INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

Interrogation of Dynamic Data Loss in Long Range Wireless Sensor Networks by Utilizing CatBoost-MLGBA to Detect Anomalies and Unusual Patterns

R Kowsalya[1]*, C V Banupriya[2]

1 Assistant Professor and Head, Department of Computer Science (PG), PSGR Krishnammal College for Women, Coimbatore, Tamil Nadu, India
2 Assistant Professor, Department of Computer Science, PSG College of Arts & Science, Coimbatore, Tamil Nadu, India

OPEN ACCESS

Received: 02-10-2023
Accepted: 16-10-2023
Published: 30-12-2023

Citation: Kowsalya R, Banupriya CV (2023) Interrogation of Dynamic Data Loss in Long Range Wireless Sensor Networks by Utilizing CatBoost-MLGBA to Detect Anomalies and Unusual Patterns. Indian Journal of Science and Technology 16(47): 4547-4560. https://doi.org/10.17485/IJST/v16i47.2496

* **Corresponding author**.

rkowsalya31psg@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Objectives: To propose a novel AI-based quantum key distribution optimization model to detect abnormal sensor readings, communication pattern between nodes, and intrusions during data transformation in long-range wireless sensor networks (LoRA-WSNs). In order to optimize the QKD in WSNs, machine learning boosting techniques are employed to minimize data loss and maximize data integrity. **Methods:** The CatBoost machine learning-based gradient boosting algorithm (CatBoost-MLGBA) is employed for QKD optimization and to detect abnormal node communications and patterns during data transfer by training historical network data. The linear regression method (LRM) with key generation rates is used to predict network attacks, which helps optimize the QKD more effectively. Lasso Regularization (L1R) is utilized to spot and recover the data in networks, and Deep Q-Networks (DQN-WSN) combined with the shortest path method is used to find alternate routing for the finest node search and data transfer. The WSN-DS historical dataset is utilized to train the CatBoost-MLGBA model to detect anomalies effectively. The OMNET++ tool is used to assess the performance of the proposed CatBoost-MLGBA model by comparing it with prevailing protocols such as ReLeC-WSN, RTM-ANN, and DL-IDSWSN. **Findings:** The new AI based optimization model, CatBoost-MLGBA outperforms the existing protocols in preventing data loss by enhancing security features. The proven results show that the data loss is minimized to 10%, with a 9% energy consumption rate, 95% network lifetime, 97% PDR rate, 91% robustness to anomaly attacks, and 6 seconds data transmission speed rate. **Novelty:** The CatBoost-MLGBA model has the ability to enhance security features and prevent data loss during data transfer in LoRa-WSNs. The new method effectively optimizes the key distribution for secured data transmission and improves the packet delivery ratio. The challenges of the prevailing security protocols, such as ReLeC-WSN, RTM-ANN, and DL-IDSWSN, are addressed.

Keywords: WSN Security; CatBoost Model; Advanced Networking; Energy Efficiency; Key Optimization Model; Gradient Boost Method

1 Introduction

In the development of 5G technology, the deployment of sensor nodes in WSNs has a significant impact on how data is sent from one end to the other. An efficient WSN protocol is required for IoT and LoRa (Long Range WSNs) environments to sense, capture, and transfer the data reliably with low latency and high bandwidth. Various ML algorithms are employed to improve the efficiency LoRa and IoT networks. The major gaps found in the prevailing methods are: i) key optimization; ii) identifying topology changes, iii) intrusion detection & abnormal patterns; and iv) dynamic error recovery. This study claims the effective optimization of quantum key distribution to assign the optimized encrypted key to data and pass through nodes securely in the LoRa WSNs and addresses a few other challenges of prevailing approaches. The new QKD optimization protocol is proposed to detect abnormal patterns and intrusions in LoRa-WSNs to minimize data loss and maximize data integrity. CatBoost-MLGBA is employed along with LRM, L1R, and DQN-WSN to ensure that data is transmitted from one end to the other with high security and the finest path without error. The CatBoost model provides a dynamic solution for key optimization, energy savings, PDR, and robustness to attacks and addresses all the limitations of existing models. During the topology changes and link failures, the error is recovered and data is passed in LoRa, which enhances the delivery ratio.

Various security complexity issues in wireless network architecture[(1)] have been discussed by the authors, which clearly show the types of attacks, risks, threats, collisions, intrusions, eavesdroppers, etc. and how they are detected using adaptive ML models, which helps to study the ML models and propose the new method in a systematic manner. An empirical component analysis-based IDS[(2)] was introduced to detect the vulnerabilities and challenges faced by intrusions into WSNs with the help of deep learning models. The manual selection of features is overcome by the model. Though it has effective detection capability, the model has drawbacks in identifying the topology changes in LoRa. LSTM is used to identify the relevant features in the validated WSN datasets, such as KDD, UNSW, CICIDS, NSI-KDD, etc. The false prediction rate, accuracy rate, and collisions are the PEMs of the ECA-IDS model. The authors also pointed out the challenges and methods to overcome issues in node deployments, underwater node communications, noisy data removals, etc.

The ML-PANN[(3)] technique was proposed to enhance the security features of WSNs. Sybil attacks, malicious node attacks, DoS attacks, etc. will be detected by using a multi-layer perceptron ANN model. It also identifies the unknown sensor node position to transfer the data effectively. All the harmful nodes are localized for high data integrity. The only shortcoming of this method is that it will not detect malicious nodes in LoRa, where the distance and data travel time are high, which leads to a high energy depletion rate in ML-PANN. All the WSN security perspectives were discussed by the authors, along with techniques to enhance the security limitations[(4)]. Almost all network architectures have security issues with data transmission due to lack of memory, deployment of nodes in harsh environments, etc. The authors narrated how the protection mechanism secures the data packets travels from one end to other end. The results portrayed how the advanced AI and ML methods are utilized to optimize the performance of protocols in WSNs, which gives a detailed overview of security enhancements using AI and ML. Energy-aware sleep scheduling protocol (EAB-IFBA)[(5)] was introduced to identify network topology changes and dynamic sleep scheduling by sensor nodes to save energy at a maximum level. The model works well in an IoT environment with minimal limitations, like processing data in LoRa-WSNs,

overfitting errors, malicious node communications, threat nodes, the installation of high-end sensor nodes in insensitive environments, etc. An IDS model was introduced for IoT devices and packets[6] to boost the security features and to transfer the data more reliable. DL-KNN method was employed to focus more on capturing the significant patterns which seems to be unusual and the same is sent to BS for action taken. The routing was done by ABS-PN model which take decisions for data communications from each sensor node. The energy is optimized, security is enhanced and data loss is minimized. The drawback of this model is, it doesn't work in LoRa-WSNs as it has only short range network power. A modified reptile search[7] method along with a deep learning technique was employed for intrusion detection and data security, where node communications are traced and recorded every time and information is directly sent to the base station by a volatile node. The retile model searches the malicious and infected nodes and gives the information to neighboring nodes to take action on passing the data to the destination. This model is hyper-tuned by DL-RNN to identify abnormal features in the historic WSNS-DS dataset during training, testing, and validation. ML with block-chain security[8,9] model for WSNs and QoS-IWAP was introduced to detect dynamic link failure during data transmission from S → D. Data authentication, key distribution, encryption, traceability, and operation of decentralized network nodes reduce the network load and data loss, which enhances data integrity and reliable communication between nodes. Energy efficiency techniques in WSN and security enhancements[10] are the key elements taken by the author, who introduced a protocol using ML techniques. The protocol initially exchanges keys between the data for integrity and passes through the finest route for data delivery. The local interactions between the sensor nodes are managed by the CHs, and server interactions are managed by the end-point node, which was deployed to manage the information received by the BS. The drawback of this protocol is that the packet delivery and data transmission speed are not remarkable, which are considered crucial things in WSNs. The trust-based approach using ML[11] was developed to evaluate unknown scenarios like adaptability, coverage, scalability, limitless energy, etc. Trust-based ML security mechanisms enable the system to predict anomalies periodically and secure the data through successful data transmission. Multi-path routing[12] was suggested to secure data in underwater communications, and RSV-RP[13] was proposed to detect link failures in large-scale WSNs. Both models secure data in a robust way and ensure successful transmission of packets to the specified location. The latency time is a little high in this model compared to other protocols, but the PDR rate is 97%, which is a remarkable result that is highly needed in WSNs. DL-IDS-WSN[14] was proposed to detect anomalies in WSNs that occur during large packet deliveries. Malicious nodes and unusual patterns are identified for every transaction, which makes the system slow in data delivery. The process time to deliver the packets in MAN is high, and the error rate is high due to the lack of a conventional method for error recovery. RM-ANN[15] works on real-time adaptive networks and ensures a robust security level for the data sent to the server. ANN works as a node agent and decides the finest route on which the data is to be transferred from one end to the other. RSA with QKD is used for security purposes, and robustness is achieved up to 89%. The limitations are that the protocol does not work on LoRa as it has minimum power and lacks memory. ReLeC-WSN[16] was introduced for energy savings, network optimization, and error recovery to maintain and track node communications in all types of actions. The method works efficiently and achieves high PDR transmission speeds with less energy depletion. The limitations of the model are network adaptability, pattern recognition, and topology change prediction. Dynamic signal optimization in communication devices, which is considered a dynamic model, along with the machine learning technique named enhanced principal component analysis (ML-EPCA) to boost accuracy and delivery rate. This model serves as a robust approach to optimizing signal processing within communication devices.[17–19]. Robust deep learning optimization models are employed to detect anomalies in the WSNs, which improve complex pattern matching, real-time detection, adaptability of networks, FPR, and scalability. Here, all the malicious nodes are identified and marked, and duplicate data has been sent to those nodes for security[20–22]. Feed forward ANN, SMOTE & RFA, and ML Quantum Key Selector based protocols[23–25] are utilized to achieve the performance of WSNs in terms of energy depletion, PDR, minimized data loss, and high transmission speed. The limitations are LoRa deployments, data integrity, and robustness to attacks, which are considered significant things in all WSNs. All the existing models have not performed with impressive outputs in terms of security, integrity, robustness, re-transmission, rekeying, etc. To overcome the drawbacks of the existing models, the new anomaly detection model for LoRa is proposed to optimize QKD in WSNs by employing the CatBoost-GBA model. The unique features of CatBoost are:

- **Optimization of QKD:** The quantum key is optimized to secure the data transfer from node to node, to ensure systematic protection in LoRa-WSNs.
- **Error Detection and Recovery:** Errors are detected and recovered for a smooth transition of packets from the source node to the destination node.
- **Key Generation with Encryption**: An encrypted key is generated and appended to the data to prevent intruders from stealing it.
- **Network Topology Monitoring:** Regularly monitors the network topology changes and sends the data through the finest route, which leads to minimal energy consumption.

2 Methodology

The newly suggested QKD-optimized model CatBoost-MLGBA focuses on optimizing dynamic key distribution for end-to-end data security in LoRa WSNs and IoT. During the data transmission process, the quantum key is optimized and appended to the data, which cannot be decrypted by the intruders. The CatBoost model, along with encryption, access control, and node authentication, provides comprehensive data security by detecting unusual patterns, compromised node communications, and data injection attacks to ensure successful transmission from source to destination. The LRM with key generation method is employed to generate the random Q-key for all the data that is ready to transfer, which helps the CatBoost model optimize the instant key more effectively and encrypt the data for transmission. Lasso Regularization (L1R) handles node signal processing, spotting, and recovering the network errors in LoRa WSNs indirectly by reducing the overfitting. Deep Q-Networks combined with SPM is utilized for the finest routing, where the data is transferred in a robust manner. This CatBoost approach will enhance the security of IoT and LoRa WSNs for effective data transmission.

2.1 Proposed Methodology

The new quantum key distribution optimization model specifically works on data security during the end-to-end transmission process by employing the CatBoost machine learning-based gradient boosting method. Here, the sensor nodes are deployed in LoRa WSNs (long-range) and IoT environments. The number of nodes in the historical dataset trained, validated, and tested in this study is 9000–13000 by using the OMNET++ simulator. The CatBoost model works by distributing the key randomly to each piece of data that is ready to transfer. The five main features that CatBoost collects from the sensor node are, node transmission power, IoT or LoRa bandwidth, base station code, initial energy level, and nearest neighbor node information to forward the appropriate data by using the available application server. Once the quantum key is appended with data, the key is optimized by using encryption and a node authentication model to make it more secure. The optimization takes place in two ways: i) by installing chip code, and ii) by allocating bit-keys in the data to secure it from intruders. After the optimization, the data is boosted with two-factor authentication, which will be known by the sender and receiver. The details of the transmission gateway, application server, and network server are collected by the CatBoost protocol for the transmission process. LRM, L1R, and DQM methods are additionally employed for protocol enhancement and to achieve performance in terms of robustness, data delivery, network lifetime, and energy consumption. Assume $nodes(n)$ is deployed in LoRa environment. The $initial\_energy$ of the node is measured and $node\_id$ is allotted to each and every node deployed. The distance between each node is $node\_s$ and $node\_d$ which refers the source and destination of each node, the addition nodes will be incremented up to last node installation *will be as* $node_{s1}$, $node_{s2}...etc$. The access points, BS, key values are measured. The transferred bit rate is calculated initially by using the below equation,

$$Node_{TransferBitRate} = \left(\frac{Number\ of\ data\ Transferred}{Total\ Data\ in\ Network}\right] + (initial_{energy} - Balance_{energy}) \tag{1}$$

where, the balance energy of node and data transfer bit rate is measured. Also, successful delivery rate of data is calculated using the below equation using modulation scheme, LoRa bandwidth and signal noise ratio.

$$R = BW.log_2(1 + SNR) * Node_{TransferBitRate} \tag{2}$$

2.2 Training of WSNs-DS for Anomaly Detection & Simulation Settings

The historic WSNs-DS network LoRa data is used to train the newly proposed QKD optimization protocol to detect anomalies, which include gray-hole (GH) attacks, black-hole (BH) attacks, flooding attacks, scheduling attacks, and normal attacks. The model is trained using the gradient boosting method, where the quantum key is optimized to ensure secured data transmission from one end to the other. 19 features are extracted from WSNs-DS and normalized as data parameters for the training and testing processes in the OMNET testbed. 2250 nodes are employed and tested with 25 clusters, and the time limit is set to 3600 seconds with a maximum transmission range of 1000m. OMNET calls the CatBoost-MLGBA library and employs the machine learning techniques within the simulation to train and test the WSNs-DS data for effective learning of anomaly detection and network behavior. Real-time optimization can be done effectively with the help of training the CatBoost model to detect unusual node communications and abnormal patterns to improve the security feature. The transfer rate is increased during the testing process in the multi-hop LoRa networks, which gives promising results in terms of transmission speed, energy consumption, PDR, and security robustness.

Table 1. Parameter Settings for Simulation

Parameters	Values
LoRa WSNs Network Range	1000 m
Sensor Nodes Count	500-2250
Number of Clusters	25
Size of data packet	1000 Bytes
Mobility Model	Random Way Point
Nature of Traffic	LoRa Constant Bit Rate
Nature of Medium	IoT - Wireless Medium
Base Station Location	(50, 75)
Initial energy of each node	80 J
Sensing Range	10 m
Threshold Distance	130 m
Simulator Name and Version	OMNET C++
Learning Rate	0.5 (alpha)
Simulation Time	3600 seconds
Attackers Intensity	10%, 40%, 60%
Transmission Range	200 m
QKD Optimizing Protocol	CatBoost-MLGBA
Routing Protocol	Deep-QN with SPM

WSN Network Dataset/Simulation Source: https://opendatalab.com/WSN-DS

Training of the Historic WSNs-DS dataset includes,

- Data cleaning: handling missing data and outliers.
- Selection of 19 WSNS-DS LoRa datasets to detect unusual patterns & node communications.
- Labeling the data to train with the CatBoost model with different scenarios.
- Employ QKD for each and every piece of data that transmits from source to destination.
- Optimize the key using a node authentication and encryption model.
- Model integration with LoRa-WSNs to make real-time anomaly detection.
- Feedback loop to adapt to LoRa network dynamics.

2.3 QKD - Optimization with LRM

The anomaly detection CatBoost-MLGBA model works on optimizing the key generated by the QKD method. The Linear Regression Method with Key Generation (LRM) helps to optimize the key effectively in the source node. A key exchange model is employed to establish random security keys between communication nodes. Once the key is assigned to the data, it is encrypted and secured, which means it cannot be hacked by intruders. By optimizing the key, it ensures the distribution to the correct node, which will minimize the risk of compromise. CatBoost ensures that keys are generated with a high degree of randomness to withstand attacks based on predictability. Here, the data with keys is exchanged securely between two communication nodes. Assume that the two entities are communicating in the un-trusted LoRa networks. The key is established for communication from the key exchange model is implemented here using Diffe-Hellman or QKD. The LRM will optimize the key, which is generated by Diffe-Hellman or QKD and exchanged between the nodes. Once the key is optimized, it will detect the anomalies easily in LoRa-WSNs as it takes a long travel time between source and destination. Even if an eavesdropper intercepts, the data is securely transferred from S $\rightarrow$ D in a robust manner. The CatBoost model follows the below steps for QKD key optimization:

- Identify the source node that has data ready to transmit to destiny
- Generate a random key with QKD and distribute it to nodes
- Exchange random keys using Diffe-Hellman or QKD
- Optimize the key using LRM
- Measure the distance from source to destination
- Communicate with the neighbor node to exchange optimized keys
- Detect an anomaly and send a signal to the base station
- Transmit the encrypted data with optimized key values

- Calculate the data transmission speed and time
- Decrypt the data at the destination
- Calculate the energy level of nodes
- Initiate rekeying and rotation
- Repeat the process.

To calculate the anomaly detection score, identify the threat severity, complexity, affected node, and likely-to-be affected node. PDR is evaluated to calculate the ratio of successfully transmitted data from end to end in LoRa-WSNs. The latency time is assessed to identify the severity score of anomalies and intrusions from which node the key sends the signal to the base station. The PDR, anomaly detection, and latency are calculated by using the below equation.

$$Number\ of\ Packets \frac{Received}{Delivered}\ x100\ \&\ Calculate\ \{Node_{Key} |\ Anomaly_{Node}\} \tag{3}$$

where, $Anomaly_{Node}$ refers to the node that is affected in LoRa-WSNs where the intruders fail to steal the data due to the optimized key, which has highly encrypted data boosted by the CatBoost model. The historic WSNs-DS data is trained, validated, and tested in the testbed by using the function points of five types of attacks, which include gray-hole (GH) attacks, black-hole (BH) attacks, flooding attacks, scheduling attacks, and normal attacks. The LRM optimizes the QKD keys distributed in the node for successful transmission.

CatBoost -MLGBA Algorithm
Input: OMNET++ Simulation settings with Parameter Values
Begin: Initialize the nodes and set $num\_nodes = 2250$
Set $Communication\_Range = 10m$
Calculate $Initial_{Fitness} of\ Node$
Set Unique identifier Point in Network
Initialize WSNs-DS Dataset
Set train_ratio = 0.8 (As per network testbed)
Set $train\_size = int(train\_ratio * len(simulation\_data))$
$train_{data} = simulation_{data\ (train_{size}]}$
$test_{data} = simulation\_data\ (train_{size}:]$
Initialize CatBoost GBA Model
$Assign\ Q - Key$ for each node
Calculate node delivery = $(Assigning\_Node_{Data}]$
Encrypt $(Assigned\_Node_{Data}]$
confusion = confusion_matrix(y_test, y_pred)
$anomalies = [\,]$ and $anomaly\_keys = [\,]$
For i = (true_label, pred_label, data_id, timestamp) enumerate(zip(y_test, y_pred,
test_data[data_id], test_data[timestamp])): if pred_label == 1:
anomalies.append(data_id)
key = f_Anomaly_{data_id}_{timestamp}
anomaly_keys.append(key)
for anomaly, key (node, anomaly_keys)
Calculate the PDR and DL
Compute*Robustness Level*
Calculate $FinalNode_{Energy}$ based on active nodes in LoRa
Measure $Transmission\ Speed$ and Get the $Optimal\ Value$
End

2.4 CatBoost-MLGBA Key Flow Diagram

Here, the diagram shows the phase-level data flow from client to server as a model to present the work of the CatBoost-MLGBA model. The phase 1 deals with assigning keys to every node that has data and the QKD is optimized using the proposed model with authentication and encryption, which allows the node to travel to its destination. The intruders are detected by the key, and the signal has been sent to BS for recovery.

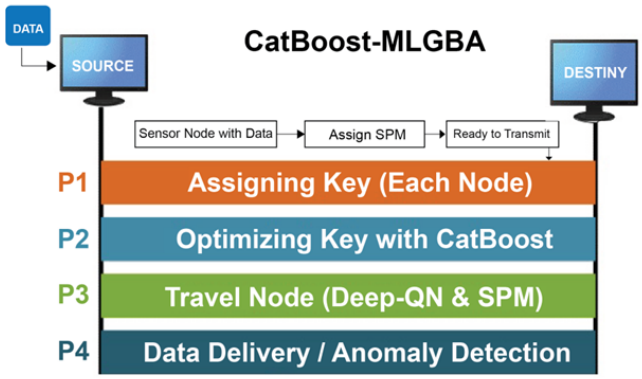

Fig 1. CatBoost Phase Flow Diagram

2.5 Error detection & recovery using Lasso Regularization (L1R)

L1R is a machine learning technique used to improve error detection and recovery in the CatBoost QKD optimization model. During the training and testing process of WSNs-DS, features like sensor readings, relevant error corrections, and critical data points are selected and trained for robust error detection. Before feature selection, the noisy data is removed for effective detection and recovery of errors in WSNs and to avoid false alarm triggers. L1R reads the patterns of intruders and matches in real-time for data integrity issues. L1R also reduces data dimensionality during the stringent energy constraints found in LoRa-WSNs. L1R assists in checking the overfitting, which helps to detect errors more effectively. During topology changes, the data is transferred with high security from source to destination. If an error is spotted or a dynamic link failure occurs, the L1R immediately takes an alternate path to deliver the data as part of the detection and recovery method. If unusual node communications are found, the L1R matches the patterns, detects the anomalies, and notifies the base station for quick resolution.

For efficient detection and recovery, the following steps are used.

- Employ L1R in the CatBoost model.
- Measure the node distance in LoRa-WSNs or IoT.
- Monitor the topology changes during data transmission.
- Identify the L1R patterns to detect unusual anomaly detections or node communications.
- If an error is found, detect the error data.
- Notify the base station.
- Recover using L1R for smooth transmission.

If there is an error, L1R indicates 1 and no error it indicates 0 to the base station to initiate the status of data transmission.

$$Error_{Detection} = Sum(Data\ Bytes) | Modulus\ some\ value \tag{4}$$

where, $(Data\ Bytes)$ denotes the data sent and received by each node and received checksum excluding the checksum appended by the sender.

2.6 Finest Routing with Deep-QN & SPM

To enhance routing decisions in LoRa WSN environments, DQN learns the optimal path for data delivery based on various factors like signal strength, node energy, traffic, topology changes, error rate, etc. DQN includes 4 steps: i) representing states ii) action space; iii) reward; and iv) training, which will maximize the routing decisions for successful data delivery. The quality of choosing the path reflects the delivery of data in a robust manner. DQN acts as a routing agent, while SPM measures the shortest path for reference. The DQN suggests the path and aligns with SPM, which meets the LoRa-WSNs goals of effective data transmission, reliable routing, and energy efficiency. As DQN evaluates the state of the network in a dynamic manner, the routing path will be suggested to the protocol in which the data has to be delivered. SPM also assigns weights to edges based on LoRa node distance or network graph. It acts as a baseline for routing decisions for data delivery. In equation 5, DQN SPM is calculated based on distance from each node in LoRa WSNs.

$$DQN - SPM\ d(v) = min\{d(v),\ d(u) + w(u,v)\} \tag{5}$$

where, $d(v)$, $d(u)$, $w(u,v)$ represents current distance estimation & weight of edges of nodes u and v. The initial distance of source node is set to 0.

2.7 Comparative Analysis using OMNET++

The comparative analysis is done for the proposed QKD optimization model CatBoost-MLGBA against the prevailing models such as DL-IDS-WSN[14], RM-ANN[15] and ReLeC-WSN[16] using the OMNET++ simulator tool, which helps to create realistic simulation scenarios with real-time and historical network data. The tool is integrated with ML and DL algorithms for optimization, anomaly detection, real-time simulations, etc. to produce key performance results. The comparative results obtained in OMNET++ can be inherited by future researchers. Visualization is done effectively, which helps the users analyze and simulate network protocols in a robust manner. As it supports parallel and distributed simulations, large-scale and LoRa-WSN model simulation processes can be done. Also, it identifies load balancing problems, node deployment issues, resource allocation, sleep and alive nodes, etc. to test the protocol effectively in the testbed.

2.8 CatBoost-MLGBA Evaluation Metrics

The performance and comparative analysis of the proposed CatBoost-MLGBA model is done against the baseline models such as DL-IDS-WSN[14], RM-ANN[15] and ReLeC-WSN[16] which was chosen in the preceding section. The following are PEM and equation to measure the protocol.

- **(EDR) Energy Depletion Rate:** The overall energy spent to transmit the data from S → D by the proposed QKD optimization protocol CatBoost-MLGBA and for data sensing and capturing.

$$EDR = \frac{(Initial\ Energy - Remaining\ Energy)}{(Time\ (Duration))} \tag{6}$$

- **(LN) Lifespan of Network:** Calculates the sensor node lifespan after the overall transmission process by dividing the total energy and consumed energy by CatBoost-MLGBA.

$$Network\ LS = \frac{Total\ Energy}{(Energy\ Depletion\ Rate)} \tag{7}$$

- **(PDR) Packet Delivery Ratio:** Calculates the amount of packets or data delivered against the total number of packets by the proposed optimization model during the testing process.

$$Packet\ Delivery\ Ratio = \frac{No.of\ Successful\ Packets}{(Total\ No\ of\ Packets)}\ x\ 100 \tag{8}$$

- **(TS) Transmission Speed:** Measures the speed in time of the data transferred by utilizing the finest path for successful delivery in the proposed model.

$$Data\ Transmission\ Speed = \frac{Data\ Size}{(Transmission\ Time)} \tag{9}$$

- **(DLR) Data Loss Rate :** Calculates the total amount of data sent and received from S → D during data sensing and capturing process by the CatBoost model.

$$Data\ Loss\ Rate = \frac{(Data\ Sent - Data\ Received)}{Total\ Amount\ of\ Data\ Sent} \tag{10}$$

- **(SR) Security Robustness :** To assess the robustness & calculate vulnerability and threat severity score of the CatBoost-MLGBA model against the attacks and topology changes.

$$SR\ (Robustness) = (1 - VS)\,x\,(1 - TSS) \tag{11}$$

3 Results and Discussions

This chapter deals with the various parameter evaluation results of the novel QKD optimization CatBoost-MLGBA and how it overcomes and addresses the drawbacks of the prevailing models such as DL-IDS-WSN[14], RM-ANN[15] and ReLeC-WSN[16] in terms of detecting anomalies and unusual patterns in LoRa-WSNs. The data loss and error rate are drastically minimized, and the robustness level to withstand attacks is improved. The new model works in an IoT environment and transmits the data more securely. In addition, a security error detection and recovery method is utilized to spot the error occurring in sensor nodes and notify the base station or neighbor node to find the best alternative route to minimize data loss. The unique feature of this model is that it detects abnormal communications and unusual patterns during topology changes and secures the data by optimizing QKD for effective transmission with a zero error rate. Figures 2, 3, 4, 5, 6 and 7 shows the simulation results where X axis shows the node counts and Y axis shows the percentage values of the existing and proposed protocols.

3.1 Energy Depletion Rate - Comparative Analysis

Figure 2 presents the amount of energy consumed by the sensor nodes deployed in an IoT environment by utilizing the novel QKD optimization CatBoost-MLGBA model. The performance of the new model is compared against the existing methods to measure the energy depletion rate of CatBoost. It is observed that the new model outperforms the existing model by consuming less energy during data transmission from source to destination. The model works well in taking the finest route for data delivery in a robust manner. During simulation testing, only 9% of energy was consumed by a sensor node after the delivery process. The sensor node will be idle and save energy until the next data arrives.

Table 2. Rate of Energy Depletion (%)

Node Counts / Protocols	500	1000	1500	2000	2250
ReLeC-WSN[16]	44	46	48	50	52
RM-ANN[15]	40	42	44	46	49
DL-IDS-WSN[14]	35	37	38	39	40
CatBoost-MLGBA (Proposed)	15	13	11	10	09

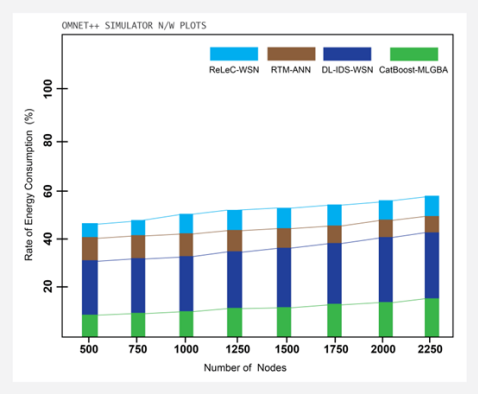

Fig 2. Performance Analysis of Energy Depletion

3.2 Network Lifespan - Comparative Analysis

The performance analysis of the QKD optimization model CatBoost-MLGBA in terms of network lifetime is portrayed in Figure 3. It is observed that the new optimization model outperforms on data transmission at a high speed, the lifetime of the sensor nodes increases. The nodes will stay idle or sleep during nil transmission, and only when a communication signal arises will they wake up and work for data transmission. The results are compared against the prevailing network optimization protocols, such as DL-IDS-WSN[14], RM-ANN[15] and ReLeC-WSN[16]. As the sensor nodes consume less energy while using the CatBoost technique with DQN to take the shortest path for data delivery, the lifetime of the sensor node is maximized to 95%.

Table 3. Network Lifetime (%)

Node Counts / Protocols	500	1000	1500	2000	2250
ReLeC-WSN [16]	30	28	22	18	16
RM-ANN [15]	72	68	65	60	58
DL-IDS-WSN [14]	82	78	76	73	70
CatBoost-MLGBA (Proposed)	**95**	**91**	**87**	**85**	**81**

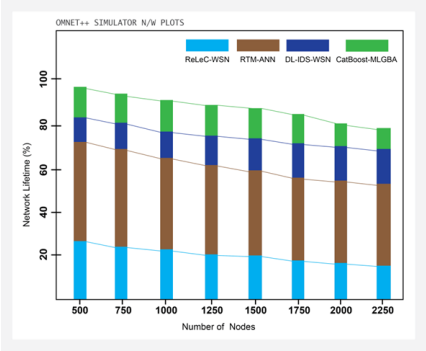

Fig 3. Performance Analysis of N/W Lifetime

3.3 Packet Delivery Ratio - Comparative Analysis

The packet delivery rate of the new and existing models is showcased in Figure 4. The total number of packets delivered from S $\rightarrow$ D is calculated against the total number of packets. The CatBoost optimization model outperforms in delivering the packet in a robust manner as it takes the shortest path to deliver the data without any noise or error. As the sensor node has the maximum energy to transmit the data at a high speed, the PDR is maximized up to 97%. The PDA results of CatBoost-MLGBA are compared with baseline approaches and the results are portrayed below. The new proposed model works well in IoT and LoRa-WSNs, which help end users, capture the data robustly without any dynamic link failure. If an error occurs, it is detected and recovered by the lasso regularization method.

Table 4. PDR Rate Analysis (%)

Node Counts / Protocols	500	1000	1500	2000	2250
ReLeC-WSN [16]	30	26	20	18	18
RM-ANN [15]	74	68	62	60	58
DL-IDS-WSN [14]	85	80	76	74	74
CatBoost-MLGBA (Proposed)	**97**	**94**	**90**	**88**	**84**

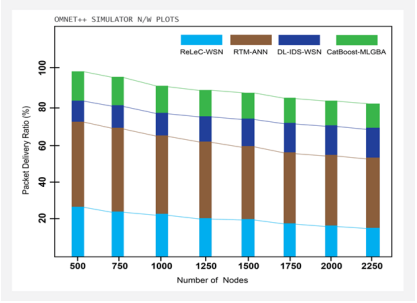

Fig 4. Performance Analysis of PDR Rate

3 .4 Data Transmission Speed - Comparative Analysis

The simulation results of the data transmission speed of the proposed QKD optimization protocol are shown in Figure 5. The results are compared against the other ML quantum key distribution models & the results are shown below. As the new model distributes the key in an optimized manner by employing the CatBoost model, the sensor node gets the data with a secured key that cannot be stolen by intruders. The DQN with shortest path model works to identify the finest route to transfer the data, and in case a dynamic link failure occurs, the alternate route will be taken by the node itself in the shortest oath to deliver the data in a robust way. The data is transferred in 6 seconds with 500 node counts, which is comparatively high compared to existing models.

Table 5. Data Transmission Speed (in Seconds)

Node Counts / Protocols	500	1000	1500	2000	2250
ReLeC-WSN[16]	30	32	33	33	35
RM-ANN[15]	24	26	26	28	30
DL-IDS-WSN[14]	20	20	22	24	25
CatBoost-MLGBA (Proposed)	**06**	**07**	**08**	**08**	**10**

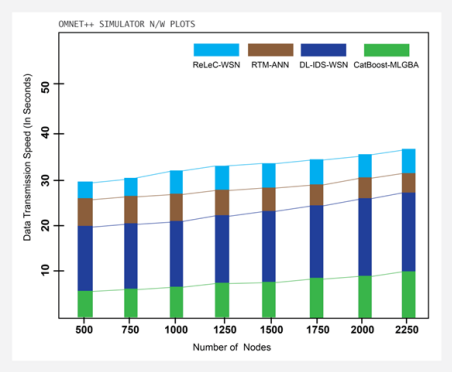

Fig 5. Performance Analysis of Transmission Speed

3.5 Security Robustness - Comparative Analysis

Figure 6 presents the security robustness of data transmission from source to destination without any loss of the proposed novel QKD dynamic key distribution optimization protocol, CatBoost-MLGBA. The simulation results are compared against the existing models, such as DL-IDS-WSN[14], RM-ANN[15] and ReLeC-WSN[16]. It is noted that the new model detects abnormal node communications, unusual patterns, and anomalies during the data transmission and secures the data for successful delivery. The model is trained with historical DS-WSN data for greater robustness to work on LoRa-WSNs and IoT environments.

Table 6. Robustness to Attacks (%)

Node Counts / Protocols	500	1000	1500	2000	2250
ReLeC-WSN[16]	24	20	20	18	18
RM-ANN[15]	64	60	60	58	56
DL-IDS-WSN[14]	78	76	75	74	72
CatBoost-MLGBA (Proposed)	**91**	**88**	**86**	**82**	**80**

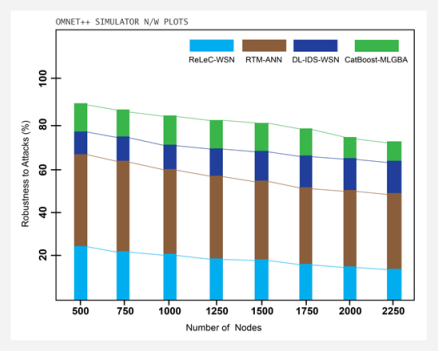

Fig 6. Performance Analysis of SRA

3.6 Data Loss – Comparative Analysis

The simulation results of data loss are assessed by employing the new CatBoost-MLGBA model to measure the amount of data loss during data transmission. The comparative results are presented below. Node placement, communication channels, and power levels are constantly increased in this model as the EDR is very less due to the high speed and finest route process for successful data delivery. The data loss is minimized to 10%, which is comparatively low compared to the prevailing models. The sensor node hibernates itself when there is no communication from the base station or neighbor node. Only the nodes will work during the data transmission process. The link failure is detected in a dynamic manner, which is one of the major reasons for the high amount of data transmission.

Table 7. Data Loss Rate (%)

Node Counts / Protocols	500	1000	1500	2000	2250
ReLeC-WSN [16]	58	60	62	62	64
RM-ANN [15]	45	48	52	53	54
DL-IDS-WSN [14]	37	39	42	44	45
CatBoost-MLGBA (Proposed)	**10**	**12**	**16**	**18**	**20**

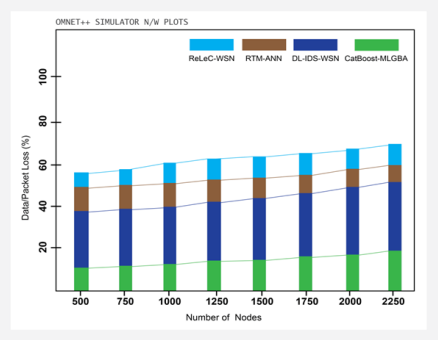

Fig 7. Performance Analysis of Data Loss

4 Conclusion

The proposed novel artificial intelligence-based WSN security protocol, CatBoost-MLGBA (Machine Learning Gradient Boost Algorithm), is used to enhance security features by detecting abnormal patterns and intrusions in WSNs to minimize data loss and maximize the packet delivery ratio in LoRa-WSNs. As LoRa has more nodes deployed, there is a high possibility of malicious nodes affecting data transfer and network shutdown. To overcome the issue, quantum key distribution is optimized by the CatBoost model, where the network data is trained to detect anomalies to secure data transfer from S → D in a robust manner. To predict the type of network attack, LRM is used, which optimizes the QKD in the form of dynamic encryption and prevents data L1R is employed for error recovery, and DQN combined with the shortest path is used to find the alternate finest route when node failure occurs during data transfer. The WSN-DS historical dataset is used with 19 features to train the model effectively for anomaly detection. CatBoost-MLGBA is tested in the network testbed, where 1000+ sensor nodes are deployed in long-range networks to collect real-time traffic data. The model captures the type of attacks, amount of data, number of active nodes, etc., and optimizes the quantum key to add a security feature for data transfer. The promising results of CatBoost-MLGBA show that the model addresses security features in terms of minimizing data loss and maximizing the PDR. The OMNET++ simulator is used to assess the performance of the CatBoost-MLGBA model. Energy consumption is reduced to 9%, data loss is minimized to 10%, network lifespan and PDR are maximized to 95% and 97%, robustness to attacks is increased to 91%, and DTR speed is minimized to 6 seconds.

The model has a few limitations, such as: i) high overhead where it deals with computational resources, bandwidth, and memory; ii) challenges in heterogeneity; iii) frequent monitoring of security breaches; and iv) complex routing decisions, etc. The recommendations are that, in the future, the model can be enhanced by adding real-time LoRa-WSNs data to build a breach assumption AI model to address heterogeneity and other challenges.

References

1) Ahmad R, Wazirali R, Abu-Ain T. Machine Learning for Wireless Sensor Networks Security: An Overview of Challenges and Issues. *Sensors*. 2022;22(13):1–35. Available from: https://doi.org/10.3390/s22134730.
2) Zhiqiang L, Mohiuddin G, Jiangbin Z, Asim M, Sifei W. Intrusion detection in wireless sensor network using enhanced empirical based component analysis. *Future Generation Computer Systems*. 2022;135:181–193. Available from: https://doi.org/10.1016/j.future.2022.04.024.
3) Gebremariam JGG, Panda S, Indu. Localization and Detection of Multiple Attacks in Wireless Sensor Networks Using Artificial Neural Network. *Wireless Communications and Mobile Computing*. 2023;2023:1–29. Available from: https://doi.org/10.1155/2023/2744706.
4) Gutierrez-Portela F, Almenarez-Mendoza F, Calderon-Benavides L, Romero-Riano E. Security perspective of wireless sensor networks. *Revista UIS ingenierias*. 2021;20(3):189–202. Available from: https://doi.org/10.18273/revuin.v20n3-2021014.
5) Nithyanandh S, Omprakash S, Megala D, Karthikeyan MP. Energy Aware Adaptive Sleep Scheduling and Secured Data Transmission Protocol to enhance QoS in IoT Networks using Improvised Firefly Bio-Inspired Algorithm (EAP-IFBA). *Indian Journal Of Science And Technology*. 2023;16(34):2753–2766. Available from: https://doi.org/10.17485/IJST/v16i34.1706.
6) Kaushik A, Al-Raweshidy H. A novel intrusion detection system for internet of things devices and data. *Wireless Networks*. 2023;p. 1–10. Available from: https://doi.org/10.1007/s11276-023-03435-0.
7) Dahou A, Elaziz MA, Chelloug SA, Awadallah MA, Al-Betar MA, Al-Qaness MAA, et al. Intrusion Detection System for IoT Based on Deep Learning and Modified Reptile Search Algorithm. *Computational Intelligence and Neuroscience*. 2022;2022:1–15. Available from: https://doi.org/10.1155/2022/6473507.
8) Ismail S, Dawoud DW, Reza H. Securing Wireless Sensor Networks Using Machine Learning and Blockchain: A Review. *Future Internet*. 2023;15(6):1–45. Available from: https://doi.org/10.3390/fi15060200.
9) Nithyanandh S, Jaiganesh V. Quality of service enabled intelligent water drop algorithm based routing protocol for dynamic link failure detection in wireless sensor network. *Indian Journal of Science and Technology*. 2020;2020(16):1641–1647. Available from: https://doi.org/10.17485/IJST/v13i16.19.
10) Manikandan S, Suganthi S, Gayathiri R. Optimal Energy Efficiency Techniques and Security Enhancement in Wireless Sensor Network Using Machine Learning. In: 2022 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS), 08-09 December 2022, Chennai, India. IEEE. 2023;p. 1–5. Available from: https://ieeexplore.ieee.org/abstract/document/10046790.
11) Khan T, Singh K, Shariq M, Ahmad KS, Savita KS, Ahmadian A, et al. An efficient trust-based decision-making approach for WSNs: Machine learning oriented approach. *Computer Communications*. 2023;209:217–229. Available from: https://doi.org/10.1016/j.comcom.2023.06.014.
12) Prakash K, Sathya S. A Deep Learning-based Multi-Path Routing Protocol for Improving Security using Encryption in Underwater Wireless Sensor Networks. In: 2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC), 06-08 July 2023, Coimbatore, India. IEEE. 2023;p. 581–588. Available from: https://ieeexplore.ieee.org/document/10193733.
13) Nithyanandh S, Jaiganesh V. Dynamic Link Failure Detection using Robust Virus Swarm Routing Protocol in Wireless Sensor Network. *International Journal of Recent Technology and Engineering*. 2019;8(2):1574–1579. Available from: https://www.ijrte.org/wp-content/uploads/papers/v8i2/B2271078219.pdf.
14) Salmi S, Oughdir L. Performance evaluation of deep learning techniques for DoS attacks detection in wireless sensor network. *Journal of Big Data*. 2023;10(1):1–25. Available from: https://doi.org/10.1186/s40537-023-00692-w.
15) Hassan KMA, Madkour MA, Nouh SAEH, Realtime. A Realtime Adaptive Trust Model Based on Artificial Neural Networks for Wireless Sensor Networks. *Journal of Cyber Security and Mobility*. 2023;12(04):519–546. Available from: https://journals.riverpublishers.com/index.php/JCSANDM/article/view/19103.
16) Sharma T, Balyan A, Nair R, Jain P, Arora S, Ahmadi F. ReLeC: A Reinforcement Learning-Based Clustering-Enhanced Protocol for Efficient Energy Optimization in Wireless Sensor Networks. *Wireless Communications and Mobile Computing*. 2022;2022:1–16. Available from: https://doi.org/10.1155/2022/3337831.

17) Banupriya CV, Kowsalya R. Machine Learning-Driven Robust Optimization of Communication Signals in Sensor Wearable Devices for Early Stage Epilepsy Seizure Prediction using EPCA. *Indian Journal Of Science And Technology*. 2023;16(25):1898–1909. Available from: https://doi.org/10.17485/IJST/v16i25.1290.
18) Abhale AB, Reddy AJ. Deep Learning Perspectives to Detecting Intrusions in Wireless Sensor Networks. *International Journal of Intelligent Systems and Applications in Engineering*. 2023;11(2S):18–26. Available from: https://ijisae.org/index.php/IJISAE/article/view/2504.
19) Song Y, Liu Z, He X. A Data Transmission Path Optimization Protocol for Heterogeneous Wireless Sensor Networks Based on Deep Reinforcement Learning. *Journal of Computer and Communications*. 2023;11(08):165–180. Available from: https://www.scirp.org/journal/paperinformation?paperid=127402.
20) Ren J, Li S, Song Y, Li M. Deep Learning Based Identification Method for Signal-Level Wireless Protocol. *IEEE Access*. 2022;10:118187–118197. Available from: https://ieeexplore.ieee.org/document/9942820.
21) Falahkheirkhah K, Yeh K, Mittal S, Pfister L, Bhargava R. Deep learning-based protocols to enhance infrared imaging systems. *Chemometrics and Intelligent Laboratory Systems*. 2021;217:104390. Available from: https://doi.org/10.1016/j.chemolab.2021.104390.
22) Banupriya CV, Devi AD. Robust Optimization of electroencephalograph (EEG) Signals for Epilepsy Seizure Prediction by utilizing VSPO Genetic Algorithms with SVM and Machine Learning Methods. *Indian Journal of Science and Technology*. 2021;14(16):1250–1260. Available from: https://doi.org/10.17485/IJST/v14i16.625.
23) Muruganandam S, Joshi R, Suresh P, Balakrishna N, Kishore KH, Manikanthan SV. A deep learning based feed forward artificial neural network to predict the K-barriers for intrusion detection using a wireless sensor network. *Measurement: Sensors*. 2023;25:1–9. Available from: https://doi.org/10.1016/j.measen.2022.100613.
24) Okey OD, Maidin SS, Rosa RL, Toor WT, Melgarejo DC, Wuttisittikulkij L, et al. Quantum Key Distribution Protocol Selector Based on Machine Learning for Next-Generation Networks. *Sustainability*. 2022;14(23):1–15. Available from: https://doi.org/10.3390/su142315901.
25) Tan X, Su S, Huang Z, Guo X, Zuo Z, Sun X, et al. Wireless Sensor Networks Intrusion Detection Based on SMOTE and the Random Forest Algorithm. *Sensors*. 2019;19(1):1–15. Available from: https://doi.org/10.3390/s19010203.