Applied Data Science and Analysis
Vol.2025, **pp**. 187–200
DOI: https://doi.org/10.58496/ADSA/2025/016 ; ISSN: 3005-317X
https://mesopotamian.press/journals/index.php/ADSA

Research Article

Data Classification and Emerging Encryption Technologies in Big Data and Cloud Computing : A Systematic Review

Karthik Kumar Vaigandla[1],*,

[1] *Electronics and Communication Engineering, Balaji Institute of Technology and Science, Warangal, Telangana, India.*

ARTICLE INFO

Article History

Received 20 Jul 2025
Revised 10 Aug 2025
Accepted 28 Aug 2025
Published 30 Sep 2025

Keywords

Data classification

data security

encryption

big data

cloud computing

ABSTRACT

This article provides a comprehensive review of recent advancements in data classification methods within the context of big data and cloud computing. As organizations increasingly rely on massive volumes of digital information, robust classification techniques have become essential, particularly for handling sensitive or confidential data. The study explores key approaches, including automated document classification and encryption-based strategies, each addressing distinct challenges related to security and efficiency. Emphasis is placed on how these methods safeguard data confidentiality, integrity, and availability critical factors in mitigating unauthorized access and cyber threats. The review also identifies pressing research gaps, such as the need for more scalable, efficient, and user-friendly classification systems that can adapt to the evolving nature of big data. The objective is to provide an in-depth overview of current practices, highlight persistent challenges, and outline promising directions for future research in this crucial field.

1. INTRODUCTION

Organizations face the challenge of managing vast volumes of data due to digitization. Data privacy regulations exist to protect sensitive information; however, manually identifying private documents is impractical. Therefore, systems that can efficiently classify documents and detect confidential content are essential [1], ensuring security and confidentiality. Traditional security solutions often struggle with large datasets exceeding database capacities. Sensitive data attracts threats that can harm an organization's reputation and stakeholder trust. Big data analysis, such as examining email behavior, can facilitate phishing attacks, highlighting the importance of securing data in cloud systems [2]. Protecting information throughout storage, management, analysis, and transfer is critical, with robust solutions emphasizing confidentiality, integrity, and availability. Data security involves preventing unauthorized access and maintaining authenticity through encryption and signature systems [3]. Attribute-Based Encryption (ABE) enables secure communication with multiple receivers and access control. Attribute-Based Signatures (ABS) allow verification of document legitimacy without revealing the signer's identity. Attribute-Based Signcryption (ABSC) combines ABE and ABS features efficiently, with lower computational and transmission costs, providing strong data security [4].

In large-scale data analysis, three factors are crucial: accessibility, privacy, and consistency. Confidentiality restricts access to authorized personnel, integrity allows controlled modification, and availability ensures data usability. Centralized storage of sensitive data increases risks of loss, sabotage, and hacking. Risk evaluation methods classify hazards based on assets, vulnerabilities, threats, and likelihood [5]. Encryption and hidden access controls enhance data security, though managing large datasets complicates implementation [6–7]. Smart security strategies monitor users exhibiting abnormal activity, using logs, behavioral cues, and keyword libraries to detect potential threats. However, cloud reliability remains a concern, with 70% of users expressing data security worries. Innovative cloud-based encryption methods aim to address these issues [8]. Homomorphic encryption (HE) allows computation on encrypted data without decryption. RSA2 demonstrates multiplicative HE, while Paillier supports additive HE [9]. Fully Homomorphic Encryption (FHE) enables all operations on encrypted data while preserving encryption integrity, but designing practical FHE systems remains challenging. Conventional encryption relies on key exchanges, which can be vulnerable if keys are compromised. Even when leaving cloud services, users' sensitive data may remain accessible to others [10]. FHE, introduced by Gentry in 2009, is promising but requires further advancement for universal compatibility.

*Corresponding author. Email: vkvaigandla@gmail.com

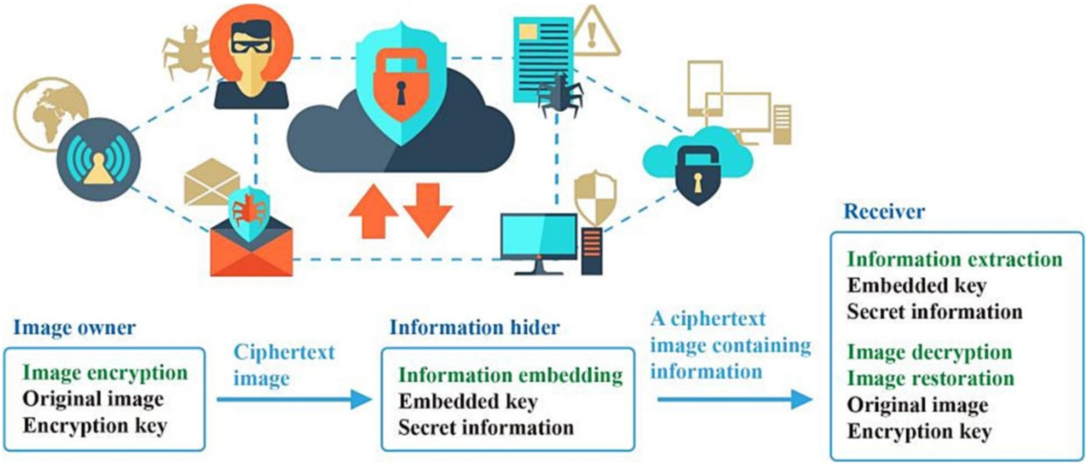

Fig. 1. Big Data Encryption Technology for Image Encryption.

The fundamental issue is maintaining data security and privacy amid rapidly growing digital data. Conventional security measures, designed for static data, are inadequate for the dynamic and extensive nature of multimodal data [11], increasing risks of unauthorized access, breaches, and cyberattacks like phishing. Reliance on often-untrusted cloud services further exacerbates these concerns. Homomorphic Encryption (HE) enables computations on encrypted data, preserving anonymity, but its practical application and performance remain limited. The key challenge is developing robust, scalable, and efficient security methods that ensure data privacy, integrity, and availability, while accommodating the specific demands of big data and cloud computing environments.

The main objectives for review paper:

1) To review and categorize existing data classification methods and emerging encryption technologies in the context of big data and cloud computing.
2) To evaluate their effectiveness, scalability, and limitations in ensuring data security, confidentiality, and integrity.
3) To identify key research gaps and propose future directions for developing more robust and adaptive solutions.

2. LITERATURE REVIEW

Researchers have increasingly focused on data security, yet many protocols are designed for static data and struggle with large datasets exceeding traditional database capacities. Sensitive personal information remains a prime target for threats that can damage organizational trust and reputation. For instance, big data exploitation can enable phishing attacks by analyzing users' email habits, posing risks to communication security. Securing big data in cloud computing is critical, as breaches can harm a company's reputation [12]. Table 1 summarizes the literature survey. Recent studies [13] highlight growing interest in secure data outsourcing to untrusted cloud servers. Homomorphic Encryption (HE) enhances data privacy for both consumers and cloud providers and has been applied in various computing scenarios to preserve anonymity. For example, [14] proposed a privacy-preserving solution for IoT applications using communication-efficient BGN HE techniques, while [15] introduced a secure cloud computing system leveraging HE to protect user data confidentiality.

The unpredictable nature of multi-modal data presents unique challenges for classification models, often reducing their performance [16]. HE allows secure computations on encrypted data without revealing content. Unlike traditional encryption methods that rely on key sharing, HE addresses confidentiality concerns even when access is restricted to authorized parties. However, widely used cloud services pose risks: if encryption keys are compromised, unauthorized users may access sensitive data, and former service providers' employees or contractors may retain access [17–18]. Fully Homomorphic Encryption (FHE) faces inefficiency due to reliance on matrix-based computations; minor decryption errors can result in incorrect messages. Using simpler encryption/decryption methods can mitigate such issues while maintaining security [19].

The model securely transfers input to a nonlinear transformation, decrypts the owner's data, performs computation, re-encrypts the output, and transmits it. However, this process imposes additional delays on the user [20]. Security measures using arbitrary execution commands address these concerns, ensuring data reaches the authorized recipient safely. In cloud environments, encryption ensures that only recipients with the sender's private key can decode the data. Users must provide a private key for secure computation, and as computations become complex, key exposure risk increases [21]. Symmetric encryption requires compatible keys for both parties and secure key backup. Security breaches can extend processing times and computational load, while encryption mechanisms provide robust privacy and protection [22].

Recent research explores diverse approaches to big data security and analytics. Studies [23–27] propose integrated methods for classifying and securing big data, emphasizing risk-based mobility control. [28] summarizes cybersecurity trends and challenges. Security-by-design frameworks for cloud-based big data deployment are introduced in [29–31], complemented by reviews of database security and privacy attributes [32]. Blockchain-based storage security and flexible protocols are proposed in [33], while [34–37] focus on cloud data protection using partitioning, partial decryption, and analysis. Classification methods for evolving data and phishing detection are discussed in [38–39], including weak KNN-based random chunk selection. Deep learning models for encrypted mobile traffic classification are presented in [40–42], alongside blockchain-auditable privacy-preserving schemes for IoT [43–44]. Integrated methodologies for classification and security are reiterated, with studies on adversarial attacks in fault detection systems [45–48] and MDSA algorithms for real-time measurement data classification [49]. Identity-based dynamic data auditing schemes for medical big data, risk indicator systems, and early economic security analysis via big data are explored in [50–51].

TABLE I. LITERATURE SURVEY

Reference	Method	Research Gap	Merits	Demerits
[11]	Secure Data Outsourcing	Need for improved cloud server reliability	Enhances privacy between cloud servers and users	Reliance on cloud servers, potential inefficiency
[12]	Fully Homomorphic Encryption	Applicability and efficiency of FHE	Enables computations on encrypted data	Still needs major Improvements for broad use
[13]	Proprietary Encrypted Patterns	Secure and efficient Encryption techniques	Solves confidentiality concerns in data sharing	Potential security risks in encryption process
[14]	Privacy- Preserving Mechanisms	More efficient privacy-preserving solutions	Preserves data privacy	Increases Computational overhead and time
[15]	CryptoNets	Enhanced security in neural network models	Allows encrypted predictions on cloud	Complex and potential Information leakage
[16]	HE-based Framework for Big Data	More efficient and user friendly solutions	Safeguards sensitive data	Response time issues, Interaction complexity
[17]	BGN HE Techniques	Improved IoT data security	Privacy preserving For IoT applications	High complexity, potential data leakage
[18]	Secure Cloud Computing Platform	Secure data processing Methods	Utilizes HE to protect user data privacy	Requires access to decryption keys
[19]	Multi-modal Data Classification	Enhanced classification models	Special treatment for higher variability of data	Performance issues with regular models
[46]	Automatic Document Classification	Accuracy and reliability in diverse data sets	Efficient handling of large volumes of data	Potential misclassification issues
[47]	Big Data Security Mechanisms	Scalable and adaptable security for big data	Protects fixed data against threats	Insufficient for dynamic nature of big data
[48]	Phishing Detection Techniques	Broader applicability and detection capabilities	Targets email based threats	Limited to specific types of threats
[49]	Cloud Computing Security Solutions	Comprehensive security solutions for cloud systems	Enhances organizations' reputation and trust	May not be fully effective against all threats
[50]	Data Storage and Management	Secure decentralized data storage methods	Ensures data confidentiality, integrity, availability	Concentration of data increases risk of attacks
[51]	Intelligentdriven Security Model	Improved data loss and leakage prevention	Monitors users for abnormal behaviors	No protection against data loss and leakage
[52]	NP-Hard Data Analysis	More efficient computational methods	Addresses crucial issues in big data analysis	Computationally intensive
[53]	Confidentiality Techniques	Feasible and efficient big data confidentiality	Protects big data From unauthorized access	Implementation difficulties for big data
[55]	Risk Metrics-Based Assessment	Comprehensive risk assessment for big data	Promotes risk management	May not cover all potential risks

TABLE II. RESEARCH GAP

Parameter	Description
Improved Reliability and Efficiency of Cloud Servers	Developing more reliable and efficient cloud server solutions to enhance data security and privacy
Applicability and Efficiency of FHE	Advancing FHE to make it broadly applicable and efficient for various platforms and data types.
Secure and Efficient Encryption Techniques	Creating encryption methods that are both secure and efficient, particularly for proprietary encrypted patterns, to mitigate potential security risks in the encryption process
Cost-effective Data Handling Methods Ensuring Security and Privacy	Finding ways to ensure data security and privacy in a cost-effective manner, especially in the context of large scale data handling, encryption, and signature techniques
Comprehensive Risk Management Techniques for Big Data	Creating more comprehensive and all-encompassing risk management techniques that address a wider range of potential threats and vulnerabilities in big data.
User-Friendly and Efficient Big Data Security Solutions	Developing solutions that are both user-friendly and efficient in handling the unique complexities and response time issues associated with big data security.
Enhanced Security in Neural Network Models	Improving the security aspects of neural network models, such as CryptoNets, to prevent potential information leakage and complexity issues.
Efficient Privacy-Preserving Solutions	Designing more efficient privacy-preserving mechanisms that reduce computational overhead and processing time while maintaining high levels of data privacy.

3. METHODOLOGY

3.1. DATA CLASSIFICATION TECHNIQUES AND EMERGING ENCRYPTION TECHNOLOGIES FOR BIG DATA

Effective data classification and encryption are essential for managing and securing big data, addressing challenges related to the volume, variety, and velocity of modern datasets. The following classification techniques are commonly applied in big data environments, along with guidance on their suitability in different scenarios. Emerging encryption technologies are summarized in Table 3.

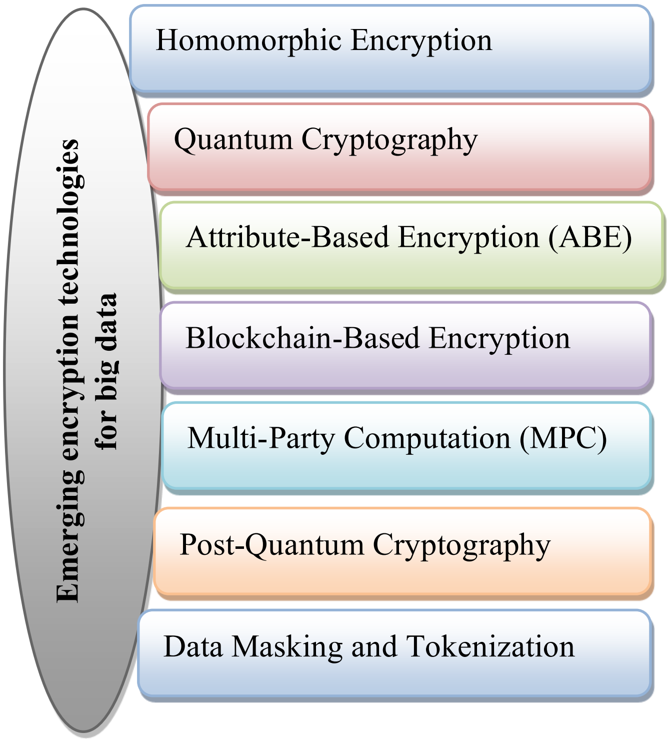
Emerging encryption technologies for big data
Homomorphic Encryption
Quantum Cryptography
Attribute-Based Encryption (ABE)
Blockchain-Based Encryption
Multi-Party Computation (MPC)
Post-Quantum Cryptography
Data Masking and Tokenization

Fig. 2. Emerging Encryption Technologies for Big Data

Manual Classification: Involves human effort to classify data based on predefined criteria. It is most effective for small datasets or highly specialized data where domain expertise is critical. However, it is time-consuming and not scalable for large datasets.

Automated Classification**:** Utilizes algorithms to classify data efficiently, making it suitable for large-scale environments. Key techniques include:

- Rule-Based Systems (RBS): Apply predefined rules set by domain experts. Effective when rules are stable but less adaptive to dynamic datasets.
- Machine Learning (ML): Trains models on labeled datasets to classify new data. Best suited for structured and semi-structured data where historical patterns exist.
- Natural Language Processing (NLP): Analyzes text data to determine categories. Particularly useful in unstructured data scenarios like social media analysis or customer feedback.

Supervised Learning**:** ML models trained on labeled data. Common algorithms include Decision Trees, Random Forests, Support Vector Machines, and Neural Networks. Highly effective for anomaly detection and predictive analytics when high-quality labeled datasets are available.

Unsupervised Learning: Finds patterns without labeled data using techniques such as K-Means, Hierarchical Clustering, Apriori, and FP-Growth. Suitable for discovering hidden structures, clustering, and anomaly detection in unlabeled datasets.

Semi-Supervised Learning: Combines a small amount of labeled data with a large volume of unlabeled data. Useful when labeling is costly, such as in medical or financial datasets, offering a balance between accuracy and resource efficiency.

Deep Learning: Employs multi-layer neural networks to classify complex data, including images, speech, and text. Particularly effective in big data applications with high-dimensional and unstructured datasets but requires significant computational resources.

Big Data-Specific Techniques:

- MapReduce: Distributes data processing across a cluster of computers, enabling efficient handling of large-scale data. Best suited for batch processing and structured datasets.
- Apache Spark: A unified analytics engine for big data with modules for streaming, SQL, machine learning, and graph processing. Offers high-speed in-memory computation, making it suitable for real-time analytics.

Critical Analysis: Supervised and semi-supervised learning methods are generally more accurate for tasks like anomaly detection but require labeled data. Unsupervised methods are flexible and scalable for discovering patterns in raw data but may yield less precise results. Deep learning excels in complex, unstructured data scenarios but comes with high computational costs. Big data-specific frameworks like MapReduce and Spark improve scalability and processing speed, complementing classification techniques for practical cloud and distributed environments.

TABLE III. EMERGING ENCRYPTION TECHNOLOGIES FOR BIG DATA

Technique	Description	Advantages	Challenges
Homomorphic Encryption	Allows computations on encrypted data without decrypting it first.	Enhances data security and privacy, particularly in cloud computing.	Computationally intensive and slower compared to traditional encryption.
Quantum Cryptography	Utilizes quantum mechanics principles to secure data.	Theoretically unbreakable by conventional computers.	Requires specialized hardware and infrastructure.
Attribute-Based Encryption (ABE)	Encrypts data such that decryption depends on user attributes	Fine-grained access control	More complex key management
Blockchain-Based Encryption	Uses blockchain technology to secure and verify transactions.	Decentralization, immutability, transparency	Scalability, energy consumption
Multi-Party Computation (MPC)	Allows multiple parties to jointly compute a function over their inputs while keeping those inputs private	Enhances privacy and security in collaborative computations	Computationally intensive, requires sophisticated protocols
Post-Quantum Cryptography	Develops cryptographic algorithms that are secure against quantum computer attacks.	Future-proof against quantum computing threats	Implementation and standardization are ongoing
Data Masking and Tokenization	Replaces sensitive data with non-sensitive equivalents	Protects sensitive information in non-production environments	Requires secure handling of original data and mappings

3.2 DATA CLASSIFICATION TECHNIQUES AND EMERGING ENCRYPTION TECHNOLOGIES FOR CLOUD COMPUTING

The data classification techniques and emerging encryption technologies are essential for ensuring the security, privacy, and compliance of data in cloud computing environments, addressing the unique challenges posed by the scalability, multi-tenancy, and distributed nature of cloud platforms. The data classification techniques are listed in Table 4 and emerging encryption technologies for cloud computing are listed in Table 5.

TABLE IV. DATA CLASSIFICATION TECHNIQUES FOR CLOUD COMPUTING

Technique	Description	Techniques	Use Cases
Rule-Based Classification	Uses predefined rules to classify data based on content, context, and metadata.	Regular expressions, keyword matching, pattern recognition	Effective for structured and semi-structured data
ML based Supervised Learning	Models are trained on labeled datasets to classify new data	Decision Trees, Random Forests, Neural Networks	Classifying emails as spam or not, sentiment analysis
ML based Unsupervised Learning	Models identify patterns in data without predefined labels	Clustering (K-Means, DBSCAN), Anomaly Detection	Identifying anomalous data points, segmenting users
NLP	Analyzes text data to classify and extract relevant information	Text categorization, sentiment analysis, topic modeling	Automating customer service responses, analyzing customer feedback
Content-Based Classification	Analyzes the actual content of data files to determine their classification	File type identification, data fingerprinting, pattern matching	Identifying confidential documents, detecting sensitive information
Context-Based Classification	Considers the context in which data is created, accessed, and used to classify it	Access patterns, user roles, data provenance	Dynamic access control, compliance monitoring
Big Data and Cloud-Specific Techniques	Leveraging cloud-native tools and services to classify large volumes of data	AWS Macie, Google Cloud Data Loss Prevention (DLP), Azure Information Protection	Automating data classification and protection in cloud environments
Data Tagging and Labeling	Tagging data with metadata to categorize it based on sensitivity, compliance requirements, or business relevance	-	Enables easy identification and management of data across cloud environments

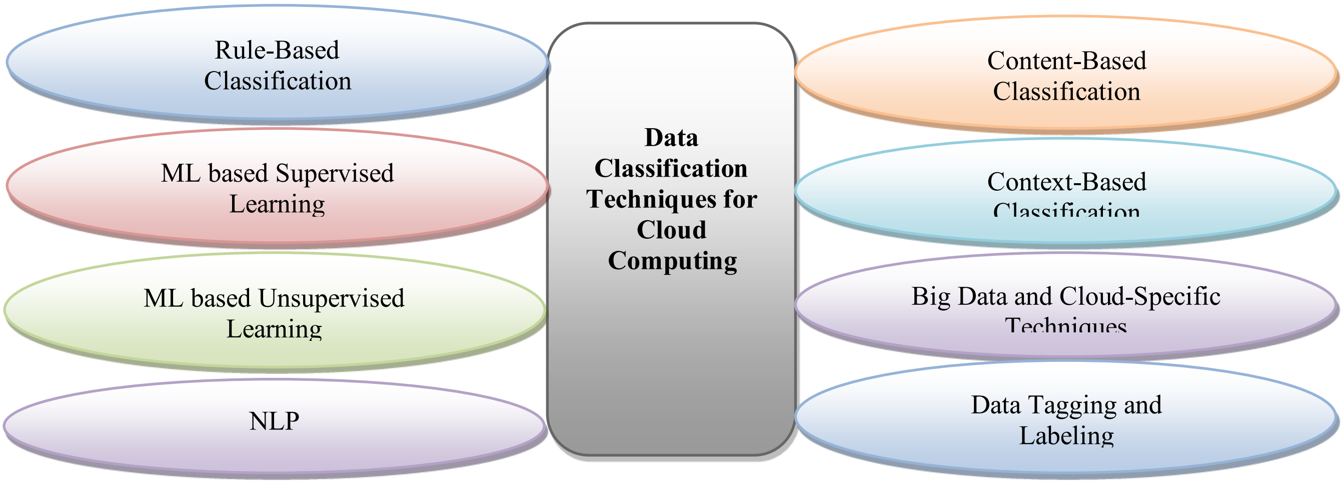
Rule-Based Classification
ML based Supervised Learning
ML based Unsupervised Learning
NLP
Data Classification Techniques for Cloud Computing
Content-Based Classification
Context-Based Classification
Big Data and Cloud-Specific Techniques
Data Tagging and Labeling

Fig. 3. Data Classification Techniques for Cloud Computing

TABLE V. EMERGING ENCRYPTION TECHNOLOGIES FOR CLOUD COMPUTING

Technique	Description	Advantages	Challenges
Homomorphic Encryption	Allows computations on encrypted data without decrypting it first.	Enhances data security and privacy, particularly in cloud computing.	Computationally intensive and slower compared to traditional encryption.
Quantum Cryptography	Utilizes quantum mechanics principles to secure data	Provides theoretically unbreakable security against conventional attacks.	Requires specialized hardware and infrastructure, not yet widely adopted.
Attribute-Based Encryption (ABE)	Encrypts data such that decryption depends on user attributes	Fine-grained access control tailored to user roles and attributes	Complex key management and policy enforcement.
Blockchain-Based Encryption	Uses blockchain technology to secure and verify data transactions.	Decentralized, immutable, and transparent ledger for data security.	Scalability issues and high energy consumption

Multi-Party Computation (MPC)	Enables multiple parties to jointly compute a function over their inputs while keeping those inputs private	Enhances privacy and security in collaborative computations without revealing individual data.	Computationally intensive and requires sophisticated protocols
Post-Quantum Cryptography	Develops cryptographic algorithms that are secure against quantum computer attacks.	Future-proof against potential quantum computing threats	Ongoing research and standardization efforts needed
Data Masking and Tokenization	Replaces sensitive data with non-sensitive equivalents to protect it in non-production environments	Protects sensitive information while allowing functional use of data for development and testing	Secure handling of original data and mappings, potential performance impact
Secure Multi-Tenancy	Ensures data isolation and protection in multi-tenant cloud environments	Enhances security and privacy for different tenants sharing the same cloud infrastructure.	Complex key management and maintaining performance.

3.2.1 HOMOMORPHIC ENCRYPTION FOR CLOUD COMPUTING AND BIG DATA

A form of encryption that allows computations to be performed on encrypted data without needing to decrypt it first. The results of these computations remain encrypted and can be decrypted later to reveal the correct result. Overall, homomorphic encryption is a powerful tool for ensuring data security and privacy in cloud computing and big data applications, despite its current performance and complexity limitations.

Applications in Cloud Computing : Homomorphic encryption provides a robust solution for maintaining data privacy in cloud computing environments. It allows users to store and process sensitive information on cloud servers without exposing the underlying data to the service provider, ensuring that confidentiality is preserved at all times. In addition, it enables secure computations on encrypted data, including operations such as searching, filtering, and data analysis, without requiring decryption. This capability ensures that sensitive information remains protected even while being actively processed. Moreover, homomorphic encryption supports regulatory compliance by guaranteeing that data privacy and security are maintained throughout the computational process, helping organizations meet legal and industry standards for data protection.

Applications in Big Data : In the context of big data, homomorphic encryption enables secure analytics by allowing organizations to analyze large datasets without revealing the underlying sensitive information. This is particularly valuable in sectors such as healthcare, finance, and other industries that handle confidential data. It also facilitates secure data sharing, allowing multiple parties to collaboratively analyze and utilize big data while maintaining strict privacy protections. Furthermore, homomorphic encryption supports outsourced computation, enabling organizations to delegate data processing tasks to third-party service providers without exposing raw data, thereby combining the benefits of cloud-based resources with robust data security.

Advantages & Challenges : Homomorphic encryption offers several significant advantages that make it a valuable tool for secure data processing. It enhances security by protecting data not only during storage and transmission but also while it is being actively processed. This ensures that sensitive information remains confidential throughout its lifecycle. Additionally, homomorphic encryption helps organizations maintain regulatory compliance by keeping data encrypted, thereby meeting legal and industry standards for data protection. Despite these benefits, the technology also presents notable challenges. Homomorphic encryption is computationally intensive, resulting in performance overhead that can make it slower than traditional encryption methods. Its implementation is complex, requiring specialized knowledge and resources, and scalability can be an issue when working with very large datasets due to the high computational demands involved.

3.2.2 ATTRIBUTE-BASED ENCRYPTION (ABE) FOR CLOUD COMPUTING AND BIG DATA

A type of encryption where the decryption of data depends on user attributes rather than specific keys. Access to the encrypted data is granted based on the attributes or policies defined by the data owner. Overall, ABE is a powerful encryption technique for enforcing fine-grained access control in cloud computing and big data environments, offering both security and flexibility despite some implementation challenges.

Types of ABE : ABE comes in two primary types, each designed to enforce fine-grained access control over encrypted data. In Key-Policy ABE (KP-ABE), the access policy is embedded in the decryption key rather than the ciphertext. Data is encrypted with specific attributes, and only users possessing a decryption key whose policy matches these attributes can access the data. This approach is particularly suitable for scenarios where data producers determine who can access the information based on predefined attributes. In contrast, Ciphertext-Policy ABE (CP-ABE) embeds the access policy directly into the ciphertext. Data is encrypted according to a specified policy, and only users whose attributes satisfy the policy can decrypt it. CP-ABE is ideal for situations where data owners want to maintain direct control over who can access the data they share, allowing them to enforce access policies even after the data has been distributed.

Applications in Cloud Computing : ABE provides robust solutions for secure data management in cloud computing by enabling fine-grained access control. It allows organizations to precisely determine who can access specific data based on user roles, attributes, and predefined policies, ensuring that sensitive information is only available to authorized individuals. Additionally, ABE facilitates secure data sharing among multiple users without the need to distribute encryption keys directly, reducing the risk of unauthorized access. This capability also supports regulatory compliance by ensuring that sensitive data remains accessible only to authorized users, helping organizations meet legal and industry standards for data protection while maintaining strong security in cloud environments.

Applications in Big Data : ABE enables privacy-preserving analytics by ensuring that sensitive information is accessible only to users with the appropriate attributes. This allows organizations to securely analyze large datasets without exposing confidential data to unauthorized parties. ABE also supports scalable data access, as access policies based on attributes can be efficiently applied across vast datasets, providing secure and flexible control over who can view or manipulate the data. Furthermore, ABE facilitates secure data collaboration, allowing multiple parties to share and work with data while ensuring that only authorized participants can access the information, thereby maintaining both privacy and security in large-scale data environments.

Advantages & Challenges : ABE offers several advantages that make it highly suitable for secure data management in cloud computing and big data environments. It provides enhanced security by tying access control to user attributes, ensuring that only authorized individuals can access sensitive information. ABE also offers flexibility, allowing dynamic access control policies that can be easily updated as organizational requirements change. Additionally, it is scalable, making it well-suited for large-scale environments where managing individual encryption keys would be impractical. Despite these benefits, ABE presents certain challenges. Managing and distributing attribute-based keys can be complex and resource-intensive, particularly in environments with many users. The encryption and decryption processes can also be computationally demanding, which may affect system performance. Furthermore, defining and maintaining access policies can be difficult, especially in dynamic or rapidly changing organizational settings, requiring careful planning and management to ensure consistent security.

3.2.3 BLOCKCHAIN-BASED ENCRYPTION FOR CLOUD COMPUTING AND BIG DATA

Blockchain-Based Encryption utilizes blockchain technology to secure and verify data transactions, ensuring integrity, transparency, and immutability of data. Overall, blockchain-based encryption provides a robust and secure framework for managing and protecting data in cloud computing and big data environments, despite challenges related to scalability, performance, and complexity.

Key Concepts : Blockchain technology is built upon several key concepts that collectively ensure secure and reliable data management. Decentralization distributes data and transaction records across multiple nodes in a network, reducing the risk of a single point of failure and increasing system resilience. Immutability ensures that once data is written to the blockchain, it cannot be altered or deleted, thereby preserving the integrity of records. Transparency allows all transactions to be recorded on a public ledger accessible to all participants, promoting accountability and trust among users. Finally, cryptographic security protects the data on the blockchain using advanced cryptographic algorithms, making it tamper-proof and secure from unauthorized access.

Applications in Cloud Computing & Big Data : Blockchain technology enhances data security and management by enabling secure data storage, ensuring that information stored in the cloud is tamper-proof and protected from unauthorized access. It also strengthens data integrity by recording all transactions on the blockchain, allowing users to verify the accuracy and authenticity of stored data. Additionally, blockchain facilitates access control through smart contracts, automatically enforcing policies to ensure that only authorized users can access specific datasets. Blockchain provides mechanisms for data provenance, tracking the origin and history of data to ensure authenticity and integrity. It supports secure data sharing, allowing multiple parties to collaborate on large datasets while maintaining transparency and protecting against tampering. Moreover, blockchain creates immutable audit trails of all data transactions, which are valuable for compliance, regulatory reporting, and accountability in large-scale data environments.

Advantages & Challenges : Blockchain technology offers several advantages that make it a powerful tool for secure and reliable data management. It provides enhanced security through decentralization and cryptographic techniques, ensuring that data is protected against tampering and unauthorized access. Data integrity is guaranteed because once information is recorded on the blockchain, it cannot be altered. Additionally, blockchain promotes transparency, fostering trust and accountability by maintaining a visible record of all transactions. Decentralized control reduces reliance on a central authority, increasing system robustness and resilience. Despite these benefits, blockchain also presents certain challenges. Scalability can be a concern, as managing large volumes of data on a blockchain may be limited by storage and processing capacities. Performance overhead is another issue, since blockchain transactions can be slower and more resource-intensive compared to traditional databases. The complexity of implementing and managing blockchain-based systems requires specialized expertise, and some consensus mechanisms, such as proof-of-work, can result in high energy consumption, adding further operational considerations.

3.2.4 QUANTUM CRYPTOGRAPHY FOR CLOUD COMPUTING AND BIG DATA

Quantum Cryptography utilizes principles of quantum mechanics to secure data, providing theoretically unbreakable encryption. Overall, quantum cryptography provides cutting-edge security for cloud computing and big data environments, offering a high level of protection against current and future threats despite the need for specialized hardware and ongoing research.

Key Concepts : Quantum cryptography leverages the principles of quantum mechanics to achieve highly secure communication. Quantum Key Distribution (QKD) uses quantum properties, such as superposition and entanglement, to securely distribute encryption keys between parties. Any attempt to intercept or measure the keys alters their quantum state, immediately alerting the communicating parties to potential eavesdropping. Quantum Random Number Generation (QRNG) enhances cryptographic security by producing truly random numbers through quantum processes. By exploiting phenomena such as photon behavior or radioactive decay, QRNG ensures that cryptographic keys are unpredictable and highly resistant to attacks, providing a robust foundation for secure encryption.

Applications in Cloud Computing & Big Data : Quantum cryptography enhances security by enabling secure key exchange, ensuring that encryption keys are shared between cloud service providers and users without the risk of interception. It also strengthens data transmission security, protecting information sent between cloud servers and clients from tampering or eavesdropping. Additionally, quantum-based techniques support authentication, verifying the identities of users and devices in the cloud to prevent unauthorized access and enhance overall system security. Quantum cryptography ensures secure data transfer, protecting large volumes of information exchanged between nodes or data centers from interception. It also provides integrity assurance, guaranteeing that data remains unaltered during storage and transfer through secure quantum key distribution mechanisms. Furthermore, quantum techniques enable privacy-preserving analytics, allowing organizations to perform computations on sensitive big data while ensuring that encryption keys and data remain secure throughout the process.

Advantages & Challenges : Quantum cryptography offers several compelling advantages that make it a powerful tool for secure communication. It provides unbreakable security, as any attempt to intercept quantum keys is immediately detectable, ensuring that encryption remains theoretically invulnerable. The technology is also future-proof, offering protection against potential threats from advanced computing, including quantum computers. Additionally, quantum systems utilize enhanced randomness, producing truly random cryptographic keys that further strengthen security. Despite these benefits, quantum cryptography faces several challenges. Implementation requires specialized hardware, such as sophisticated and often costly quantum devices. Distance limitations constrain the effective range of quantum key distribution, which can impact long-distance communication. Integrating quantum cryptography into existing cloud and big data infrastructures presents complexity, requiring careful planning and adaptation. Furthermore, many aspects of quantum cryptography remain in ongoing research, and practical deployment continues to evolve as the technology matures.

3.2.5 DATA MASKING AND TOKENIZATION FOR CLOUD COMPUTING AND BIG DATA

Data Masking: The process of obscuring original data with modified content (characters or other data) to protect sensitive information. Ensures sensitive data is not exposed to unauthorized users while allowing it to be used in non-production environments such as development, testing, or training. Overall, data masking and tokenization are crucial techniques for protecting sensitive information in cloud computing and big data environments, offering robust security, compliance, and usability despite challenges related to performance and complexity.

Techniques : Data masking employs various techniques to protect sensitive information by replacing or obfuscating real data with fictional but realistic values. Static Data Masking (SDM) involves masking data in a static copy, typically used in non-production environments such as development or testing, to prevent exposure of sensitive information. Dynamic Data Masking (DDM) operates in real-time, masking data as it is accessed based on user roles, ensuring that unauthorized users cannot view confidential information while still allowing legitimate access. On-the-Fly Data Masking protects data during transfer or migration between environments, ensuring that sensitive information remains secure throughout the movement process without disrupting normal operations.

Applications in Cloud Computing & Big Data : In cloud computing, data masking enhances security by enabling secure development and testing, protecting sensitive information in cloud-based non-production environments. It also supports compliance by ensuring that confidential data is not exposed, helping organizations meet regulatory requirements. Additionally, data masking facilitates data sharing, allowing realistic but anonymized data to be shared with third parties for analysis or collaboration without risking sensitive information. For big data applications, data masking enables secure analytics, allowing organizations to analyze large datasets without revealing sensitive details. It also supports data sharing across departments or with external partners while maintaining privacy. Furthermore, data masking protects sensitive information stored in data lakes, ensuring that large-scale big data repositories used for analytics remain secure and compliant with data protection standards.

Advantages & Challenges : Data masking offers several advantages that make it an effective method for protecting sensitive information. It enhances security by preventing unauthorized access to confidential data and supports compliance by helping organizations meet regulatory requirements for data protection. Additionally, data masking provides flexibility, allowing realistic data to be safely used in non-production environments, such as development, testing, or analytics, without risking exposure of sensitive information. However, implementing data masking also presents certain challenges. It can introduce a performance impact, as masking operations may add overhead and affect system efficiency. The process can be complex, requiring careful design and management of masking rules and techniques. Ensuring consistency is another challenge, as masked data must remain logically coherent and useful for its intended purposes, such as analytics, reporting, or testing.

Tokenization : Replaces sensitive data with unique identification symbols (tokens) that retain essential information without exposing the original data. Protects sensitive data by substituting it with a token, which can only be mapped back to the original data through a secure tokenization system.

Techniques : Tokenization protects sensitive data by replacing it with non-sensitive equivalents, called tokens, while preserving the usability of the data for processing and analytics. Vault-Based Tokenization stores the mapping between tokens and the original data in a secure database, or vault, ensuring that sensitive information can be retrieved only through controlled access. In contrast, Vaultless Tokenization generates tokens using algorithms without relying on a central database, reducing potential single points of failure and simplifying scalability. Both techniques provide robust methods for protecting sensitive information in cloud computing and big data environments.

Applications in Cloud Computing & Big Data: In cloud computing, tokenization enhances security by protecting sensitive information throughout its lifecycle. It enables secure payment processing by replacing credit card numbers with tokens, reducing the risk of exposure. Tokenization also safeguards data storage in cloud databases, ensuring that sensitive information is never directly stored. Additionally, it supports secure data transmission, allowing tokens to be transmitted instead of actual sensitive data, thereby minimizing the risk of interception.For big data applications, tokenization ensures data privacy by protecting sensitive information within large datasets. It facilitates analytics by allowing organizations to analyze data without exposing the original confidential information. Moreover, tokenization supports regulatory compliance, helping organizations meet data protection requirements by consistently substituting sensitive information with secure tokens across cloud and big data environments.

Advantages & Challenges: Tokenization provides several advantages that make it an effective method for protecting sensitive data in cloud computing and big data environments. It enhances security by reducing the risk of data breaches, as sensitive information is replaced with non-sensitive tokens. Tokenization also supports compliance, helping organizations meet regulatory requirements for data protection. Additionally, tokens retain the usability and format of the original data, allowing them to be used seamlessly in applications such as analytics, storage, and payment processing. Despite these benefits, tokenization also presents certain challenges. Token management is critical, as secure handling of tokens and their mappings is essential to prevent unauthorized access. Integration can be complex, particularly when introducing tokenization into existing systems and workflows. Furthermore, performance may be affected, as the tokenization process can introduce computational overhead, potentially impacting system efficiency.

5. DISCUSSION

This review highlights persistent research gaps in big data and cloud security, including inefficiencies in FHE, limited scalability of ABE, and integration challenges with blockchain and quantum cryptography. Current methods struggle with balancing performance, usability, and security for large-scale, dynamic datasets. Future research should prioritize scalable post-quantum cryptographic solutions to counter emerging quantum threats, along with lightweight and efficient FHE implementations. Additionally, developing hybrid frameworks that integrate machine learning with encryption can enhance adaptability. User-friendly, low-overhead privacy-preserving systems remain critical for practical adoption in real-world cloud and big data environments.

6. CONCLUSION

This systematic review provides a comprehensive analysis of current classification approaches and encryption technologies in big data and cloud computing, emphasizing their critical role in ensuring data security and privacy. The study highlights key findings, including the limitations of existing methods in terms of efficiency, scalability, and usability, and the ongoing need for adaptive solutions that can handle the dynamic and heterogeneous nature of big data. Emerging approaches, such as FHE, ABE, and blockchain-based techniques, offer promising avenues but face practical challenges, including high computational overhead and integration complexities in real-world cloud environments. The review identifies specific research gaps, including the need for scalable post-quantum encryption, more efficient FHE implementations, and user-friendly security solutions that do not compromise performance. By synthesizing these insights, the study provides a roadmap for future research, directing efforts toward the development of resilient, efficient, and comprehensive data classification and encryption systems. Overall, this work contributes to advancing the security and privacy of digital data, offering valuable guidance for researchers and practitioners seeking to address the evolving challenges of big data environments.

Acknowledgments

The author gratefully acknowledges the support and encouragement received from colleagues, and institutional resources that contributed to the successful completion of this review article.

Funding Information

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author Contributions Statement

The author solely conceived, designed, and prepared the manuscript, including literature review, writing, and final approval of the submitted version.

Conflict of Interest Statement

The author declares no conflict of interest.

Informed Consent

Not applicable, as this article does not involve human participants.

Ethical Approval

Not applicable, as this study is a review article and does not involve experiments on humans or animals.

Data Availability

No primary datasets were generated or analyzed for this study. All data supporting this review are available in the cited references.

References

[1] M. Abadi, A. Chu, I. Goodfellow, H. B. Mcmahan, I. Mironov, K. Talwar, and L. Zhang, ‘‘Deep learning with differential privacy,’’ in Proc. ACM SIGSAC Conf. Comput. Commun. Secur., Oct. 2016, pp. 308–318, doi.org/10.48550/arXiv.1607.00133.

[2] M. Choraś and M. Pawlicki, ‘‘Intrusion detection approach based onoptimised artificial neural network,’’ Neurocomputing, vol. 452, pp. 705–715, Sep. 2021, doi.org/10.1016/j.neucom.2020.07.138.

[3] A. Acar, H. Aksu, A. S. Uluagac, and M. Conti, ‘‘A survey on homomorphic encryption schemes: Theory and implementation,’’ ACM Comput. Surv., vol. 51, no. 4, pp. 1–35, 2018, doi.org/10.48550/arXiv.1704.03578.

[4] B. Li and D. Micciancio, ‘‘On the security of homomorphic encryption on approximate numbers,’’ in Advances in Cryptology—EUROCRYPT 2021 (Lecture Notes in Computer Science), vol. 12696, A. Canteaut and F. X. Standaert, Eds. Cham, Switzerland: Springer, Oct. 2021, doi: 10.1007/978- 3-030-77870-5_23.

[5] F. Boemer, A. Costache, R. Cammarota, and C. Wierzynski, ‘‘NGraphHE2: A high-throughput framework for neural network inference on encrypted data,’’ in Proc. 7th ACM Workshop Encrypted Comput. Appl. Homomorphic Cryptogr. (WAHC), 2019, pp. 45–56, doi.org/10.48550/arXiv.1908.04172.

[6] Z. Du et al., "Merge Loss Calculation Method for Highly Imbalanced Data Multiclass Classification," in IEEE Transactions on Neural Networks and Learning Systems, doi: 10.1109/TNNLS.2023.3321753.

[7] S. K. V, M. K. V, C. N. Azmea, and K. K. Vaigandla, “BCSDNCC: A Secure Blockchain SDN framework for IoT and Cloud Computing”, Int. Res. J. multidiscip. Technovation, vol. 6, no. 3, pp. 26–44, Apr. 2024, doi: 10.54392/irjmt2433.

[8] B. K. Sethi, D. Singh, S. K. Rout and S. K. Panda, "Long Short-Term Memory-Deep Belief Network based Gene Expression Data Analysis for Prostate Cancer Detection and Classification," in IEEE Access, doi: 10.1109/ACCESS.2023.3346925.

[9] M. J. Zideh, P. Chatterjee and A. K. Srivastava, "Physics-Informed Machine Learning for Data Anomaly Detection, Classification, Localization, and Mitigation: A Review, Challenges, and Path Forward," in IEEE Access, doi: 10.1109/ACCESS.2023.3347989.

[10] C. Madhu and S. M.S., "An Interpretable Fuzzy Graph Learning for Label Propagation Assisting Data Classification," in IEEE Transactions on Fuzzy Systems, doi: 10.1109/TFUZZ.2023.3323093.

[11] K. K. Vaigandla, "Communication Technologies and Challenges on 6G Networks for the Internet: Internet of Things (IoT) Based Analysis," *2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM)*, 2022, pp. 27-31, doi: 10.1109/ICIPTM54933.2022.9753990.

[12] A. Alabdulatif, I. Khalil, and X. Yi, ''Towards secure big data analytic for cloud-enabled applications with fully homomorphic encryption,'' J. Parallel Distrib. Comput., vol. 137, pp. 192–204, Mar. 2020, doi.org/10.1016/j.jpdc.2019.10.008.

[13] J. Park, D. S. Kim and H. Lim, "Privacy-Preserving Reinforcement Learning Using Homomorphic Encryption in Cloud Computing Infrastructures," in IEEE Access, vol. 8, pp. 203564-203579, 2020, doi: 10.1109/ACCESS.2020.3036899.

[14] H. Pang and B. Wang, "Privacy-Preserving Association Rule Mining Using Homomorphic Encryption in a Multikey Environment," in IEEE Systems Journal, vol. 15, no. 2, pp. 3131- 3141, June 2021, doi: 10.1109/JSYST.2020.3001316.

[15] Li, X. Kuang, S. Lin, X. Ma, and Y. Tang, ''Privacy preservationfor machine learning training and classification based on homomorphicencryption schemes,'' Inf. Sci., vol. 526, pp. 166–179, Jul. 2020, doi.org/10.1016/j.ins.2020.03.041.

[16] A. Agarwal, M. Khari, and R. Singh, ''Detection of DDOS attack using deep learning model in cloud storage application,'' Wireless Pers. Commun., Mar. 2021, doi: 10.1007/s11277-021-08271-z.

[17] Z. Zhang, C. Li, B. B. Gupta and D. Niu, "Efficient Compressed Ciphertext Length Scheme Using Multi-Authority CP-ABE for Hierarchical Attributes," in IEEE Access, vol. 6, pp. 38273-38284, 2018, doi: 10.1109/ACCESS.2018.2854600.

[18] B. Gupta, "An efficient KP design framework of attribute-based searchable encryption for user level revocation in cloud," Concurrency Comput. Pract. Exp., vol. 32, no. 18, 2020, Art. no. e5291, doi.org/10.1002/cpe.5291.

[19] B. Joshi, B. Joshi, A. Mishra, V. Arya, A. K. Gupta, and D. Perakovic,´ "A comparative study of privacy-preserving homomorphic encryption techniques in cloud computing," Int. J. Cloud Appl. Comput., vol. 12, no. 1, pp. 1–11, 2022, DOI:10.4018/IJCAC.309936.

[20] H. Chen, W. Dai, M. Kim, and Y. Song, "Efficient multi-key homomorphic encryption with packed ciphertexts with application to oblivious neural network inference," in Proc. ACM SIGSAC Conf. Comput. Commun. Security, 2019, pp. 395–412, DOI:10.1145/3319535.3363207.

[21] C. Marcolla, V. Sucasas, M. Manzano, R. Bassoli, F. H. P. Fitzek and N. Aaraj, "Survey on Fully Homomorphic Encryption, Theory, and Applications," in Proceedings of the IEEE, vol. 110, no. 10, pp. 1572-1609, Oct. 2022, doi: 10.1109/JPROC.2022.3205665.

[22] M. Ali, J. Mohajeri, M.-R. Sadeghi, and X. Liu, "A fully distributed hierarchical attributebased encryption scheme," Theor. Comput. Sci., vol. 815, pp. 25–46, May 2020, doi.org/10.1016/j.tcs.2020.02.030.

[23] W. Xu, Y. Zhan, Z. Wang, B. Wang and Y. Ping, "Attack and Improvement on a Symmetric Fully Homomorphic Encryption Scheme," in IEEE Access, vol. 7, pp. 68373-68379, 2019, doi: 10.1109/ACCESS.2019.2917028.

[24] D. B. Rawat, R. Doku and M. Garuba, "Cybersecurity in Big Data Era: From Securing Big Data to Data-Driven Security," in IEEE Transactions on Services Computing, vol. 14, no. 6, pp. 2055-2072, 1 Nov.-Dec. 2021, doi: 10.1109/TSC.2019.2907247.

[25] F. M. Awaysheh, M. N. Aladwan, M. Alazab, S. Alawadi, J. C. Cabaleiro and T. F. Pena, "Security by Design for Big Data Frameworks Over Cloud Computing," in IEEE Transactions on Engineering Management, vol. 69, no. 6, pp. 3676-3693, Dec. 2022, doi: 10.1109/TEM.2020.3045661.

[26] G. D. Samaraweera and J. M. Chang, "Security and Privacy Implications on Database Systems in Big Data Era: A Survey," in IEEE Transactions on Knowledge and Data Engineering, vol. 33, no. 1, pp. 239-258, 1 Jan. 2021, doi: 10.1109/TKDE.2019.2929794.

[27] Sivapriya, N. ., Mohandas, R. ., & Vaigandla, K. K. . (2023). A QoS Perception Routing Protocol for MANETs Based on Machine Learning. International Journal of Intelligent Systems and Applications in Engineering, 12(1), 733–745. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/4171

[28] A. Sasikumar, L. Ravi, K. Kotecha, A. Abraham, M. Devarajan and S. Vairavasundaram, "A Secure Big Data Storage Framework Based on Blockchain Consensus Mechanism With Flexible Finality," in IEEE Access, vol. 11, pp. 56712-56725, 2023, doi: 10.1109/ACCESS.2023.3282322.

[29] R. Gupta, I. Gupta, A. K. Singh, D. Saxena and C. -N. Lee, "An IoT-Centric Data Protection Method for Preserving Security and Privacy in Cloud," in IEEE Systems Journal, vol. 17, no. 2, pp. 2445-2454, June 2023, doi: 10.1109/JSYST.2022.3218894.

[30] R. Abdillah, Z. Shukur, M. Mohd, T. S. M. Z. Murah, I. Oh and K. Yim, "Performance Evaluation of Phishing Classification Techniques on Various Data Sources and Schemes," in IEEE Access, vol. 11, pp. 38721-38738, 2023, doi: 10.1109/ACCESS.2022.3225971.

[31] A. S. Tarawneh, E. S. Alamri, N. N. Al-Saedi, M. Alauthman and A. B. Hassanat, "CTELC: A Constant-Time Ensemble Learning Classifier Based on KNN for Big Data," in IEEE Access, vol. 11, pp. 89791-89802, 2023, doi: 10.1109/ACCESS.2023.3307512.

[32] M. Dener, S. Al and G. Ok, "RFSE-GRU: Data Balanced Classification Model for Mobile Encrypted Traffic in Big Data Environment," in IEEE Access, vol. 11, pp. 21831-21847, 2023, doi: 10.1109/ACCESS.2023.3251745.

[33] Q. Liu, P. Li, W. Zhao, W. Cai, S. Yu and V. C. M. Leung, "A Survey on Security Threats and Defensive Techniques of Machine Learning: A Data Driven View," in IEEE Access, vol. 6, pp. 12103-12117, 2018, doi: 10.1109/ACCESS.2018.2805680.

[34] Y. Zhao, X. Yang, Y. Yu, B. Qin, X. Du and M. Guizani, "Blockchain-Based Auditable Privacy-Preserving Data Classification for Internet of Things," in IEEE Internet of Things Journal, vol. 9, no. 4, pp. 2468-2484, 15 Feb.15, 2022, doi: 10.1109/JIOT.2021.3097890.

[35] I. Hababeh, A. Gharaibeh, S. Nofal and I. Khalil, "An Integrated Methodology for Big Data Classification and Security for Improving Cloud Systems Data Mobility," in IEEE Access, vol. 7, pp. 9153-9163, 2019, doi: 10.1109/ACCESS.2018.2890099.

[36] Y. Zhuo, Z. Yin and Z. Ge, "Attack and Defense: Adversarial Security of Data-Driven FDC Systems," in IEEE Transactions on Industrial Informatics, vol. 19, no. 1, pp. 5-19, Jan. 2023, doi: 10.1109/TII.2022.3197190.

[37] S. Liu et al., "Model-Free Data Authentication for Cyber Security in Power Systems," in IEEE Transactions on Smart Grid, vol. 11, no. 5, pp. 4565-4568, Sept. 2020, doi: 10.1109/TSG.2020.2986704.

[38] A. Zigomitros, F. Casino, A. Solanas and C. Patsakis, "A Survey on Privacy Properties for Data Publishing of Relational Data," in IEEE Access, vol. 8, pp. 51071-51099, 2020, doi: 10.1109/ACCESS.2020.2980235.

[39] T. Shang, F. Zhang, X. Chen, J. Liu and X. Lu, "Identity-Based Dynamic Data Auditing for Big Data Storage," in IEEE Transactions on Big Data, vol. 7, no. 6, pp. 913-921, 1 Dec. 2021, doi: 10.1109/TBDATA.2019.2941882.

[40] R. Jiang, M. Shi and W. Zhou, "A Privacy Security Risk Analysis Method for Medical Big Data in Urban Computing," in IEEE Access, vol. 7, pp. 143841-143854, 2019, doi: 10.1109/ACCESS.2019.2943547.

[41] Y. Liang, D. Quan, F. Wang, X. Jia, M. Li and T. Li, "Financial Big Data Analysis and Early Warning Platform: A Case Study," in IEEE Access, vol. 8, pp. 36515-36526, 2020, doi: 10.1109/ACCESS.2020.2969039.

[42] C. Yang, X. Xu, K. Ramamohanarao and J. Chen, "A Scalable Multi-Data Sources Based Recursive Approximation Approach for Fast Error Recovery in Big Sensing Data on Cloud," in IEEE Transactions on Knowledge and Data Engineering, vol. 32, no. 5, pp. 841-854, 1 May 2020, doi: 10.1109/TKDE.2019.2895612.

[43] R. Yadav, Ritambhara, K. K. Vaigandla, G. S. P. Ghantasala, R. Singh and D. Gangodkar, "The Block Chain Technology to protect Data Access using Intelligent Contracts Mechanism Security Framework for 5G Networks," *2022 5th International Conference on Contemporary Computing and Informatics (IC3I)*, Uttar Pradesh, India, 2022, pp. 108-112, doi: 10.1109/IC3I56241.2022.10072740.

[44] K. K. Vaigandla, R. Karne, M. Siluveru, and M. Kesoju, "Review on blockchain technology: Architecture, characteristics, benefits, algorithms, challenges and applications," Mesopotamian Journal of CyberSecurity, pp. 73–85, 2023, doi: 10.58496/MJCS/2023/012.

[45] S. Funde and G. Swain, "Big Data Privacy and Security Using Abundant Data Recovery Techniques and Data Obliviousness Methodologies," in IEEE Access, vol. 10, pp. 105458- 105484, 2022, doi: 10.1109/ACCESS.2022.3211304.

[46] K. R. Sollins, "IoT Big Data Security and Privacy Versus Innovation," in IEEE Internet of Things Journal, vol. 6, no. 2, pp. 1628-1635, April 2019, doi: 10.1109/JIOT.2019.2898113.

[47] S. Seo and J. -M. Chung, "Adaptive Trust Management and Data Process Time Optimization for Real-Time Spark Big Data Systems," in IEEE Access, vol. 9, pp. 156372-156379, 2021, doi: 10.1109/ACCESS.2021.3129885.

[48] S. Han, K. Han and S. Zhang, "A Data Sharing Protocol to Minimize Security and Privacy Risks of Cloud Storage in Big Data Era," in IEEE Access, vol. 7, pp. 60290-60298, 2019, doi: 10.1109/ACCESS.2019.2914862.

[49] C. Banapuram, A. C. Naik, M. K. Vanteru, V. S. Kumar, and K. K. Vaigandla, "A comprehensive survey of machine learning in healthcare: Predicting heart and liver disease, tuberculosis detection in chest X-ray images," SSRG International Journal of Electronics and Communication Engineering, vol. 11, no. 5, pp. 155–169, 2024.

[50] X. Yang, R. Lu, K. K. R. Choo, F. Yin and X. Tang, "Achieving Efficient and Privacy- Preserving Cross-Domain Big Data Deduplication in Cloud," in IEEE Transactions on Big Data, vol. 8, no. 1, pp. 73-84, 1 Feb. 2022, doi: 10.1109/TBDATA.2017.2721444.

[51] A. D’Alconzo, I. Drago, A. Morichetta, M. Mellia and P. Casas, "A Survey on Big Data for Network Traffic Monitoring and Analysis," in IEEE Transactions on Network and Service Management, vol. 16, no. 3, pp. 800-813, Sept. 2019, doi: 10.1109/TNSM.2019.2933358.

[52] M. Ali, M. -R. Sadeghi and X. Liu, "Lightweight Revocable Hierarchical Attribute-Based Encryption for Internet of Things," in IEEE Access, vol. 8, pp. 23951-23964, 2020, doi: 10.1109/ACCESS.2020.2969957.

[53] M. Ali, M.-R. Sadeghi, X. Liu, Y. Miao, and A. V. Vasilakos, “Verifiable online/offline multikeyword search for cloud-assisted Industrial Internet of Things,” J. Inf. Security Appl., vol. 65, Mar. 2022, Art. no. 103101, doi.org/10.1016/j.jisa.2021.103101.

[54] K. Munjal and R. Bhatia, “A systematic review of homomorphic encryption and its contributions in healthcare industry,” Complex Intelligent Systems, vol. 9, pp. 3759–3786, 2023, doi: 10.1007/s40747-022-00756-z.

[55] Z. Zhang, C. Li, B. B. Gupta and D. Niu, "Efficient Compressed Ciphertext Length Scheme Using Multi-Authority CP-ABE for Hierarchical Attributes," in IEEE Access, vol. 6, pp. 38273-38284, 2018, doi: 10.1109/ACCESS.2018.2854600.

[56] K. K. Vaigandla, T. Mounika, N. Azmi, U. Urooj, K. Chenigaram, and R. K. Karne, “Investigation on machine learning towards future generation communications,” in AIP Conference Proceedings, vol. 2965, no. 1, July 2024, AIP Publishing.