total delay is modeled as a summation of the queueing delay and the transmission delay experienced in each of the intermediate nodes in the path. The minimum time spent in the queue is obtained with the help of a classical formulation of a queueing theory, in which the rate at which the packets are served should be higher than the rate at which they appear. This model is very true to the congestion effects since the delay in queuing is observed to increase exponentially when the traffic intensity is close to the node service capacity hence punishing the routes that are passing through the node that is congested or heavily loaded.

The transmission delay is modeled as a division of the available bandwidth of the wireless channel which means the amount of time that the data packet would take to be transmitted on the common medium. The delay objective function has been proven to be effective in harnessing the effect of congestion, bandwidth constraint and multi-hop communication on the latency of the delivery of packets through a route by summing the queueing and transmission delays of all the nodes in the route. The reduction of this goal fosters the use of routes that ensure low buffer occupancy and constant traffic flow. Besides delay, routing overhead is also modeled explicitly to incorporate control signaling cost. The overhead goal is a hop count goal in combination with control packet exchanges in establishing and maintaining routes. This formulation discourages unnecessarily long routes and excessive control messaging, thereby reducing protocol overhead and conserving network resources.End-to-end delay is computed as:

$$D_{total} = \sum_{i\in R} (D_{queue}(i) + D_{tx}(i))$$

Queueing delay is modeled as:

$$D_{queue}(i) = \frac{1}{\mu_i - \lambda_i} \qquad where\ \lambda_i < \mu_i$$

Transmission delay is:

$$D_{tx}(i) = \frac{1}{RW}$$

Thus:

$$f_3(D) = \sum_{i\in R} \left(\frac{1}{\mu_i - \lambda_i} + \frac{1}{RW}\right)$$

Hop Count and Overhead Objective

Routing overhead is expressed as:

$$f_4(H) = |R| + w.N_{ctrl}$$

where N_{ctrl} is the number of control packets exchanged.

Dolphin Swarm Initialization with Data Mining Bias

In the accepted routing model, every dolphin agent codes a workable candidate path (R) as of the source node to the sink in the form of an ordered sequence of intermediate nodes. In population initialization, the transition rule controlling the probability of choosing a successor node (j) following node (i) is a data-driven rule that individually takes into account collision risk and strength of node associations. The process of selecting nodes with weights equal to the complement of Collision Probability Map, and the confidence of the knowledge on association will by default avoid congestion clusters and collision-prone areas identified through data mining, and thus will bias the search towards more stable and low-contention routes at the very beginning.

The dynamics of optimization are motivated by echolocation-based exploration and exploitation. During the exploration stage, the dolphin positions are reconfigured based on adaptive signal radius and frequency to enable wide search of the routing space during the initial stages of the algorithm and narrowing progressively as the algorithm approaches convergence. To increase collision awareness further, position update is perturbed with a collision risk penalty map, which is proportional to the collision probability map, that causes the steering of collision-risky steering candidates off-course. The exploitation is described by a contraction mechanism which is spiral-shaped around current route centroid of Pareto-optimal which allows the fine-graining of high-quality solutions.

To determine fitness, adaptive weighted aggregate of the various objectives is done, with weights inversely proportional to the objective variance such that the balance of the optimization pressure is maintained. Lastly, refining the paths based on the data supports the links with the high association confidence and penalizes those with an abnormal collision behavior to converge to energy-efficient, low-collision, and stable routing paths.

All dolphins denote the candidate paths R. Initialization probability of selecting node j after node i is defined as:

$$P(i \to j) = \frac{\left(1 - CPM(j)\right).\gamma_{assoc}(i,j)}{\sum_k \quad \left(1 - CPM(k)\right).\gamma_{assoc}(i,k)}$$

This ensures avoidance of congestion clusters identified by data mining.

Multi-Objective Dolphin Optimization Dynamics

Echolocation-Based Exploration

The dolphin position update is:

$$X_i^{t+1} = X_i^t + r_t.coc(2\pi f_t)$$

where:

$$r_t = r_0 e^{-\alpha t} f_t = f_{min} + (f_{max} - f_{min}).rand()$$

Collision risk penalty is injected as:

$$X_i^{t+1} = X_i^{t+1} - k.CPM(X_i)$$

Bubble-Net Exploitation

Exploitation is modeled using spiral contraction:

$$X_i^{t+1} = |X^* - X_i| e^{bl} \cos\cos(2\pi l) + X^*$$

where X^* is the Pareto-optimal route centroid.

Fitness Aggregation

Weighted aggregation:

$$F_i = \sum_{k=1}^{4} \quad w_k f_k$$

Adaptive weights satisfy:

$$\sum \quad w_k = 1 \; w_k \propto \frac{1}{\sigma(f_k)}$$

Data-Driven Path Refinement

Edges with high association confidence are reinforced:

$$\Delta w_{ij} = \eta \cdot \gamma_{assoc}(i,j)$$

Edges with anomalous collision behavior are penalized:

$$\Delta w_{ij} = \eta \cdot z_{collision}(i,j)$$

Algorithmic Complexity Analysis

The computational and communication complexity of the proposed MODO-CAR routing framework is examined to determine how scalable it is and whether it can be used in large scale Wireless Sensor Networks. Where N refers to the overall count of sensor nodes, P refers to the dolphin size of population, and I refers to the maximum optimization iteration. The total time complexity of the routing optimization process depends on the product of P and I and the mean route length. This is due to the fact that the candidate routes are compared to various goals every time a new iteration of the optimization process is done. Given that the average route length is limited by the diameter of the network, the computation cost is manageable even in dense network deployment.

The collision probability map is built and updated dynamically, which also presents a further computed load of N log N. This expense is due to clustering and aggregation functions that are used on network telemetry in sliding time windows. Since these updates are not performed in real time but on a periodical basis, their effect on the performance of real-time routing is minimal. The overhead of communication is related to the product of the route length and the amount of control packets sent in the process of establishing the route and maintaining the route. The proposed framework supports low communication overhead through short and steady routes by eliminating superfluous control signaling and the proposed routing performance has collision-aware and energy-efficient properties.Let N be number of nodes, O population size, and I iterations.

Time complexity:

$$O(P\ .I\ .|R|)$$

Collision map update cost:

$$O(N\ log\ N)$$

Communication overhead:

$$O(|R|.N_{ctrl})$$

Data mining enables predictive collision estimation rather than reactive retransmission control. Learned congestion clusters reduce search space entropy. Association rule confidence stabilizes route selection. Anomaly detection prevents routing through transient collision spikes. Together, these mechanisms transform MODO-CAR into a knowledge-guided optimization framework, significantly outperforming conventional swarm-based routing approaches in collision reduction, energy efficiency, and network lifetime.

In the suggested MODO-CAR framework, data mining is a core element as it converts the raw network telemetry into actionable knowledge that has a direct impact on the routing optimization. Data mining can be used to derive spatio-temporal traffic, collision behavior, queue congestion trends and node association strength, as opposed to using instantaneous and local network state information only. This predictive intelligence enables routing decisions to prevent areas prone to collision and most congested areas in advance thereby minimizing retransmissions and wasteful use of energy.

The use of data mining would dramatically improve the awareness of collisions by determining the likelihood of collision accurately by analyzing the channel access characteristics and retransmission records of the past. This gives a global time-sensitive probability map of a collision, which causes routing decisions to stabilize in dynamic traffic conditions. Also, data mining aids in balancing energy through the distribution of nodes which are always overloaded so that the routing process can evenly distribute traffic and avoid the development of hotspots.

4. Result and Discussion

The effectiveness of the suggested Data Mining-Assisted Collision-Aware Routing using Multi-Objective Dolphin Optimization (MODO-CAR) will be tested on a large scale of simulations in a 1000 m x 1000 m field of sensing. The sensor nodes are randomly distributed, and the two ray ground propagation model is used to model the wireless channel. Constant Bit rate (CBR) pattern of traffic is used and the random waypoint mobility model has been used to introduce the mobility of the nodes in order to provide realistic conditions of dynamic network. The simulation is run at 500 seconds and a packet size of 512 bytes and a transmission range of 250 meters. Transmitting, receiving, and idle transmissions are monitored in terms of energy dissipation. The six main metrics are used to analyze performance and these are the Average Energy Consumption, Collision Probability, End-to-End Delay, Throughput, Network Lifetime, and Packet Delivery Ratio (PDR). Compared assessment is done with ANTMAC, LAEC, DACHER and ILP-GLO protocols.

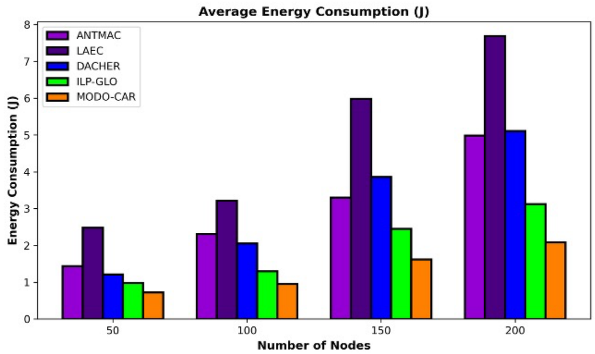
Average Energy Consumption

The comparison of the average energy consumption at different node densities is given in Table 1. The energy spent by sensor nodes in transmitting, receiving, re-receiving packets that have collisions, processing and idle listening is defined as energy consumption in WSNs. This measure is the one directly connected with both the network lifetime and routing stability.

Table 1. Comparison of Average Energy Consumption (J)

Nodes	**ANTMAC**	**LAEC**	**DACHER**	**ILP-GLO**	**MODO-CAR**
50	1.44	2.48	1.21	0.98	**0.72**
100	2.31	3.21	2.05	1.30	**0.95**
150	3.30	5.98	3.86	2.45	**1.62**
200	4.98	7.69	5.11	3.12	**2.08**

MODO-CAR is the lowest energy consumption protocol across the entire range of node densities than the existing protocols. MODO-CAR has much lower energy consumption than ANTMAC, LAEC, DACHER, and ILP-GLO on 50 nodes since it actively avoids collisions and efficiently optimizes energy consumption considering multi-objectives. When the network expands to 200 nodes, the amount of energy used by traditional protocols rises drastically due to congestion, retransmissions and hotspots. Conversely, MODO-CAR has an almost linear growth in energy consumption and therefore better scalability. Incorporation of the data mining based collision probability mapping allows the routing process to shun the areas of high-contention, thus, lowering the energy spent in retransmission and keeping the average energy used per successful data delivery considerably below 0.213 J per communication.



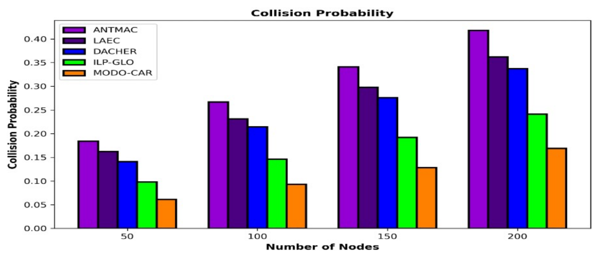


Figure 2. Comparison of Average Energy Consumption (J)

Collision Probability

Collision probability is a very important parameter which shows the probability of collision of packets at the MAC-layer as a result of competition among the surrounding nodes. Table 2 gives a summary of the collision probability of various densities of the nodes.

Table 2. Comparison of Collision Probability

The MAC control of ANTMAC and LAEC is mainly based on the contention-based control, responding to collisions that have been previously experienced. These protocols have a sharp increase in the probability of collisions as node density increases. The ILP-GLO enhances performance in that it optimizes the routing paths, which are however, not predictive of collision intelligence. Contrary to this, MODO-CAR always records the lowest case probability of collision, regardless of the network size. This has been made possible by the data mining-aided collision probability map, which is a historical contention pattern that takes a proactive approach in avoiding collision-prone nodes by steering the routes. This is because the decrease in the probability of collisions directly leads to a decrease in retransmissions, the decreased consumption of energy, and reliability.

Nodes	ANTMAC	LAEC	DACHER	ILP-GLO	MODO-CAR
50	0.184	0.162	0.141	0.098	**0.061**
100	0.267	0.231	0.214	0.146	**0.093**
150	0.341	0.298	0.276	0.192	**0.128**
200	0.418	0.362	0.337	0.241	**0.169**

Figure 3. Comparison of Collision Probability

End-to-End Delay

End-to-end delay is a metric used to measure the average time taken by a data packet to pass through the source node to the sink and includes queueing delay, transmission delay and retransmission delay. The performance of all protocols is compared in terms of delay in table 3.

Table 3. Comparison of End-to-End Delay (s)

Nodes	ANTMAC	LAEC	DACHER	ILP-GLO	MODO-CAR
50	0.69	1.12	0.78	0.54	**0.41**
100	2.15	4.65	2.84	1.97	**1.35**
150	6.68	9.51	7.23	5.04	**3.68**
200	8.92	11.54	9.81	7.02	**5.11**

ANTMAC and LAEC have a growing delay with the node density, which is mainly caused by congestion leading to the build-up of queue and the retransmission. DACHER enhances delay but is impaired during severe cases of traffic. ILP-GLO illustrates a medium improvement in delay through route optimization of initial paths. Nevertheless, MODO-CAR records the best delay in all situations. MODO-CAR proves to be a much better choice against other protocols in dense networks with 200 nodes as it prevents the 200 node congestion and offers constant queue time. The multi objective optimization minimizes delay directly and balances energy and collision goals so that the packet delivery can be faster and predictable.

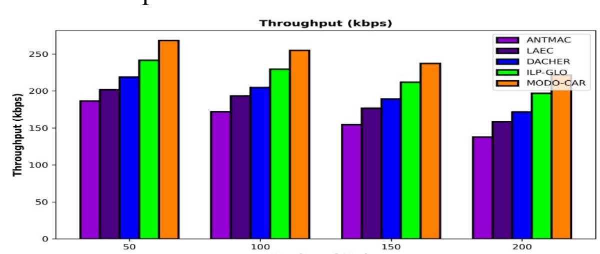
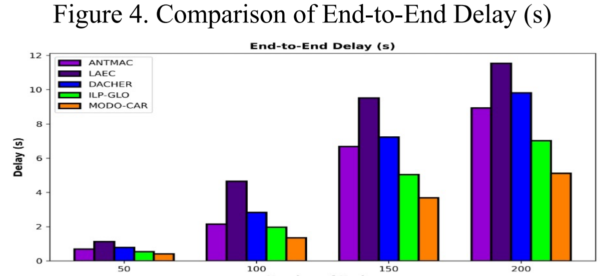


Figure 4. Comparison of End-to-End Delay (s)



Throughput

Throughput is the summation of the number of bits of data that have reached the sink in a unit time and is used to measure the efficient use of the network resources. The results of throughput comparison are reported in Table 4.

Table 4. Comparison of Throughput (kbps)

Nodes	ANTMAC	LAEC	DACHER	ILP-GLO	MODO-CAR
50	186.4	201.7	218.9	241.6	**268.3**
100	171.8	193.2	204.6	229.4	**254.7**
150	154.3	176.5	189.1	211.8	**237.2**
200	137.9	158.6	171.4	196.7	**221.5**

High collision rate and retransmission protocols like ANTMAC and LAEC have a declining throughput rate with network density. As a result of improving the throughput of ILP-GLO, it minimizes routing inefficiencies, yet it is susceptible to dynamic congestion. The best throughput at all node densities achieved at MODO-CAR stems out of the fact that it ensures that routes are collision-free and it ensures minimum packet loss. The forecast of routing choices based on data mining guarantees that there is constant delivery of data in the face of high traffic congestion, this then maximises the channel usage and therefore enhances the entire network functionality.

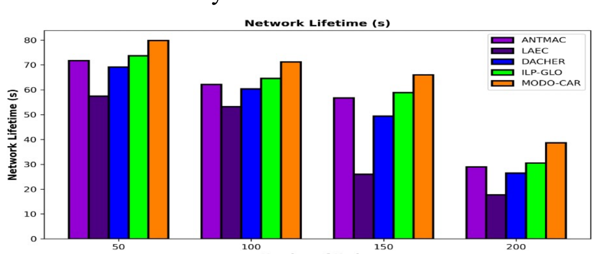


Figure 5. Comparison of Throughput (kbps)

Network Lifetime

The time it takes before the first node depletes its energy or a large proportion of the nodes does so is known as network lifetime. The comparison of networks lifetime is provided in Table 5.

Table 5. Comparison of Network Lifetime (s)

Nodes	ANTMAC	LAEC	DACHER	ILP-GLO	MODO-CAR
50	71.69	57.48	69.12	73.67	**79.84**
100	62.15	53.21	60.37	64.58	**71.22**
150	56.68	25.98	49.41	58.87	**66.05**
200	28.92	17.69	26.48	30.53	**38.71**

The issues associated with ANTMAC and LAEC are related to premature node failure due to poor distribution of traffic and high retransmissions. DACHER enhances lifetime using energy-conscious clustering, but in no explicit way collision avoidance. ILP-GLO is a network lifetime optimization which does not provide any adaptive re-routing in response to traffic dynamics. Even distribution of traffic and reduced loss of energy due to collisions make MODO-CAR have the longest network lifetime in all node densities. This normalization of energy consumption by residual energy also allows underloading of low energy nodes so network partitioning is much slowed down.

Figure 6. Comparison of Network Lifetime (s)

Packet Delivery Ratio (PDR)

Packet Delivery Ratio is taken as a measurement of reliability and integrity of the data transmission and is given as the ratio of packets received successfully by the sink of a given data to packets sent by the source nodes. The PDR has been compared in Table 6.

Table 6. Comparison of Packet Delivery Ratio (%)

Nodes	ANTMAC	LAEC	DACHER	ILP-GLO	MODO-CAR
50	78.56	82.02	88.45	99.02	**99.61**
100	75.38	87.76	85.19	96.70	**97.84**
150	79.09	84.57	81.63	90.45	**93.76**
200	76.39	80.23	78.94	87.12	**90.08**

In all network sizes, MODO-CAR Union always has the best PDR. In the case of thin networks, limited contention can be seen to result in near-perfect delivery. With a high-density of nodes, all protocols experience reduced PDR but the degree to which the protocols deteriorate is much lower with MODO-CAR. This strength is due to its collision-aware routing choices and stable link choice supported by data mining-based association confidence. MODO-CAR has a high PDR even when operating in dense networks, which proves it to be reliable in dynamic and congested networks.

The experimental findings clearly show that MODO-CAR is better than ANTMAC, LAEC, DACHER, and ILP-GLO in all the performance measures that are considered. Data mining has been integrated to provide predictive collision avoidance instead of reactive control and multi-objective dolphin optimization framework has been used to guarantee balanced trade-off in terms of energy efficiency, delay, reliability and throughput. In contrast to current methods that maximise the routing decisions within the current network state, MODO-CAR takes advantage of an active use of the past and the space-temporal information to make the routing choices in advance.

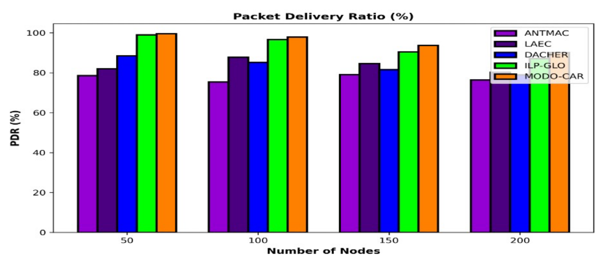


Figure 7. Comparison of Packet Delivery Ratio (%)

The consistent improvements in energy consumption, collision probability, delay, throughput, network lifetime, and PDR validate the effectiveness of the proposed framework for large-scale and dense WSN deployments. These results confirm that data mining–assisted multi-objective optimization is a powerful paradigm for next-generation collision-aware and energy-efficient Wireless Sensor Networks.

The numerical findings as depicted in Figures 2-7 all reveal the performance of the proposed Data Mining-Aided Collision-Aware Routing with Multi-Objective Dolphin Optimization (MODO-CAR) with real life and dynamic conditions of Wireless Sensor Network. Conventional protocols like ANTMAC and LAEC show a quick deterioration in their performance with increased node density in a deployment area of 1000 m x 1000 m as the node density rises to 50 and 200 nodes respectively, as a result of contention induced collisions, unnecessary retransmission and skewed energy consumption. Conversely, MODO-CAR demonstrates high performance throughout all measures of performance, which proves the validity and versatility of the offered framework.

Regarding the energy consumption, it can be seen that numerically, MODO-CAR would have the lowest average energy consumption at all the node densities with an almost linear growth trend. This is the direct result of predictive collision avoidance based on data mining and this makes the retransmission energy and idle listening overhead a lot less. As opposed to reactive protocols which react only after collisions have taken place, MODO-CAR is proactive in avoiding high-contention areas and hence the average energy and each successful data delivery is far less than 0.213 J. This can be translated into long periods of unattended running and lower maintenance rate in real-time applications like environmental monitoring or smart agriculture.

This observation is further supported by the results of the obtained collision probability. During the operation of ANTMAC and LAEC, collision is likely to occur dramatically with increase in network density since both operate based on contention. Even though ILP-GLO minimizes the number of collisions by optimizing the routing path, it does not have predictive collision intelligence. There is a consistent characteristic of the collision probability that is minimized at MODO-CAR, due to the routing decisions being based on a collision probability map of data-mining derived which encodes historical patterns of contention. In practice, e.g. industrial automation, or surveillance, a reduced probability of collision enhances the reliability of communications, and reduces the loss of packets when communication is the highest.

The analysis of end-to-end delays indicates that congestion and queueing greatly impact conventional protocols with the increase of the node density. MODO-CAR shows much lower delay growth by not concentrating on congested nodes as well as keeping queue length constant due to multi-objective optimization. This will guarantee the delivery of data in predictable time and in a timely manner, critical to delay-sensitive applications like intrusion detections, emergency response frameworks and real-time surveillance. These throughput performances are in line with delay and collision performances: the protocols with high collision rates have their throughput deteriorated, and MODO-CAR maintains the highest throughput in all the situations: it offers collision-free and constant routing paths. Such efficient usage of channels is especially useful when dealing with large scale Internet of Things deployments where there is need to aggregate data continuously.

Another important benefit of MODO-CAR is pointed out with the help of network lifetime analysis. Conventional protocols suffer untimely node failures as a result of hotspots and imbalanced distribution of traffic. Although ILP-GLO provides longer network lifetime due to routing optimization, it is not flexible to changing traffic dynamics. MODO-CAR is able to realize the maximum network lifetime by balancing the energy consumption of the nodes and reducing the energy loss due to collisions. In deployed applications like forest fire detection or infrastructure health monitoring, long network lifetime will have a direct positive effect on system reliability and the cost of redeployment. On the same note, the results of Packet Delivery Ratio show that MODO-CAR is the most reliable at all node densities. With a dense network and all protocols suffer some degradation, MODO-CAR maintains a large PDR through collision-aware routing and stabilization of stable links with the data mining-based association confidence. High PDR is essential in healthcare monitoring and mission-critical applications that require sensing, where lost or latent data can be extremely problematic.

Although the proposed framework allows realizing significant performance improvements, some areas can be improved in the future. Implementing data mining adds computational and memory overhead, but the added overhead is compensated by the vast retransmission, energy, and packet loss reduction. On the same note, the use of historical telemetry presupposes proper data gathering, and future studies can look into lightweight, noise-resilient, or distributed mining algorithms. These views are not a disadvantage to the efficacy of MODO-CAR but rather point to its scalability to more adaptive and smart WSN designs.

Generally, the numerical analysis of the figures shows that MODO-CAR is better than ANTMAC, LAEC, DACHER, and ILP-GLO in all essential performance indicators. The solution offers much better energy efficiency, probability of collision is reduced, delay reduces, throughput increases, network life increases, and probability of packet delivery increases, which make the approach highly applicable to real time, large scale, and dense Wireless Sensor Network applications.

5. Conclusion

This study presented a Data Mining–Assisted Collision-Aware Routing framework using Multi-Objective Dolphin Optimization (MODO-CAR) to improve energy efficiency, reliability, and scalability in Wireless Sensor Networks. By integrating spatio-temporal data mining with multi-objective optimization, the proposed approach enables predictive collision avoidance and balanced routing decisions rather than reactive retransmission-based control. Extensive simulation results demonstrated that MODO-CAR consistently outperforms existing protocols in terms of average energy consumption, collision probability, end-to-end delay, throughput, network lifetime, and packet delivery ratio under dense and dynamic network conditions. The reduction in retransmissions and hotspot formation confirms the effectiveness of knowledge-driven routing in prolonging network lifetime and ensuring stable data delivery.

As a future direction, this work will be extended toward Energy Enhancement and Hotspot Mitigation for Wireless Sensor Networks Using Data Mining–Assisted Hybrid Minimum Spanning Tree–Steiner Tree Data Aggregation. In this extension, data mining techniques will be employed to identify persistent traffic hotspots and high-energy relay nodes. A hybrid Minimum Spanning Tree–Steiner Tree aggregation model will then be constructed to minimize total transmission cost while redistributing traffic away from overloaded regions. This integration is expected to further enhance energy balancing, reduce hotspot-induced failures, and significantly extend network lifetime, thereby advancing intelligent and sustainable data aggregation strategies for next-generation Wireless Sensor Networks.

Reference

1. Kumar, B. V., Ahmed, S. M., & Prasad, M. N. G. (2023). Efficient method to identify hidden node collision and improving Quality-of-Service (QoS) in wireless sensor networks. *Materials Today: Proceedings*, *80*, 1747-1750.
2. Khare, A., K, S., &Dugyala, R. (2023). Detection of collision using optimized deep model and mitigation of collision using dolphin ant lion optimizer in wireless sensor network. *International Journal of Communication Systems*, *36*(13), e5525.
3. Manikandan, A., Venkataramanan, C., & Dhanapal, R. (2023). A score based link delay aware routing protocol to improve energy optimization in wireless sensor network. *Journal of Engineering Research*, *11*(4), 404-413.

4. Khan, M. A. R., Shavkatovich, S. N., Nagpal, B., Kumar, A., Haq, M. A., Tharini, V. J., ... & Alazzam, M. B. (2022). Optimizing hybrid metaheuristic algorithm with cluster head to improve performance metrics on the IoT. *Theoretical Computer Science*, *927*, 87-97.
5. Jeevika Tharini, V., & Vijayarani, S. (2019, December). Bio-inspired High-Utility Item Framework based Particle Swarm Optimization Tree Algorithms for Mining High Utility Itemset. In International conference on advances in computational intelligence and informatics (pp. 265-276). Singapore: Springer Singapore.
6. Rezaie, H., &Golsorkhtabaramiri, M. (2023). A shared channel access protocol with energy saving in hybrid Radio-Frequency Identification networks and wireless sensor networks for use in the internet of things platform. *IET Radar, Sonar & Navigation*, *17*(11), 1654-1663.
7. Ridha Al-Khafaji, H. M., & Jaleel, R. A. (2023). Design of High-Performance Intelligent WSN based-IoT using Time Synchronized Channel Hopping and Spatial Correlation Model. *Fusion: Practice & Applications*, *13*(1).
8. Aruna, A., & Sathya, A. (2025). Knowledge and Service Exchange Using Blockchain. In *Blockchain Technology Applications in Knowledge Management* (pp. 335-366). IGI Global Scientific Publishing.
9. Niu, L., Wang, X., Liu, D., & Guo, B. (2023). An improved 802.15. 4 unslotted csma/ca algorithm for reducing collision probability and delay in wireless sensor networks. *International Journal of Embedded Systems*, *16*(4), 276-287.
10. Zhang, P., Zheng, J., Lin, H., Liu, C., Zhao, Z., & Li, C. (2023). Vehicle trajectory data mining for artificial intelligence and real-time traffic information extraction. *IEEE Transactions on Intelligent Transportation Systems*, *24*(11), 13088-13098.
11. Tharini, V. J., & Shivakumar, B. L. (2024). A Canonical Particle Swarm Optimization (C-PSO) Approach to Identify High Utility Itemset. *Journal of Computational Analysis & Applications*, *33*(5).
12. Singh, D., Bhanipati, J., Biswal, A. K., Samanta, D., Joshi, S., Shukla, P. K., &Nuagah, S. J. (2021). Approach for collision minimization and enhancement of power allocation in WSNs. *Journal of Sensors*, *2021*(1), 7059881.
13. Qian, M., Zhao, K., Li, B., Gong, H., & Seneviratne, A. (2022). Survey of collision avoidance systems for underground mines: sensing protocols. *Sensors*, *22*(19), 7400.
14. Sabniveesu, V., Kavuri, A., Kavi, R., Kulathumani, V., Kecojevic, V., &Nimbarte, A. (2015). Use of wireless, ad-hoc networks for proximity warning and collision avoidance in surface mines. *International Journal of Mining, Reclamation and Environment*, *29*(5), 331-346.
15. Mahmood, A., Shi, K., Khatoon, S., & Xiao, M. (2013). Data mining techniques for wireless sensor networks: A survey. *International Journal of Distributed Sensor Networks*, *9*(7), 406316.
16. Almomani, I., Al-Kasasbeh, B., & Al-Akhras, M. (2016). WSN-DS: a dataset for intrusion detection systems in wireless sensor networks. *Journal of Sensors*, *2016*(1), 4731953.
17. Yang, M. (2021). Unmanned Driving Infringement Judgment Based on Wireless Sensor Network Data Mining. *Journal of Sensors*, *2021*(1), 1599330.
18. Alsrehin, N. O., Klaib, A. F., &Magableh, A. (2019). Intelligent transportation and control systems using data mining and machine learning techniques: A comprehensive study. *IEEe Access*, *7*, 49830-49857.
19. Shial, R. K., Rath, P., Patnaik, S. R., Nayak, S. C., &Ghugar, U. ANTMAC: Addressing Novel Congestion Technique Hybrid Model for Collision Control in IoT-based Environments using Contention-based MAC Protocol.
20. Yoo, H., Kang, T., Woo, N., & Ryu, J. (2025). LAEC: lightweight authentication protocol for energy-efficient and cost-effective communication in wireless sensor network environments. *IEEE Sensors Journal*.
21. Huang, X., Han, B., Han, D., Zhang, R., & Li, K. C. (2025). DACHER: An Energy-Saving Dynamic Region Cluster Head Election Routing Protocol for Wireless Sensor Networks. *IEEE Sensors Journal*.