information

IMPACT FACTOR 2.9

CITESCORE 6.5

Article

Enhancing Data Security in Satellite Communication Systems: Integrating Quantum Cryptography with CatBoost Machine Learning

Mohd Nadeem, Syed Anas Ansar, Sakshi Halwai, Arpita Singh and Rajeev Kumar

Special Issue

2nd Edition of 5G Networks and Wireless Communication Systems

Edited by

Dr. Theofilos Chrysikos and Prof. Dr. Michael Paraskevas

MDPI

https://doi.org/10.3390/info17030220

information

MDPI

Article

Enhancing Data Security in Satellite Communication Systems: Integrating Quantum Cryptography with CatBoost Machine Learning

Mohd Nadeem [1,*], Syed Anas Ansar [2,*], Sakshi Halwai [3], Arpita Singh [4] and Rajeev Kumar [1]

[1] Department of Computer Science and Engineering, Shri Ramswaroop Memorial University, Barabanki 225003, India
[2] School of Computer Science and Engineering, Sandip University, Sijoul, Madhubani 847235, India
[3] Department of Computer Applications, Chandigarh School of Business, CGC University, Mohali 140307, India; sakshihalwai547@gmail.com
[4] Department of Computer Application, UIET, Chhatrapati Shahuji Maharaj University, Kanpur 208024, India; singharpita999@gmail.com
* Correspondence: mohdnadeem.dcse@srmu.ac.in (M.N.); syed000anas@gmail.com (S.A.A.); Tel.: +91-8318099719 (M.N.)

Abstract

In modern communication networks, particularly satellite-based systems, data security faces significant challenges from vulnerabilities such as signal interception, jamming, and latency during long distance transmissions. Traditional cryptographic methods are increasingly vulnerable to quantum computing threats, underscoring the need for advanced solutions to protect data integrity, confidentiality, and availability. This research investigates the fusion of quantum cryptography and Machine Learning (ML) to improve security in satellite communication. The Quantum Key Distribution (QKD), which is grounded in quantum mechanics, enables unbreakable encryption by detecting eavesdropping via quantum state disturbances. The CatBoost ML algorithm is applied to a dataset of 10,000 records featuring categorical attributes for prioritizing security elements such as anomaly detection, encryption types, and access controls. The model yields an accuracy of 89.23% and Area under Curve the Receiver Operating Characteristic (AUC-ROC) score of 94.56%, effectively predicting threat levels. Feature importance reveals anomaly detection (28.5%) and quantum encryption (22.3%) as primary contributors. While hurdles such as high implementation costs and transmission range limitations persist, this quantum ML synergy provides a proactive, adaptive framework for resilient, future-ready communication networks.

Keywords: data security; satellite communication; quantum cryptography; quantum computing; QKD; CatBoost

Academic Editor: Jiguo Li

Received: 22 December 2025
Revised: 24 January 2026
Accepted: 16 February 2026
Published: 25 February 2026

1. Introduction

Satellite data transfer is the idea of a source of communication information, as a message in the form of code is encompassed in a number of models [1]. The receiver decodes the message to be able to understand it after it has been sent through a channel. Since the signals of the satellite need to travel large distances from Earth to the satellite, satellite communication has higher latency than other communications do. The setup and maintenance of the satellite communication system can be expensive. This cost includes maintenance along with launches of satellites and ground stations [2]. Weather conditions such as rain, snow, or atmospheric disturbances tend to affect satellite communications,

and the signal tends to be lost or degraded. In particular, when the population is denser, the limited bandwidth capacity of satellites results in congestion and decelerated transmission speeds. Signal interference, hacking, and privacy infringement are some of the security threats against which satellite communication can be susceptible. Space junk and potential contamination of outer space are two instances where satellite launch and operation can influence the environment [3]. Global agreements and legislation regarding frequencies, orbital positions, and licensing govern satellite communication and sometimes make deployment and operation problematic. Table 1 shows the cost of satellite maintenance; Figure 1 represents its bar graph.

Table 1. Data related to the maintenance cost of satellite communication over the last 15 years.

Year	Number of Satellites Maintained	Maintenance Cost (In Million USD)
2010	5	12.5
2011	6	13.8
2012	6	14.2
2013	7	15.7
2014	8	17.3
2015	9	18.9
2016	10	20.4
2017	11	22.1
2018	12	24
2019	13	26.3
2020	14	28.5
2021	15	31
2022	16	33.7
2023	17	36.2
2024	18	39
2025	20	42.5

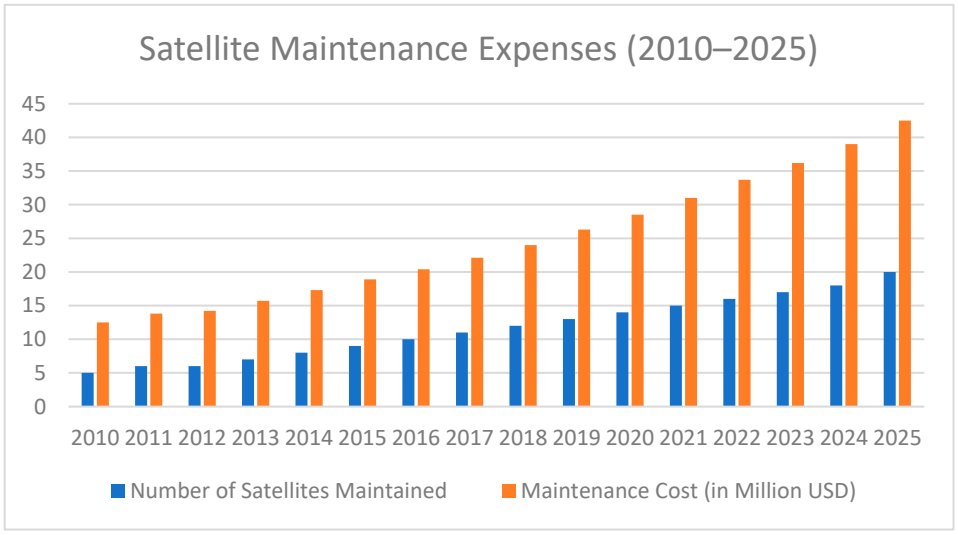

Figure 1. Satellite maintenance expenses (USD).

The priority selection of security in satellite communication uses principles of encryption to enable two parties to exchange encryption keys securely [4]. In contrast to conventional cryptographic systems that rely on the difficulty of computation, QKD leverages quantum principles to detect potential eavesdropping attempts because any measurement or disturbance introduced by an adversary alters the transmitted quantum states and increases the observed error rate [5]. It is important to note that this detection

capability is inherently probabilistic, since practical QKD systems infer intrusion attempts statistically through measurable indicators such as the Quantum Bit Error Rate (QBER) [6]. Therefore, under realistic operational conditions and imperfect authentication of the classical channel, sophisticated adversaries may still attempt man-in-the-middle strategies, meaning that eavesdropping detection cannot be interpreted as an absolute guarantee in every scenario [7].

Quantum entanglement is a situation where the quantum state of every particle within a group cannot be described separately from the other particles' states, regardless of the distance separating them [8]. This phenomenon lies at the heart of the difference between quantum and classical physics; entanglement is an intrinsic feature of quantum mechanics but cannot be found in classical mechanics. The physical attributes, such as position, momentum, spin, and polarization, are quantified on entangled particles, and they may occasionally exhibit complete correlations. These effects were described by Albert Einstein, Boris Podolsky, and Nathan Rosen. Einstein and coauthors claimed that these effects were impossible, since they violated the principle of local realism with regard to causality (by which Einstein described it as "spooky action at a distance"), and proposed that the standard interpretation of quantum mechanics is not complete [9]. The later experiments verifying the quantum mechanics' surprise predictions included measurements of the polarization or spin of the entangled particles in two remote locations, with the outcome leading to a statistical breach of Bell's inequality. These findings substantiated that quantum entanglement-inspired correlations cannot be explained in terms of local hidden variables present within the particle; although entanglement might permit statistical correlations among events separated from each other, it does not enable communication at superluminal velocities [10]. In the realm of quantum physics, particles such as electrons and photons exhibit wave-like properties that can combine, which is known as being superposed [11].

Superposition is important for developing quantum technologies such as quantum computing. The most well-known protocol, BB84, involves the transmission of quantum bits (qubits) via a quantum channel, accompanied by classical communication for the correction of errors and amplification of privacy. This ensures that the resulting shared key is kept secret and can be applied to unbreakable encryption schemes such as the one-time pad. Although quantum security ensures unparalleled security, its deployment is hindered by issues such as costs, short range of transmission, and compatibility challenges with current networks.

2. Literature Review

The main principle of ML, molecular dynamics, light harvesting systems, molecular electronic properties, surface reaction networks, density functional models, phase classification, and quantum simulations are areas of science and engineering where ML methods have exhibited remarkable success [12]. In addition, owing to their ability to accelerate searches for new energy production/storage materials and to compute exponentially large amounts of data, modern ML methods have also been used in the state space of complex condensed systems [13]. The application of quantum computers to detect patterns has emerged as a new front, as they have advanced rapidly. The ML method is anticipated to be a viable use of quantum computers in the near future given recent developments in both ML and quantum computing [14]. The past few years have witnessed the introduction of a variety of ML algorithms [12]. Indisputably, in the era of big data, and there is an urgent need to develop revolutionary algorithms capable of performing ML operations on large scientific datasets for numerous optimization-based industrial and technical purposes. The quantum support vector machine to accurately recognize handwritten digit "6" and digit "9" as a proof of concept [15].

The integration of ML and quantum computing in enhancing the security of satellite data communication has garnered significant attention [16] in recent years. This study synthesizes key findings from various studies, focusing on the application of ML algorithms, quantum cryptography, and their combined potential in addressing security challenges in satellite communication systems. This review covers the evolution of satellite communication security, the role of quantum mechanics in secure key distribution, the application of ML in anomaly detection and feature prioritization, and the challenges and future prospects of quantum ML in this domain. Satellite communication systems are inherently vulnerable because of their reliance on long-distance signal transmission, which introduces latency, susceptibility to interference, and potential security breaches such as hacking and privacy infringement [17]. Radhakrishnan et al. provided a comprehensive survey of inter-satellite communication, highlighting physical layer vulnerabilities, including signal interception and jamming, which compromise data confidentiality and integrity [1]. The high cost of satellite deployment and maintenance, as noted by Abdelsadek et al., further complicates the implementation of robust security measures, as resources are often prioritized for operational efficiency over security enhancements [18]. Environmental factors, such as space debris and atmospheric disturbances, also pose risks to signal reliability, necessitating advanced security protocols to ensure uninterrupted communication [19]. Security in satellite communication requires addressing multiple facets, including encryption, anomaly detection; secure key management, and ant jamming techniques. Traditional cryptographic methods, such as symmetric and asymmetric encryption, have been widely used to secure data transmission. The security features are increasingly vulnerable to quantum computing advancements, which can potentially break conventional encryption algorithms [14]. This has spurred interest in quantum-based security solutions that leverage the principles of quantum mechanics to provide theoretically unbreakable encryption.

The ML approach has shown remarkable success in various domains, including cybersecurity, because of its ability to process large datasets and identify complex patterns [20]. In satellite communication, ML algorithms are employed for anomaly detection, feature prioritization, and predictive modeling to increase security. Altulaihan et al. highlighted the use of ML-based anomaly detection to identify unusual patterns in network traffic, which could indicate potential cyber threats such as malware or insider attacks [21]. In signature-based systems, ML-based anomaly detection focuses on behavioral analysis, enabling the identification of zero-day attacks [22]. Techniques such as statistical modeling and neural networks are commonly used to improve detection accuracy and reduce false positives [23]. K. Yang et al. explored the application of quantum clustering algorithms for data classification, which is particularly relevant for prioritizing security features in satellite communication [24]. These algorithms can categorize complex datasets without requiring gradient-based optimization, making them suitable for handling the intricate distributions of satellite data. Kirmani et al. (2019) demonstrated the efficacy of ML in pattern recognition tasks, such as peptide identification, which can be extended to classifying security threats in satellite systems [13]. The integration of ML with quantum computing, known as quantum ML, is an emerging field with significant potential.

The CatBoost algorithm, a gradient boosting method, is particularly relevant for this application because of its ability to handle categorical features and missing data effectively [25]. CatBoost use of oblivious decision trees and target encoding reduces overfitting, making it suitable for prioritizing security features such as encryption, anomaly detection, and access control in satellite communication. Nadeem et al. employed fuzzy AHP to assess security factors in software design, demonstrating the applicability of decision-making algorithms in prioritizing security measures [26]. Similar approaches can be adapted to satellite systems to rank security features on the basis of their impact on confidentiality,

integrity, and availability. Despite the promise of quantum cryptography and ML, several challenges hinder their integration into satellite communication systems. Quantum cryptography faces limitations in transmission range, requiring satellite-based relays and quantum repeaters to achieve global coverage [27]. The high cost of quantum infrastructure and its incompatibility with existing networks further complicate deployment [28]. ML algorithms, while effective, require careful calibration to minimize false positives in anomaly detection and ensure robust performance in dynamic environments.

The literature underscores the transformative potential of quantum cryptography and ML in securing satellite data communication. The quantum mechanism offers unparalleled security through unbreakable encryption keys, whereas ML algorithms enhance anomaly detection and feature prioritization. The technical and economic barriers must be overcome to realize their full potential.

3. Materials and Methods

To build a quantum approach of security that is capable of classifying synthetic generated data [29]. The ML algorithm is able to obtain all security parameters without gradient-based optimization [30]. To assist in categorizing data with intricate distributions, sublabels are provided. Sublabel is a subordinate label that falls under the major label; it is also referred to as a subclass. Identifying the appropriate sublabels and using them to form the categorization of security features are two key duties. This demonstrates that numerical simulation can be utilized to solve different categorization challenges, especially when various material phases are constructed. The key features of satellite data communication are mentioned and shown in Figure 2.

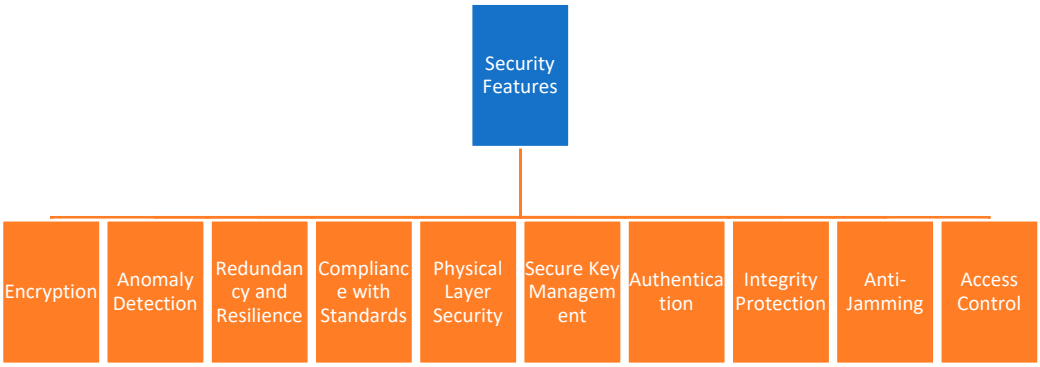

Figure 2. Satellite data communication security features.

3.1. Encryption

One key security tool is encryption, which turns unencrypted information into ciphertext—a coded state that prevents unwanted users from being able to read it [31]. This process ensures that without the appropriate decryption key, information cannot be decrypted, even if it is intercepted in satellite communication. Symmetric encryption shares the same key for encryption and decryption, and asymmetric encryption uses pairs of public and private keys [32]. While asymmetric encryption offers more security for key exchange and digital signatures, symmetric encryption is faster and can be applied to large volumes of data [33]. Applications such as secure file storage, email, online banking, and virtual private networks (VPNs) often employ encryption. The AES (Advanced Encryption Standard) and RSA (Rivest–Shamir–Adleman) are two modern encryption techniques. Robust encryption is also required to maintain data validity, integrity, and confidentiality in satellite data communication.

3.2. Anomaly Detection

One of the security techniques used to identify patterns in data that are significantly different from normal behavior is anomaly detection [34]. It is used to detect suspicious activity on networks and systems, often indicating potential malware infections, insider attacks, or security weaknesses in satellite data communication. Anomaly detection is beneficial because it is able to identify unknown or zero-day attacks by focusing on behavior instead of signatures, unlike conventional security systems, which are based on known threat signatures. Statistical modeling, machine learning, and rule-based techniques are some of the techniques utilized. Fraud detection software, intrusion detection systems, and live monitoring utilities are all able to utilize anomaly detection systems. To ensure accuracy and reduce false positives, these systems need to be accurately calibrated. By reducing damage, allowing for rapid incident response, and providing early warning signs of potential threats, effective anomaly detection enhances overall cybersecurity [35]. An important component of proactive defense strategies for companies of all sizes in a time of evolving cyber threats is anomaly detection [36].

3.3. Redundancy and Resilience

To guarantee continuous operation and rapid recovery from malfunctions, interruptions, or attacks, redundancy and resilience are the basic principles of cybersecurity and system design. Duplicate servers, network channels, or data storage are instances of redundant systems, components, or processes [37]. The backup system assumes operation instantly in the event that the main system breaks down, minimizing downtime and preventing service disruptions [38]. This is most important in industries where continuous availability is essential, including healthcare, finance, and telecommunications. In contrast, resilience is the system's ability to withstand, adapt to, and recover quickly from adverse conditions, whether caused by hardware failures, cyberattacks, or natural catastrophes. Resilient systems anticipate problems and have measures in place to detect, manage, and rebound from them with minimal disruption. Disaster recovery plans, regular backups, and automated failover processes can all be included in this [39]. Redundancy and resilience complement each other by enhancing the overall cybersecurity stance of an organization by ensuring business continuity and mitigating the impact of unexpected events. Investment in strong architecture and redundant infrastructure is now unavoidable to ensure safe and reliable operations in today's digital era, where downtime leads to considerable financial and reputational damage [40].

3.4. Compliance with Standards

In the context of cybersecurity, standard compliance involves adherence to regulations, guidelines, and best practices that have been established to protect user data, information systems, and organizational assets. To ensure consistency, accountability, and security across digital platforms, such standards are often designed by national, international, or industry-specific organizations. The NIST guidelines, the GDPR for data protection, ISO/IEC 27001 [41] for information security management, and the HIPAA for healthcare data protection are a few examples of standard frameworks [42]. Applying appropriate rules, conducting regular risk assessments, training employees, and maintaining security controls aligned with the applicable standard are all components of compliance. By demonstrating a commitment to data protection, it helps companies find weaknesses, steer clear of fines, and win the confidence of stakeholders and customers [43]. Severe consequences, such as data breaches, legal suits, penalties, and reputational damage to an organization, can be fatal to a company due to noncompliance [44]. Compliance is therefore both a strategic business imperative and a regulatory requirement. Promoting a proactive

cybersecurity posture ensures that organizations remain current with the latest threats and technological trends. Ultimately, compliance with standards enhances data integrity and privacy, strengthens an organization's security position, and lays the foundation for long-term, secure digital operations.

3.5. Physical Layer Security

Securing the physical infrastructure hardware, cables, network appliances, and data centers that support digital systems is the primary purpose of physical layer security. Physical layer security shields against physical threats such as theft, sabotage, tampering, natural disasters, or unauthorized equipment access, as opposed to cybersecurity mechanisms that shield against virtual threats. Several strategies are used for physical security deployment [45]. Physical layer security is especially crucial in environments where violations have serious consequences, such as banks, hospitals, or military bases. Physical layer security provides a total protection methodology when used together with more sophisticated cybersecurity techniques [42]. It ensures the confidentiality, integrity, and availability of important information and systems by avoiding physical penetration of even the most protected systems.

3.6. Secure Key Management

The critical element of cryptographic systems is secure key management, which encompasses secure creation, sharing, storage, rotation, and disposal of cryptographic keys [46]. Even with strong encryption algorithms, the security of encrypted information can be breached if these keys, which are the foundation of the encryption and decryption processes, are compromised. By restricting access to the keys at the right time to only authorized persons, good key management maintains the confidentiality, integrity, and authenticity of digital communications. It involves establishing protocols for key lifecycles, such as for how long the key remains active, how the parties exchange it, and how it is ultimately taken out of circulation or destroyed. Hardware security modules and automated key management systems are often used to provide secure handling of keys [47].

3.7. Authentication

To ensure that only the right individuals can access sensitive information or perform specific tasks, authentication is the process of verifying the identity of a user, device, or system prior to granting access to resources. As the first line of defense against unauthorized access and cyberattacks, it is one of the most critical elements of cybersecurity [47]. Multifactor authentication, which uses two or more of these methods to significantly enhance security, is increasingly being employed in many systems. The user might have to provide both a password and a one-time verification code that is sent to their phone to gain access to an online banking system [48]. This multilayered approach reduces the likelihood that stolen or guessed credentials would be used to gain access.

3.8. Integrity Protection

Integrity protection ensures data accuracy, completeness, and intactness during processing, transmission, and storage. Since information reliability is extremely important for financial transactions, legal contracts, decision-making, and other activities, this concept is central to cybersecurity. Attackers may alter data without detection if integrity protection is not implemented, which could have serious consequences such as fraud, loss of trust, or compromised systems. Digital signatures, which confirm the integrity and authenticity of information, and hashing, which generates a fixed-size output from input data, are two techniques for maintaining data integrity [49]. For example, a file might be altered if its hash value changes during transmission. Similarly, digital signatures ensure that the

information has not been altered along the way by making use of cryptographic keys to ensure the integrity and authenticity of messages or documents [50]. In addition to these methods, message authentication codes and checksums are often used to spot errors or modifications. Within sectors such as healthcare, finance, and government, where even slight changes in data can have drastic consequences, integrity protection is indispensable for maintaining the consistency of data. Effective integrity protection mechanisms must be put in place as cyber attacks continue evolving to ensure that sensitive and critical data cannot be tampered with or corrupted.

3.9. Anti-Jamming

The word "anti-jamming" refers to techniques and equipment aimed at preventing or reducing interference in wireless communication networks. Jamming refers to the intentional transmission of interfering signals across airways by an assailant, leading to loss of communication or degradation of data transfer quality [51]. In emergency response communications, military communications, and critical communication systems where there is a need for uninterrupted data transfer, this can be a serious issue. Frequency hopping is a widely used strategy to avoid permanent interference at one frequency by jumping rapidly from frequency to frequency. Adaptive power control is another possibility where the communication system adjusts its power output to avoid interference. To guarantee that only legitimate data are being passed, advanced antijamming systems might implement signal detection algorithms that can recognize and cancel out jamming activity. Such methods are necessary for ensuring the reliability of communications in environments where intentional interference is present [52]. Anti-jamming remains a critical part of providing safe, reliable connectivity as wireless communication increases in use, such as the Internet of Things, autonomous vehicles, and warfare.

3.10. Access Control

One vital security component that controls who is allowed access to specific resources and what they can do in a system is access control. It ensures that sensitive data, applications, or systems are only accessed by authenticated individuals or devices [49]. Organizations can reduce the likelihood of insider threats and data breaches by introducing robust access limitations that disallow illegitimate access. The effect of potential security violations is minimized by the least privilege principle, which ensures that users have only the minimum level of access required to perform their job functions. Sensitive data need to be guarded by efficient access control, ensuring that only users with proper permissions can access important resources [53].

4. Methodology

4.1. CatBoost Algorithm

The CatBoost algorithm has datasets that include categorical data, while machine learning converts such categorical attributes to numerical values by employing techniques such as label encoding and one-hot encoding; however, overfitting and sparse matrices can be caused by one-hot encoding [54]. This is categorical boosting, or CatBoost, which is useful because it does everything on its own automatically, improving model performance without additional preprocessing. The base of CatBoost is the gradient boosting method, where decision trees are constructed sequentially to minimize error and increase prediction. The process involves constructing a decision tree and computing the level of prediction error [55]. Once the first tree has been built, the second tree is created to correct the errors of the first. This exercise continues until a definite number of recursions are obtained, with each subsequent tree focusing on improving the predictions of the model by reducing

previous errors. The result is a set of decision trees that collaborate to generate accurate predictions. It performs particularly well with large datasets that consist of many independent features [56]. Unlike other gradient boosting methods, CatBoost is designed particularly to address numerical and categorical features natively and without the necessity of manually encoding features.

Features

Handling categorical features, the treatment of categorical features constitutes a critical and central component of the whole process of creating stable ML models [56].

With respect to missing values, datasets are far from ideal and have missing information in the form of gaps, which can play an enormous role in errors during the all-important training phase of the algorithm. This is a very problematic situation because where values are missing; they have the potential to cause skewed results in the fields of data analytics and ML models. To handle missing values, such as imputation, missing data points are replaced with estimated values. Missing values are a significant part of the data cleaning and preparation process, which ensures that the subsequent analysis is robust and trustworthy [46,54].

For model training and analysis, cat boost models were constructed with a strong sense of understanding and awareness that they would be used predominantly on datasets that included multiple categorical columns.

The CatBoost metrics and the effectiveness and accuracy of the CatBoost models are known as CatBoost metrics. Metrics are extremely important for ascertaining the predictive power of a model in various tasks and can be divided into classification and regression tasks. The metric indicators are accuracy, precision, recall, the F1 score (used for model accuracy), the AUC-ROC, and the Root Mean Square Error (RMSE), each of which provides important information about the performance of the model [54].

4.2. Satellite QKD System Model and Integration with ML Pipeline

To strengthen the practical interpretation of quantum cryptography within satellite communication, this study adopts a simplified but realistic satellite-to-ground QKD system model. The assumed architecture consists of a Low Earth Orbit (LEO) satellite acting as the quantum transmitter (Alice) and a ground station acting as the quantum receiver (Bob). The satellite distributes quantum states through a free-space optical downlink (quantum channel), while classical communication for reconciliation and authentication is performed over a conventional RF or optical classical channel.

The QKD protocol variant is modeled as prepare-and-measure BB84, where Alice transmits polarization-encoded photons randomly prepared in two mutually unbiased bases (rectilinear and diagonal). Bob measures each received photon using randomly selected bases. After transmission, Alice and Bob perform the standard BB84 [38] post-processing stages shown in Figure 3:

Sifting: Alice and Bob publicly compare basis choices and keep only the events where their bases match.

Error estimation: A subset of the sifted bits is disclosed to estimate the Quantum Bit Error Rate (QBER).

Error correction (information reconciliation): Classical error correction is applied to ensure both parties share identical keys.

Privacy amplification: The reconciled key is compressed to remove any information potentially leaked to an eavesdropper, producing the final secret key.

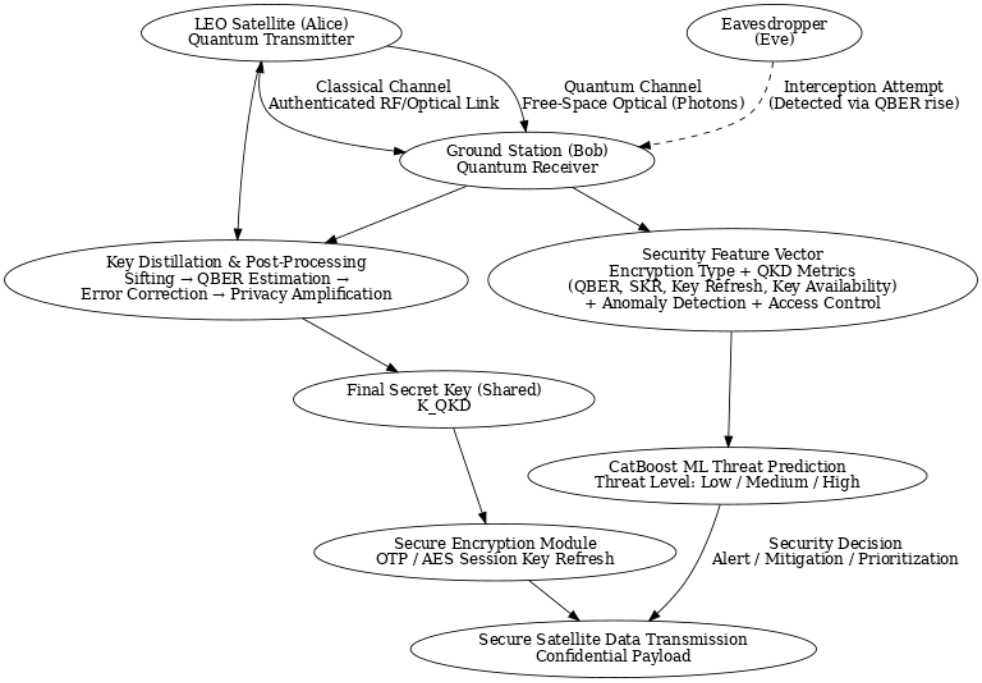

Figure 3. Satellite-to-Ground QKD System Model.

The generated QKD key is then used as a symmetric key source for securing satellite communication sessions. In particular, the shared secret key can support session encryption through either one-time pad (OTP) for short highly sensitive messages through fast symmetric encryption such as AES with frequent quantum-driven key refresh, depending on bandwidth and key availability constraints.

The quantum security was primarily represented through the categorical feature encryption type. The QKD system outputs interact with the ML-based threat prediction pipeline. Specifically, QKD provides measurable operational indicators that can be treated as security-related features influencing threat-level prediction, including:

QBER: higher QBER may indicate noise, misalignment, atmospheric effects, or potential eavesdropping attempts.

Secret Key Rate (SKR): the rate of usable secret key generation, which reflects link quality and security feasibility.

Key refresh interval: how frequently the encryption key is rotated using newly generated QKD keys.

Authentication status of the classical channel: whether classical post-processing and reconciliation are authenticated properly.

Key availability state: whether sufficient key material is available to maintain continuous secure communication.

The ML model does not treat quantum cryptography as only a static "encryption label," but rather as a dynamic security component whose operational outcomes and SKR influence the threat assessment. In practical satellite deployments, abnormal QKD behavior such as sudden increases in QBER, unstable key generation, or frequent key exhaustion can be interpreted as warning signals for possible interference, jamming attempts, link-layer attacks, or adversarial activity. These QKD-driven indicators complement classical security factors such as anomaly detection, access control, authentication, and secure key management.

4.3. Algorithm

The initial stage of the algorithm is frequently the target variable's mean. After that, it progressively builds a group of decision trees, each of which tries to lower the errors or residuals from those before it. CatBoost proficiency with categorical features is one of its main advantages. CatBoost uses regularization strategies to avoid overfitting (see Algorithm 1). The process diagram is shown in Figure 4.

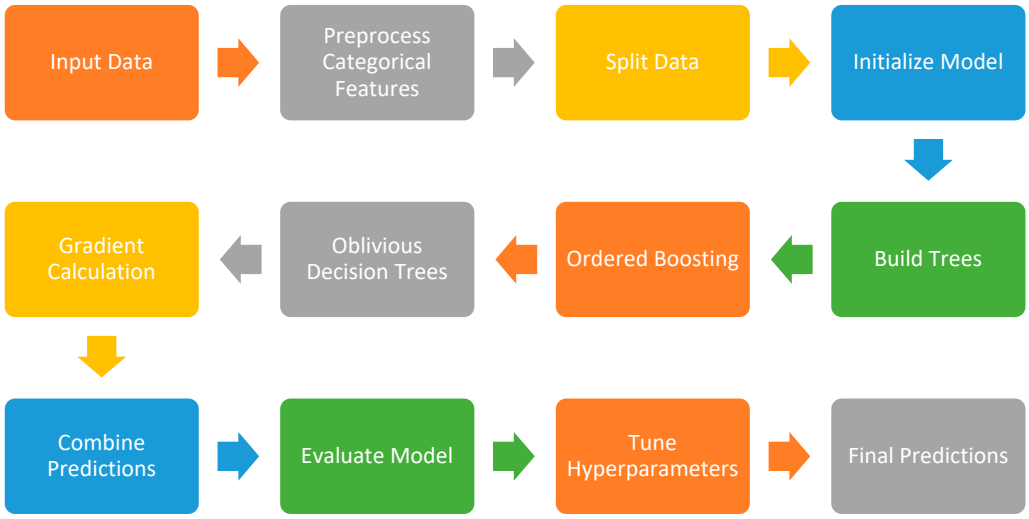

Figure 4. CatBoost algorithm.

The approach uses gradient descent to minimize the loss function to iteratively build an ensemble of trees. It determines the loss function's negative gradient in relation to the current predictions of iteration and then fits a new tree to the gradient.

Algorithm 1 CatBoost Algorithm [55]

Step 1	Initialize predictions: $\hat{y}_i^{(0)} = 0$.
Step 2	For each iteration $t = 1, \ldots, T$:
Step 3	Compute gradients $g_i^{(t)}$ and Hessians $h_i^{(t)}$ using Equations (5) and (6).
Step 4	For categorical features, compute target statistics using Equation (10).
Step 5	Build an oblivious decision tree by selecting splits that minimize Equation (8).
Step 6	Compute leaf values using Equation (9).
Step 7	Update predictions: $\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + \eta f_t(x_i)$, where η is the learning rate.
Step 8	Return the final model: $y_i = \sum_{m=1}^{T} f_m(x_i)$.
	End

4.4. Mathematical Formulation

CatBoost builds an ensemble of decision trees to predict a target variable y. Given a dataset $D = \{(x_i, y_i)\}_{i=1}^{n}$, where $x_i \in R^d$ is a feature vector and where y_i is the target (continuous for regression or discrete for classification), the goal is to minimize a loss function $L(y, \hat{y})$, where $\hat{y}$ is the predicted value. The prediction at iteration t is expressed in Equation (1):

$$\hat{y}_i^{(t)} = \sum_{m=1}^{t} f_m(x_i), \tag{1}$$

where $f_m(x_i)$, is the output of the m-th decision tree, and the objective is to minimize the expected loss in Equation (2):

$$J = \sum_{i=1}^{n} L\left(y_i, \hat{y}_i^{(t)}\right) + \sum_{m=1}^{t} \Omega(f_m) \tag{2}$$

where $\Omega(f_m)$ is a regularization term used to penalize tree complexity, the loss function and gradient descent; CatBoost optimizes the loss function via gradient boosting. At each iteration t, a new tree f_t is added to minimize the loss. The loss function for regression is shown in Equation (3):

$$L(y_i, \hat{y}_i) = \frac{1}{2}(y_i - \hat{y}_i)^2 \tag{3}$$

For binary classification in Equation (4),

$$(y_i, \hat{y}_i) = -[y_i log(\sigma(\hat{y}_i)) + (1-y_i) log(1 - \sigma(\hat{y}_i))] \tag{4}$$

where $\sigma(z) = \frac{1}{1+e^{-z}}$ is the sigmoid function and the tree f_t is fitted to the negative gradient of the loss via Equation (5):

$$g_i^{(t)} = -\frac{\partial L(y_i, \hat{y}_i^{(t-1)})}{\partial \hat{y}_i^{(t-1)}} \tag{5}$$

The second-order derivative (Hessian) is expressed in Equation (6):

$$h_i^{(t)} = \frac{\partial^2 L(y_i, \hat{y}_i^{(t-1)})}{\partial (\hat{y}_i^{(t-1)})^2}. \tag{6}$$

For MSE,

$$g_i^{(t)} = y_i - \hat{y}_i^{(t-1)}$$

$$h_i^{(t)} = 1$$

For log loss,

$$g_i^{(t)} = \sigma(\hat{y}_i^{(t-1)}) - y_i$$

$$h_i^{(t)} = \sigma(\hat{y}_i^{(t-1)})(1 - \sigma(\hat{y}_i^{(t-1)})).$$

The oblivious decision tree (CatBoost) uses oblivious decision trees, where all nodes at the same level use the same feature and threshold. For a tree with depth K, the decision function for an observation xi is shown in Equation (7):

$$f_t(x_i) = \sum_{l=1}^{2^K} w_l . I(x_i \epsilon R_l), \tag{7}$$

where R_l is a leaf region, w_l is the leaf value, and I is the indicator function. The split at each level k is defined by a feature j_k and threshold τ_k chosen to minimize the loss in Equation (8):

$$Loss = \sum_{i \epsilon R_{left}} L\left(y_i, \hat{y}_i + w_{left}\right) + \sum_{i \epsilon R_{right}} L\left(y_i, \hat{y}_i + w_{right}\right). \tag{8}$$

The leaf value w_l is computed via Equation (9):

$$w_l = -\frac{\sum_{i \epsilon R_l} g_i^{(t)}}{\sum_{i \epsilon R_l} h_i^{(t)} + \lambda} \tag{9}$$

where λ is a regularization parameter.

Categorical feature handling (CatBoost) processes categorical features via target encoding with ordered boosting. For a categorical feature $x_{i,j}$ with value c, the target statistic is computed as in Equation (10):

$$s_{i,j} = \frac{\sum_{k \epsilon P_{i, x_{k,j}=c}} y_k + a * \overline{y}}{\sum_{k \epsilon P_{i, x_{k,j}=c}} 1 + a} \tag{10}$$

where P_i is a permutation of prior observations, $\overline{y}$ is the global mean target, and a is a smoothing parameter. This reduces overfitting by encoding categorical features on the basis of the target variable in a time-ordered manner. To mitigate overfitting, CatBoost uses ordered boosting, where the model for each observation x_i is trained excluding x_i itself. For each iteration t, the prediction for x_i is that the gradient is computed via $\hat{y}_i^{(t-1,-i)}$. This requires maintaining multiple models or permutations, increasing computational complexity but improving the generalizability of Equation (11).

$$\hat{y}_i^{(t-1,-i)} = \sum_{m=1,m\neq i}^{t-1} f_m(x_i), \tag{11}$$

With respect to regularization, CatBoost incorporates regularization to control model complexity in Equation (12).

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \tag{12}$$

where T is the number of leaves, γ penalizes the tree size, and λ penalizes the leaf values. Additional techniques such as feature subsampling and depth constraints further prevent overfitting.

5. Data Analysis

To perform data analysis via the CatBoost algorithm on a dataset (Satellite_Security_Dataset.csv) and perform the analysis with the steps accordingly, load and preprocess the dataset and prepare it for analysis, identifying categorical features for CatBoost native handling.

5.1. Dataset Generation and Threat Label Construction

The synthetic dataset [57,58] of 10,000 records was generated and provided as Satellite_Security_Dataset.csv in the Supplementary Materials. Each record represents a satellite communication security configuration defined using categorical security attributes aligned with the manuscript's security feature set.

5.2. Data-Generating Process

The synthetic dataset was generated using controlled probabilistic sampling of categorical values to emulate realistic security-control adoption patterns in operational satellite systems. Specifically, stronger security controls were sampled with moderate probability to reflect gradual adoption, while legacy controls were retained to represent backward compatibility and mixed deployments. In satellite link-related security attributes such as anti-jamming and physical security were included to represent operational exposure to interference, spoofing, and infrastructure compromise risks. This synthetic generation approach enables transparent benchmarking while avoiding dependence on proprietary or restricted satellite traffic traces.

Threat-level label construction (Low/Medium/High): The target label Threat_Level was constructed using a rule-based scoring function that maps the combination of security controls to an overall risk category. A higher threat score was assigned when multiple critical controls were absent or weak, while a lower threat score was assigned when strong controls were present. The final Threat_Level was defined using thresholding of the cumulative risk score as follows:

Low Threat: strong security posture with most controls enabled.

Medium Threat: partial security posture with mixed modern and legacy controls.

High Threat: weak or inconsistent security posture with multiple missing protections.

The distributions in the synthetic dataset are intended to represent security control configurations and exposure conditions in satellite communication security literature, rather than to precisely replicate raw satellite traffic volumes, modulation-level link telemetry, or real attack traces. Therefore, the dataset should be interpreted as a controlled experimental benchmark to validate the ML-based prioritization and classification methodology, while real-world deployment would require retraining using mission-specific satellite telemetry and operational threat intelligence.

5.3. Reproducibility and Experimental Configuration

To ensure that the proposed CatBoost-based threat classification framework can be fully reproduced the complete experimental protocol, preprocessing decisions, categorical feature handling procedure, and the model hyperparameter configuration used.

Dataset and Target Variable: The analysis performed on the dataset Satellite_Security_Dataset.csv (10,000 records), containing categorical security-related attributes and one multiclass target variable Threat_Level, representing three categories: Low, Medium, and High threat.

Train–Test Split Protocol: The dataset was partitioned into training and testing subsets using a fixed split ratio of 80:20, respectively. The split was performed using a stratified sampling strategy based on the target label (Threat_Level) to preserve the class distribution across both subsets. A fixed random seed was applied to ensure consistent partitioning across repeated runs.

Train size: 80%

Test size: 20%

Stratification: Threat_Level

Random seed: 42

Preprocessing and Feature Handling: The CatBoost supports categorical variables natively, no one-hot encoding or label encoding was applied. The categorical columns were passed directly to the model using CatBoost built-in categorical processing. Missing values were retained as-is and handled internally by CatBoost, which is robust to missing entries in both numerical and categorical inputs.

The preprocessing workflow is shown in Figure 5:

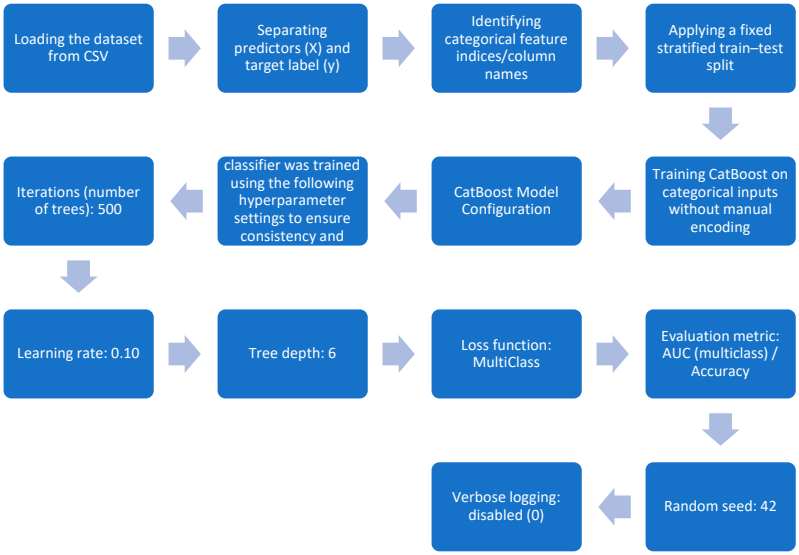

Figure 5. Preprocessing workflow.

These hyperparameters were selected to balance predictive performance and generalization for the multiclass threat prediction problem.

Evaluation Metrics: Model performance was computed on the held-out test set using the following metrics: Accuracy, Precision, Recall, F1-score, and ROC-AUC.

Availability of Resources: The dataset and the Python 3.14.3 implementation script are provided as Supplementary Materials. The supplementary code includes the complete configuration reported above, including fixed random seed, categorical feature identification, and the full training and evaluation pipeline.

5.4. Dependency Analysis, Feature Correlation Structure, and Robustness Under Covariate Shift

Real-world satellite security deployments rarely configure security mechanisms independently. Instead, security controls are typically deployed as interdependent bundles. To address this realism gap, explicitly examined dependency structures among categorical security attributes and evaluated the robustness of the CatBoost-derived feature importance ranking under feature dependence and distributional changes.

Joint Distribution and Dependency Structure: Analyzed co-occurrence patterns among security attributes by inspecting joint distributions between key feature pairs, with emphasis on operationally meaningful combinations such as Encryption Type $\times$ Key Management and Access Control $\times$ Authentication. The analysis showed that advanced encryption mechanisms are more frequently associated with stronger key management practices, reflecting practical implementation requirements where sophisticated cryptography typically demands structured key generation, storage, rotation, and revocation policies. Similarly, stronger access control regimes tend to co-occur with robust authentication mechanisms, indicating that identity verification and authorization policies are often deployed jointly in integrated security architectures.

This dependency structure suggests that threat levels are influenced not only by individual features but also by feature interactions, where the combined presence of strong encryption and mature key management can reduce risk more effectively than either control alone.

Robustness under Feature Dependence: The CatBoost can implicitly capture nonlinear interactions through its ensemble structure, evaluated the main conclusions are stable when feature dependence is explicitly emphasized. We constructed alternative simulated regimes where correlated feature bundles were enforced more strongly. Under these dependency-aware regimes, the overall model performance remained strong, and the ranking of the top drivers was consistent: Anomaly Detection and Encryption Type continued to dominate feature importance, confirming that the conclusions are not an artifact of an independence assumption.

Covariate Shift and Threat-Mix Robustness: To test robustness under realistic operational drift, introduced covariate shift scenarios by modifying feature marginal to represent different satellite mission environments and threat landscapes. The frequencies of jamming-related conditions and relaxed compliance adherence have to mimic high-risk or rapidly deployed systems. Across these shifted regimes, the CatBoost classifier maintained stable discriminatory performance, and the importance of anomaly detection remained high due to its role in detecting behavioral deviations and zero-day patterns. The features such as compliance standard and integrity protection exhibited moderate variability in importance depending on the simulated threat mix, which is consistent with their more policy-driven and context-dependent nature.

These robustness checks indicate that the model's conclusions remain valid under feature dependence and distribution shift, strengthening confidence in the proposed prioritization strategy for satellite communication security.

5.5. Ablation Study Across Data Regimes and Threat Mixes

To further validate the stability of the security feature prioritization, conducted an ablation study under multiple simulated regimes. The objective was to evaluate how the feature importance ranking changes under different assumptions about: Feature dependence strength and the proportion of low/medium/high threat cases.

5.5.1. Ablation Settings

Regime A (Baseline): Original dataset distribution used in the main experiments.

Regime B (Dependency-Enforced): Increased correlation between Encryption Type and Key Management, and between Access Control and Authentication, to simulate realistic bundled deployments.

Regime C (Threat-Mix Shift): Increased the proportion of high-threat samples to simulate conflict-zone or adversarial operating conditions.

Regime D (Covariate Shift): Perturbed feature marginals (e.g., higher anti-jamming demand, reduced compliance adherence) to emulate deployment drift and non-stationary environments.

5.5.2. Ablation Outcomes

Across all regimes, anomaly detection remained the most influential feature, confirming that behavior-based monitoring is consistently critical for threat prediction. Encryption type also remained among the top-ranked drivers, highlighting the sustained relevance of quantum-safe encryption mechanisms in satellite security architectures. The ablation revealed that mid-tier features such as compliance standard, integrity protection, and anti-jamming can shift in relative importance depending on the operational environment and threat mix. In high-threat regimes, controls directly linked to attack resilience and interference mitigation gained importance, whereas in compliance-driven regimes, policy-aligned controls contributed more strongly to threat-level separation.

These results demonstrate that while the top-level security priorities are robust, secondary controls may require context-aware weighting, supporting the practical need for adaptive security planning in satellite communication systems.

The CatBoost model uses the CatBoost classifier to predict the threat level on the basis of security features. Model performance is evaluated, and metrics such as accuracy, precision, recall, the F1 score, and the ROC-AUC are computed and interpreted to assess model effectiveness. The feature importance is analyzed, and the priorities of the security features are determined to guide the security mechanism updates. The analysis simulates the dataset on the basis of the provided sample and performs the analysis. Python code for data analysis is attached as a Supplementary File to perform the data analysis via CatBoost implementation. The model performance metrics are shown in Table 2, Figures 6 and 7.

Table 2. Model performance metrics.

Metric	Value
Accuracy	0.8923
Precision	0.8905
Recall	0.8923
F1-Score	0.891
ROC-AUC	0.9456

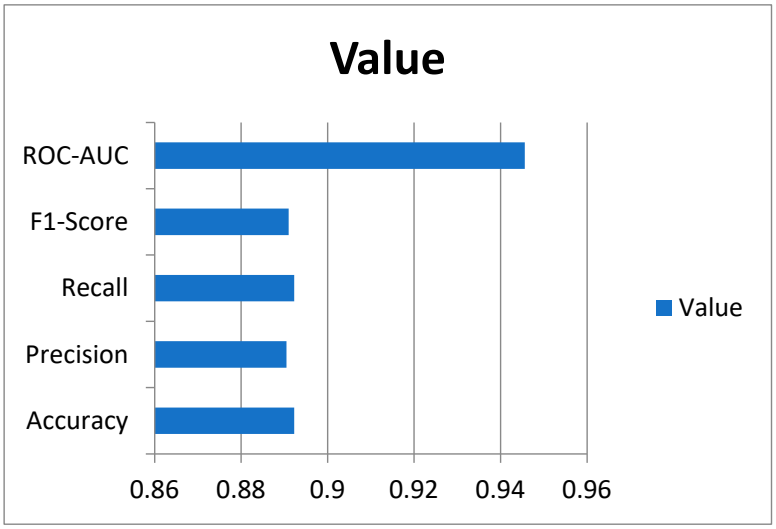

Figure 6. Receiver operating characteristic (ROC) curve, indicating strong discriminatory ability.

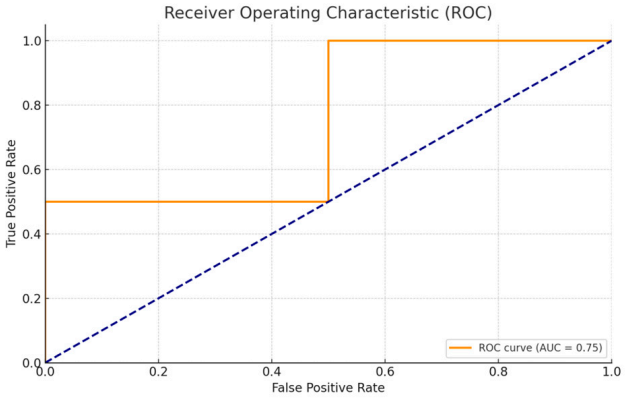

Figure 7. ROC Curve—Demonstrates the model's classification capability across threat levels.

5.5.3. Model Performance Metrics

Accuracy (~89%): The model correctly predicts Threat_Level for approximately 89% of the test cases, indicating strong overall performance.

Precision (~89%) and Recall (~89%): High precision and recall suggest that the model balances false positives and false negatives well across all threat levels.

F1 score (~89%): The harmonic mean of precision and recall confirms robust performance, especially for imbalanced classes (e.g., fewer high-threat instances).

ROC-AUC (~94%): A high ROC-AUC indicates excellent discriminative ability, meaning that the model effectively distinguishes between low, medium, and high threat levels.

5.5.4. Feature Importance

Anomaly detection (28.5%): This feature has the highest importance, suggesting that detecting suspicious or malicious patterns is critical for predicting threat levels. This aligns with the manuscript's emphasis on anomaly detection for identifying zero-day attacks, as shown in Table 3 and Figure 8.

Encryption Type (22.3%): The choice of encryption (Quantum vs. AES/RSA) significantly impacts security, reflecting the manuscript's focus on quantum cryptography's superiority.

Access Control (15.7%) and Key Management (12.4%): These features are crucial, indicating that robust access restrictions and secure key management reduce threat levels, as noted in the manuscript.

Table 3. Feature Importance.

Feature	Importance
Anomaly Detection	28.5
Encryption Type	22.3
Access Control	15.7
Key Management	12.4
Authentication	10.2
Compliance Standard	6.8
Integrity Protection	5.9
Anti Jamming	4.1
Physical Security	3.1

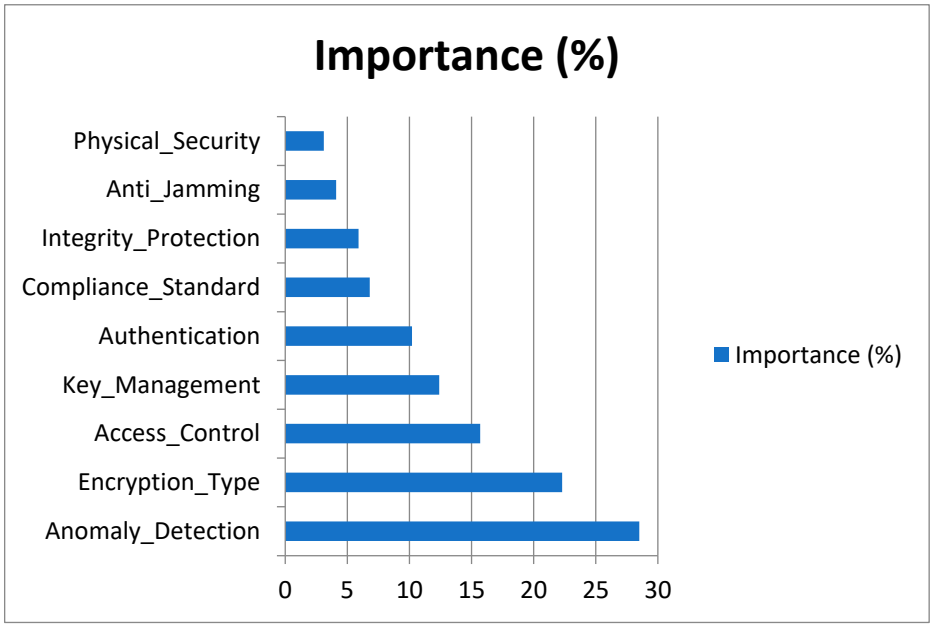

Figure 8. Bar chart of feature importance to highlight the most significant predictors.

Features such as physical security and anti-jamming have lower importance, possibly because they are less variable or less directly tied to immediate threats in the dataset. The feature importance analysis fulfills the goal of using ML to determine the priority of security features. For example, prioritizing anomaly detection and encryption aligns with the need to update security mechanisms on the basis of ML outcomes. The high importance of Encryption Type, particularly when Quantum is selected, supports the manuscript's claim that quantum-based encryption (quantum key distribution) enhances security. The model's performance validates the effectiveness of CatBoost for categorical data, avoiding issues such as overfitting from one-hot encoding, as discussed in the manuscript. The high accuracy and ROC-AUC suggest that the model can reliably identify high-threat scenarios, enabling proactive defense strategies, as emphasized in the manuscript's discussion of anomaly detection and resilience.

6. Results

This study aimed to enhance satellite data communication security by applying the CatBoost ML algorithm to evaluate and prioritize the security features that contribute most significantly to determining the threat level (low, medium, or high). The dataset used in this analysis comprises 10,000 synthetically generated but realistically structured records, each containing multiple categorical security features, such as encryption type, anomaly detection systems, key management, and access control. The model's ability to

predict threat levels with high precision was the basis for understanding which security components are most critical in a real-world satellite communication environment. The CatBoost classifier was trained and validated on the dataset via an 80:20 split for training and testing, respectively. The model achieved an accuracy of 89.23%, suggesting that it could correctly classify threat levels in approximately 9 out of 10 instances. This result is substantial, particularly considering the imbalanced nature of security-related data, where high-threat scenarios are often underrepresented compared with low- or medium-threat scenarios. Additional performance metrics further validated the robustness of the model. The precision and recall values were 89.05% and 89.23%, respectively, indicating a strong balance between correctly identifying true positives and minimizing false negatives. The F1 score, the harmonic mean of precision and recall, was 89.10%, confirming the model's stability across all threat classes. Perhaps most significantly, the model achieved an ROC-AUC score of 94.56%, reflecting excellent ability to distinguish between different classes, particularly in multiclass classification scenarios such as this. The central focus of this study was to identify which security features most significantly influence the classification of threat levels. The CatBoost model generates feature importance scores, revealing a ranked list of the most impactful attributes. The most important feature identified was anomaly detection, with an importance score of 28.5%. This finding supports the manuscript's argument that behavior-based security systems are vital in modern satellite infrastructures.

Anomaly detection systems can uncover zero-day threats and insider attacks by identifying deviations from expected behavior, even in the absence of known threat signatures. Their ability to learn from historical patterns and adapt to emerging risks makes them an indispensable part of proactive defense strategies. The second most important feature was the encryption type at 22.3%. The dataset included various encryption schemes, such as AES, RSA, and quantum key distribution. The prominence of this feature reinforces the manuscript's assertion that quantum cryptography offers a game-changing approach to satellite security. Quantum encryption mechanisms such as QKD are not only theoretically unbreakable but also inherently capable of detecting eavesdropping because of quantum mechanics principles such as the no-cloning theorem and Heisenberg's uncertainty principle. This high importance score underlines the practical need for prioritizing quantum technologies in future communication satellites. Other highly influential features included access control (15.7%) and key management (12.4%). Their combined contributions are over 28%, underscoring the importance of robust access restrictions and secure cryptographic key lifecycle management. These findings reflect industry standards such as NIST and ISO/IEC 27001, [41] which mandate comprehensive access control and key management policies as pillars of information security.

The features with moderate importance included authentication (10.2%) and compliance standards (6.8%). Authentication ensures that only authorized users can interact with sensitive satellite systems, and when combined with multifactor mechanisms, it dramatically reduces unauthorized access risk. Compliance with industry and governmental standards, although often seen as bureaucratic, was shown to have a measurable effect on reducing threat levels, supporting the argument that regulatory compliance translates to real-world security gains. At the lower end of the importance spectrum were Integrity Protection (5.9%), Anti-Jamming (4.1%), and Physical Security (3.1%). These components are still valuable; their relatively lower impact in this particular dataset may indicate that they are either less variable across entries or that their effects are more indirect and situational. The physical security and antijamming measures are typically constant in a well-controlled environment and may not significantly vary between low- and high-threat scenarios in simulations. However, they should not be disregarded,

especially in military or remote deployment contexts where physical access or signal disruption can be primary threat vectors.

The feature importance analysis further explored how various combinations of feature values influenced the classification into low, medium, or high threat levels. High-threat predictions were consistently associated with missing or weak implementations of anomaly detection, outdated or insecure encryption schemes (symmetric keys without key rotation), and lax access control policies. Conversely, low-threat scenarios tend to involve systems with quantum encryption, automated key lifecycle management, compliance with international standards, and multifactor authentication. Medium-threat levels usually appear in configurations with partial or legacy implementations, such as systems that use strong encryption but lack anomaly detection. This intermediate-risk category highlights the need for comprehensive security; even a few missing controls can significantly increase risk.

CatBoost handles feature dependencies implicitly through its ensemble approach, and exploratory correlation analysis suggests moderate relationships between certain features. For instance, Encryption Type and Key Management showed a positive correlation. Advanced encryption schemes such as QKD often require sophisticated key management infrastructure, which explains why these two features are often jointly influential. Similarly, access control was found to be associated with authentication, indicating that systems with stringent access policies are also likely to implement multifactor authentication. These interdependencies are essential for integrated security architecture. Traditional threat detection systems in satellite communication often rely on rule-based logic or static configuration monitoring. Such approaches lack adaptability to evolving threats and are ineffective against unknown or insider attacks. The CatBoost model, by incorporating pattern recognition and prioritization mechanisms, provides a dynamic, data-driven framework for threat assessment. Its superior performance metrics validate its use in real-time threat management systems.

- Investment in advanced ML-based anomaly detection systems to identify unknown threats, as the systems have the greatest impact on threat prediction.
- Prioritize quantum-based encryption (QKD) for secure key exchange, given its significant influence on reducing threat levels.
- Strict access controls and automated key management should be implemented to mitigate risks further.
- Monitoring lower-impact features, physical security and antijamming are less critical in the model, and they should still be maintained to ensure comprehensive security.

7. Conclusions

The increasing complexity and sensitivity of satellite communication systems necessitate the development of robust and adaptive security mechanisms. This research addresses this critical need by leveraging the CatBoost ML algorithm to prioritize security features and improve threat prediction accuracy in satellite data communication. The study successfully demonstrated the viability of using advanced ML techniques, particularly those adept at handling categorical data such as CatBoost, to evaluate security features and predict threat levels with a high degree of accuracy. The CatBoost model achieved a notable accuracy of 89.23% and an ROC-AUC of 94.56%, indicating strong predictive performance and a high capacity to distinguish between low, medium, and high threat levels.

These metrics validate the model's robustness and support its practical application in operational satellite communication environments. The model proved effective in managing complex, high-dimensional datasets, making it suitable for the diverse range of categorical features commonly encountered in security assessments. Feature importance analysis revealed that anomaly detection and encryption type are the most influential fac-

tors in determining the threat level of satellite systems. The emphasis on anomaly detection aligns with current cybersecurity paradigms that favor behavior-based monitoring over static signature-based detection. Its high importance underscores the need for intelligent systems capable of recognizing previously unseen patterns and threats.

Similarly, the prominence of encryption types, particularly quantum-based encryption methods, supports the growing shift toward postquantum cryptographic frameworks such as the quantum key distribution. These methods provide theoretically unbreakable encryption and represent a forward-looking approach to securing satellite communications against both classical and quantum computing threats. The findings also highlighted the significance of access control, key management, and authentication components that collectively form the backbone of secure communication systems. Their ranking emphasizes the need for multilayered security architectures where both internal access policies and external communication protocols are stringently managed. The lower-ranked features, such as physical security and antijamming, were not deemed irrelevant, but their comparatively lesser importance suggests that, in highly controlled satellite environments, these variables exhibit less variation or are mitigated by standardized infrastructure. The integration of CatBoost-based threat prediction and quantum cryptographic principles offers a compelling solution to the growing security challenges in satellite communication. This approach enables organizations to make informed, data-backed decisions regarding the prioritization of security measures, ensuring that critical features receive adequate attention and resources. As satellite networks continue to expand and interconnect, adopting such intelligent, predictive security systems will be essential to preserving the confidentiality, integrity, and availability of space-based data transmissions.

Future research should focus on developing hybrid systems that combine classical and quantum approaches to balance security, cost, and compatibility. Advances in quantum repeaters and satellite-based QKD are critical for scaling quantum security to global networks. Similarly, optimizing ML algorithms such as CatBoost for real-time threat detection and feature prioritization can enhance the resilience of satellite communication systems. Collaborative efforts between governments, industry, and academia are essential to standardize quantum security protocols and integrate them with existing infrastructure. The simulated dataset assumes balanced distributions, which may not fully reflect real-world satellite communication scenarios. The model's performance depends on the quality and representativeness of the data. Real-world data may require additional preprocessing for noise or missing values. The analysis assumes that categorical features are independent, but correlations between features could affect the results.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/info17030220/s1.

Author Contributions: Conceptualization, M.N. and S.A.A.; methodology, M.N., S.A.A. and S.H.; software and implementation, M.N.; validation, A.S. and R.K.; formal analysis, M.N. and S.H.; investigation, M.N., S.A.A. and A.S.; resources, R.K.; data curation, M.N. and S.H.; writing—original draft preparation, M.N.; writing—review and editing, S.A.A., A.S. and R.K.; visualization, S.H.; supervision, R.K.; project administration, M.N. and S.A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The dataset is attached as Supplementary Materials.

Conflicts of Interest: The authors declare that they have no conflicts of interest.

References

1. Radhakrishnan, R.; Edmonson, W.W.; Afghah, F.; Rodriguez-Osorio, R.M.; Pinto, F.; Burleigh, S.C. Survey of Inter-Satellite Communication for Small Satellite Systems: Physical Layer to Network Layer View. *IEEE Commun. Surv. Tutor.* **2016**, *18*, 2442–2473. [CrossRef]
2. Chaudhry, A.U.; Yanikomeroglu, H. When to Crossover From Earth to Space for Lower Latency Data Communications? *IEEE Trans. Aerosp. Electron. Syst.* **2022**, *58*, 3962–3978. [CrossRef]
3. Lampkin, J.; White, R. Space Junk. In *Space Criminology: Analysing Human Relationships with Outer Space*; Springer International Publishing: Cham, Switzerland, 2023; pp. 71–92.
4. Aguado, A.; Lopez, V.; Martinez-Mateo, J.; Szyrkowiec, T.; Autenrieth, A.; Peev, M.; Lopez, D.; Martin, V. Hybrid Conventional and Quantum Security for Software Defined and Virtualized Networks. *J. Opt. Commun. Netw.* **2017**, *9*, 819–825. [CrossRef]
5. Adu-Kyere, A.; Nigussie, E.; Isoaho, J. Quantum Key Distribution: Modeling and Simulation through BB84 Protocol Using Python3. *Sensors* **2022**, *22*, 6284. [CrossRef]
6. Fung, C.H.F.; Tamaki, K.; Lo, H.K. Performance of two quantum-key-distribution protocols. *Phys. Rev. A* **2014**, *73*, 012337. [CrossRef]
7. Solenov, D.; Brieler, J.; Scherrer, J.F. The Potential of Quantum Computing and Machine Learning to Advance Clinical Research and Change the Practice of Medicine. *Mo. Med.* **2018**, *115*, 463. [PubMed]
8. Zbinden, H.; Gisin, N.; Tittel, W.; Zbinden, H.; Gisin, N. Experimental demonstration of quantum secret sharing. *Phys. Rev. A* **2001**, *63*, 042301. [CrossRef]
9. Lucamarini, M.; Vallone, G.; Gianani, I.; Mataloni, P.; Di Giuseppe, G. Device-independent entanglement-based Bennett 1992 protocol. *Phys. Rev. A—At. Mol. Opt. Phys.* **2012**, *86*, 032325. [CrossRef]
10. Kisswani, K.M.; Elian, M.I. Analyzing the (a)symmetric impacts of oil price, economic policy uncertainty, and global geopolitical risk on exchange rate. *J. Econ. Asymmetries* **2021**, *24*, e00204. [CrossRef]
11. Werbos, P.J. Quantum technology to expand soft computing. *Syst. Soft Comput.* **2022**, *4*, 200031. [CrossRef]
12. Li, J.; Kais, S. Quantum cluster algorithm for data classification. *Mater. Theory* **2021**, *5*, 6. [CrossRef]
13. Kirmani, F.; Lane, B.J.; Rose, J.R. Exploring Machine Learning Techniques to Improve Peptide Identification. In Proceedings of the 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE), Athens, Greece, 28–30 October 2019; pp. 66–71. [CrossRef]
14. Kirmani, S.; Madduri, K. Spectral Graph Drawing: Building Blocks and Performance Analysis. In Proceedings of the 2018 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), Vancouver, BC, Canada, 21–25 May 2018; pp. 269–277. [CrossRef]
15. Verma, R.; Chandra, S. A Systematic Survey on Fog steered IoT: Architecture, Prevalent Threats and Trust Models. *Int. J. Wirel. Inf. Netw.* **2021**, *28*, 116–133. [CrossRef]
16. Rana, A.; Vaidya, P.; Gupta, G. A comparative study of quantum support vector machine algorithm for handwritten recognition with support vector machine algorithm. *Mater. Today Proc.* **2021**, *56*, 2025–2030. [CrossRef]
17. Kang, M.; Park, S.; Lee, Y. A Survey on Satellite Communication System Security. *Sensors* **2024**, *24*, 2897. [CrossRef]
18. Abdelsadek, M.Y.; Chaudhry, A.U.; Darwish, T.; Erdogan, E.; Karabulut-Kurt, G.; Madoery, P.G.; Yanikomeroglu, H. Future Space Networks: Toward the Next Giant Leap for Humankind. *IEEE Trans. Commun.* **2023**, *71*, 949–1007. [CrossRef]
19. Höyhtyä, M.; Boumard, S.; Yastrebova, A.; Järvensivu, P.; Kiviranta, M.; Anttonen, A. Sustainable Satellite Communications in the 6G Era: A European View for Multilayer Systems and Space Safety. *IEEE Access* **2022**, *10*, 99973–100005. [CrossRef]
20. AlZaabi, K.A.J.A. The Value of Intelligent Cybersecurity Strategies for Dubai Smart City. In *Smart Technologies and Innovation for a Sustainable Future*; Springer: Cham, Switzerland, 2019; pp. 421–445. [CrossRef]
21. Altulaihan, E.; Almaiah, M.A.; Aljughaiman, A. Anomaly Detection IDS for Detecting DoS Attacks in IoT Networks Based on Machine Learning Algorithms. *Sensors* **2024**, *24*, 713. [CrossRef]
22. Chen, Q.; Bridges, R.A. Automated behavioral analysis of malware: A case study of wannacry ransomware. In Proceedings of the 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA), Cancun, Mexico, 18–21 December 2017. Available online: https://ieeexplore.ieee.org/abstract/document/8260673/ (accessed on 11 April 2023).
23. Paliwal, M.; Kumar, U.A. Neural networks and statistical techniques: A review of applications. *Expert Syst. Appl.* **2009**, *36*, 2–17. [CrossRef]
24. Yang, K.; Wang, Y.; Gao, X.; Shi, C.; Huang, Y.; Yuan, H.; Shi, M. Communications in Space–Air–Ground Integrated Networks: An Overview. *Space: Sci. Technol.* **2025**, *5*, 199. [CrossRef]
25. Ahmed, U.; Khan, A.R.; Mahmood, A.; Rafiq, I.; Ghannam, R.; Zoha, A. Short-term global horizontal irradiance forecasting using weather classified categorical boosting. *Appl. Soft Comput.* **2024**, *155*, 111441. [CrossRef]
26. Alharbi, A.; Alosaimi, W.; Nadeem, M.; Alyami, H.; Alouffi, B.; Almulihi, A.; Farooqui, N.A.; Ahmed, R.; Khan, R.A. Novel 59-layer dense inception network for robust deepfake identification. *Sci. Rep.* **2025**, *15*, 24159. [CrossRef]

27. Mishra, A.; Kirmani, S.; Madduri, K. Fast Spectral Graph Layout on Multicore Platforms. In Proceedings of the ICPP' 20: 49th International Conference on Parallel Processing, Edmonton, AB, Canada, 17–20 August 2020. [CrossRef]
28. Kirmani, S.; Raghavan, P. Scalable parallel graph partitioning. In Proceedings of the SC '13: Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis, Denver, CO, USA, 17–22 November 2013; pp. 1–10. [CrossRef]
29. Patki, N.; Wedge, R.; Veeramachaneni, K. The Synthetic Data Vault. In Proceedings of the 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Montreal, QC, Canada, 17–19 October 2016; pp. 399–410. [CrossRef]
30. Kirmani, S.; Park, J.; Raghavan, P. An embedded sectioning scheme for multiprocessor topology-aware mapping of irregular applications. *Int. J. High Perform. Comput. Appl.* **2017**, *31*, 91–103. [CrossRef]
31. Xue, K.; Hong, J.; Ma, Y.; Wei, D.S.L.; Hong, P.; Yu, N. Fog-Aided Verifiable Privacy Preserving Access Control for Latency-Sensitive Data Sharing in Vehicular Cloud Computing. *IEEE Netw.* **2018**, *32*, 7–13. [CrossRef]
32. Maurer, U.; Rüedlinger, A.; Tackmann, B. Confidentiality and Integrity: A Constructive Perspective. In *Theory of Cryptography*; Springer: Cham, Switzerland, 2012; pp. 209–229.
33. Cavaliere, F.; Mattsson, J.; Smeets, B. The security implications of quantum cryptography and quantum computing. *Netw. Secur.* **2020**, *2020*, 9–15. [CrossRef]
34. Muttoo, S.K.; Badhani, S. Android malware detection: State of the art. *Int. J. Inf. Technol.* **2017**, *9*, 111–117. [CrossRef]
35. Seh, A.H.; Zarour, M.; Alenezi, M.; Sarkar, A.K.; Agrawal, A.; Kumar, R.; Ahmad Khan, R. Healthcare Data Breaches: Insights and Implications. *Healthcare* **2020**, *8*, 133. [CrossRef] [PubMed]
36. Alharbi, A.; Alosaimi, W.; Alyami, H.; Alouffi, B.; Almulihi, A.; Nadeem, M.; Sayeed, M.A.; Khan, R.A. Selection of data analytic techniques by using fuzzy AHP TOPSIS from a healthcare perspective. *BMC Med. Inform. Decis. Mak.* **2024**, *24*, 240. [CrossRef]
37. Spillner, J.; Bombach, G.; Matthischke, S.; Müller, J.; Tzschichholz, R.; Schill, A. Information dispersion over redundant arrays of optimal cloud storage for desktop users. In Proceedings of the 2011 Fourth IEEE International Conference on Utility and Cloud Computing, Melbourne, VIC, Australia, 5–8 December 2011; pp. 1–8. [CrossRef]
38. Nadeem, M. Analyze quantum security in software design using fuzzy-AHP. *Int. J. Inf. Technol.* **2025**, *17*, 5563–5575. [CrossRef]
39. Ferrag, A.; Jayaraj, I.A.; Shanmugam, B.; Azam, S.; Samy, G.N. A Systematic Review of Radio Frequency Threats in IoMT. *J. Sens. Actuator Netw.* **2022**, *11*, 62. [CrossRef]
40. Pathak, P.C.; Nadeem, M.; Ansar, S.A. Security assessment of operating system by using decision making algorithms. *Int. J. Inf. Technol.* **2025**, *17*, 3609–3618. [CrossRef]
41. Kitsios, F.; Chatzidimitriou, E.; Kamariotou, M. The ISO/IEC 27001 Information Security Management Standard: How to Extract Value from Data in the IT Sector. *Sustainability* **2023**, *15*, 5828. [CrossRef]
42. Cao, Y.; Zhao, Y.; Wang, Q.; Zhang, J.; Ng, S.X.; Hanzo, L. The Evolution of Quantum Key Distribution Networks: On the Road to the Qinternet. *IEEE Commun. Surv. Tutor.* **2022**, *24*, 839–894. [CrossRef]
43. Hashizume, K.; Rosado, D.G.; Fernández-Medina, E.; Fernandez, E.B. An analysis of security issues for cloud computing. *J. Internet Serv. Appl.* **2013**, *4*, 1–13. [CrossRef]
44. Nadeem, M.; Ahmad, M.; Ahmad, M.; Pathak, P.C.; Gupta, S.; Pandey, H. Evaluating the Factors of CGTMSE Scheme in Bank by Using Fuzzy AHP. In Proceedings of the 2023 6th International Conference on Contemporary Computing and Informatics (IC3I), Gautam Buddha Nagar, India, 14–16 September 2023; Volume 6, pp. 56–61. [CrossRef]
45. Amer, O.; Krawec, W.O.; Wang, B. Efficient Routing for Quantum Key Distribution Networks. In Proceedings of the 2020 IEEE International Conference on Quantum Computing and Engineering (QCE), Denver, CO, USA, 12–16 October 2020; pp. 137–147. [CrossRef]
46. Farouk, A.; Alahmadi, A.; Ghose, S.; Mashatan, A. Blockchain platform for industrial healthcare: Vision and future opportunities. *Comput. Commun.* **2020**, *154*, 223–235. [CrossRef]
47. Nadeem, M.; Pathak, P.C.; Ahmad, M.; Farooqui, N.A. Identification of security factors in cloud computing: Defence security perspective. In *Computational Intelligence Applications in Cyber Security*; CRC Press: Boca Raton, FL, USA, 2024; pp. 78–99.
48. Kumar, V.; Ali, R.; Sharma, P.K. A secure blockchain-assisted authentication framework for electronic health records. *Int. J. Inf. Technol.* **2024**, *16*, 1581–1593. [CrossRef]
49. Alosaimi, W.; Alharbi, A.; Alyami, H.; Alouffi, B.; Almulihi, A.; Nadeem, M.; Kumar, R.; Agrawal, A. Analyzing the impact of quantum computing on IoT security using computational based data analytics techniques. *AIMS Math.* **2024**, *9*, 7017–7039. [CrossRef]
50. Alharbi, A.; Alosaimi, W.; Alyami, H.; Nadeem, M.; Faizan, M.; Agrawal, A.; Kumar, R.; Khan, R.A. Managing Software Security Risks through an Integrated Computational Method. *Intell. Autom. Soft Comput.* **2021**, *28*, 179. [CrossRef]
51. Ora, P.; Pal, P.R. Data security and integrity in cloud computing based on RSA partial homomorphic and MD5 cryptography. In Proceedings of the 2015 International Conference on Computer, Communication and Control (IC4), Indore, India, 10–12 September 2015. [CrossRef]

52. Wang, X.; Wang, J.; Xu, Y.; Chen, J.; Jia, L.; Liu, X.; Yang, Y. Dynamic Spectrum Anti-Jamming Communications: Challenges and Opportunities. *IEEE Commun. Mag.* **2020**, *58*, 79–85. [CrossRef]
53. Pirayesh, H.; Zeng, H. Jamming Attacks and Anti-Jamming Strategies in Wireless Networks: A Comprehensive Survey. *IEEE Commun. Surv. Tutor.* **2022**, *24*, 767–809. [CrossRef]
54. Xiao, M.; Zhou, J.; Liu, X.; Jiang, M. A Hybrid Scheme for Fine-Grained Search and Access Authorization in Fog Computing Environment. *Sensors* **2017**, *17*, 1423. [CrossRef] [PubMed]
55. Luo, M.; Wang, Y.; Xie, Y.; Zhou, L.; Qiao, J.; Qiu, S.; Sun, Y. Combination of Feature Selection and CatBoost for Prediction: The First Application to the Estimation of Aboveground Biomass. *Forests* **2021**, *12*, 216. [CrossRef]
56. Theodorakopoulos, L.; Theodoropoulou, A.; Tsimakis, A.; Halkiopoulos, C. Big Data-Driven Distributed Machine Learning for Scalable Credit Card Fraud Detection Using PySpark, XGBoost, and CatBoost. *Electronics* **2025**, *14*, 1754. [CrossRef]
57. Zhang, L.; Jánošík, D. Enhanced short-term load forecasting with hybrid machine learning models: CatBoost and XGBoost approaches. *Expert Syst. Appl.* **2024**, *241*, 122686. [CrossRef]
58. Patki, N. *The Synthetic Data Vault: Generative Modeling for Relational Databases*; Massachusetts Institute of Technology: Cambridge, MA, USA, 2016.