A Comprehensive Survey of data Classification Techniques and Emerging Encryption Technologies

Sadiya Ansari[1], Shameem Akther[2]

[1]Department of Faculty of Engineering and Technology, Research Scholar, KBN University, Kalaburagi, India

[2]Department of Faculty of Engineering and Technology, Associate Professor, KBN University, Kalaburagi, India

Abstract

This paper presents a comprehensive review of recent advancements in data classification techniques within the realm of big data and cloud computing. As organizations increasingly rely on large volumes of digital data, the need for robust classification mechanisms has become dominant. This paper specifically addresses the challenges and solutions in classifying data that contains sensitive or private information. The paper reviews a range of classification methods, from automatic document classification to encryption and decryption techniques, each addressing different sides of the problem. The focus is on how these methods ensure the confidentiality, integrity, and availability of data, which are crucial in safeguarding against unauthorized access and cyber threats. Special attention is given to the emerging field of Homomorphic Encryption (HE) and its role in enhancing privacy in cloud-based systems. The survey identifies critical research gaps, such as the need for more efficient, scalable, and user-friendly classification solutions that can adapt to the dynamic nature of big data. This paper aims to provide a clear understanding of data classification, highlighting the ongoing challenges, and suggest directions for future research in this vital field.

Keywords: Data classification, Data security, Encryption, Big Data, Homomorphic Encrytion

1 INTRODUCTION

Organizations are faced with the task of managing vast quantities of data due to the digitalization process. There are laws in place to protect data privacy, given the sheer volume, it is not feasible to manually identify private documents. Hence, it is of utmost importance to develop a system that can efficiently classify documents and aid in identifying papers containing sensitive information [1]. Guarantee the utmost security and confidentiality of sensitive documents.

While security measures often focus on protecting static data, they often struggle to effectively handle large amounts of data that surpass the capabilities of current databases. Threats that pose a risk to the organization's reputation and the trust of its stakeholders consider private data as a valuable target.

By analyzing users email habits and interests, big data can contribute to the development of phishing websites, which may compromise the security of their correspondence. Ensuring the security of big data is of utmost importance in cloud computing systems, as it can have a significant impact on the reputation of businesses [2]. Ensuring the security and privacy of data is crucial when handling large volumes of information, encompassing storage, management, analysis, and transfer. An effective big data security solution should focus on maintaining the confidentiality, integrity, and availability of data to provide the highest level of protection.

Ensuring the security and confidentiality of data to safeguarding sensitive information from unauthorized access. In order for data to be authentic, it needs to come from a trustworthy source and stay unchanged. Encryption and signature systems are essential for maintaining confidentiality and verifying authenticity [3]. Attribute-based encryption (ABE) is an advanced encryption method that enables secure communication to multiple recipients and also provides access control. Unlike traditional digital signature schemes, attribute-based signatures (ABS) provide a solution to the problem of revealing the identity of the signer. Using ABS, users can quickly verify the authenticity of a document by ensuring that the signature attributes adhere to the necessary policy. ABSC (attribute-based signcryption) combines the properties of ABE and ABS by merging them in a unified phase. In addition, the costs associated with computing and transmitting ABSC are lower compared to using ABE and ABS directly. ABSC is an extremely dependable cryptographic source that offers strong data security [4].

Ensuring data security is crucial in large data analysis, as it has been proven to be a complex problem. When analyzing large data sets, it is crucial to consider three important factors: availability, confidentiality, and integrity. Ensuring confidentiality is of utmost importance in order to protect sensitive information and restrict access to only authorized individuals with the necessary clearance. Integrity allows authorized users to modify, edit, update, and delete data. Ensuring availability guarantees that data is easily accessible and usable. Many businesses possess a significant volume of sensitive data that they store, collect, and process in a centralized manner. It is highly desirable to avoid storing large amounts of sensitive data in a single location to minimize the risk of data loss, sabotage, and hacking. This covers a broad spectrum of data, including trade and financial information, personal data of patients and consumers, and other valuable data. In addition, a malicious attack has the potential to create a vulnerability that may lead to a denial of service. By utilizing the values of risk metrics, we offer an approach to categorizing risk assessment. This approach guarantees the reduction of possible risks linked to big data and enables effective risk management. This approach considers crucial factors such as assets, vulnerability exposure, threat intensity, and probability of threats [5].

Securing sensitive information is crucial, and data privacy is essential for achieving this goal. By incorporating encryption and hidden data access control mechanisms, data security can be greatly enhanced, guaranteeing the confidentiality of sensitive information. However, dealing with large amounts of data can present challenges when it comes to implementing security measures such as access control and encryption. [6] investigates into a smart security approach that focuses on closely monitoring users who engage in unusual behaviors. This method utilizes a cutting-edge and adaptable user log system and software to guarantee dependability. It also includes a package that integrates the behavioral cues given by users to enhance security.

Once a user's activity exceeds a specific threshold, they are flagged as a critical user due to suspicious behavior. The behavior is examined using a model to ascertain whether it is suspicious or merely motivated by curiosity. In addition, the self-assuring architecture incorporates a library that can detect and store keywords linked to unusual user behavior, along with crucial logs pertaining to these keywords.

Nevertheless, the industry is facing a major challenge when it comes to data privacy and confidentiality. This is mainly due to the concerns surrounding the trustworthiness of cloud servers. According to Gardner's research, a significant number of users, around 70%, express worries about the security of their data when it comes to sharing it on cloud servers. [7] have proposed pioneering cloud-based encryption techniques to address this issue. Although these methods do tackle the issue of data secrecy, the challenge of ensuring secure data processing remains a persistent concern in these systems. For optimal processing and calculation, it is essential to have the required access to the decryption keys for the cloud server. The relationship between ciphertext and plaintext operations in HE systems allows for the presence of homomorphic characteristics. As an illustration, the RSA2 technique demonstrates multiplicatively homomorphic properties, while the Paillier cryptosystem [8] demonstrates additively homomorphic properties. Despite the existence of partial homomorphic encryption (PHE) systems, the development of a fully functional homomorphic encryption (FHE) plan has remained a persistent challenge in recent years.

Homomorphic encryption is a highly effective method that enables secure computations to be performed on encrypted data, while maintaining the integrity of the encryption process. Traditional encryption methods rely on the exchange of keys among users to facilitate secure transmission of encrypted messages. However, these methods do raise concerns regarding the preservation of confidentiality. Whoever possesses the necessary access, whether it be a user or a service provider, has limited entry to the data. Ensuring the security of sensitive information is crucial, particularly when using a widely-used cloud service platform. The owner desires to restrict access to the information [9].

However, if the key is not kept confidential, external entities will have the ability to access the encrypted data. In addition, even if users decide to discontinue the use of unreliable cloud services, individuals such as employees, suppliers, and workers will still have access to the chosen user items for a prolonged duration. Consider using HE, a specialized encrypted pattern that enables third parties to work directly with encoded data, eliminating the requirement for prior decoding. Despite being in existence for over a decade, Fully Homomorphic Encryption (FHE) was introduced by Craig Gentry in 2009 as the pioneering encryption technique that achieved full functionality. It was designed to work specifically with encoded data. While this achievement is certainly impressive, there are other advanced applications that indicate the necessity for additional improvements to guarantee compatibility with Fully Homomorphic Encryption (FHE) on all platforms [10].

Problem Statement:

The primary problem addressed in this context is the challenge of maintaining data security and privacy in the face of rapidly expanding volumes of digital data. Traditional security mechanisms, primarily designed for fixed data, are proving inadequate for the dynamic and voluminous nature of bi data. This inadequacy presents significant risks, such as unauthorized access, data leaks, and cyber threats like phishing. Moreover, the reliance on cloud servers, often perceived as unreliable, exacerbates these concerns. Although Homomorphic Encryption (HE) offers a promising solution by enabling computations on encrypted data, thereby preserving privacy, it still faces limitations in terms of applicability and efficiency. The overarching issue is developing robust, scalable, and efficient data security methods that ensure confidentiality, integrity, and availability of data, while also being adaptable to the unique demands of big data and cloud computing environments.

2 Literature Survey

2.1 Data Security

Data security has become a major focus for researchers in recent times. However, many security measures focus on preventing attacks on static data, which may not be enough when dealing with large amounts of data that surpass the capabilities of current databases. Threats that challenge the trust and reputation of the company view personal information as a valuable target. As an illustration, the utilization of big data may result in the creation of phishing websites that exploit users' email habits and preferences, risking the security of their communications. Big data security is a significant concern for cloud computing platforms and can have detrimental effects on a company's reputation [11]. Table 1 shows the literature survey.

Table 1 Literature survey table

Ref	Method	Advantages	Disadvantages	Research Gap
[10]	Secure Data Outsourcing	Enhances privacy between cloud servers and users	Reliance on cloud servers, potential inefficiency	Need for improved cloud server reliability
[11]	Fully Homomorphic Encryption (FHE)	Enables computations on encrypted data	Still needs major improvements for broad use	Applicability and efficiency of FHE
[12]	Proprietary Encrypted Patterns	Solves confidentiality concerns in data sharing	Potential security risks in encryption process	Secure and efficient encryption techniques
[13]	Privacy-Preserving Mechanisms	Preserves data privacy	Increases computational	More efficient privacy-

			overhead and time	preserving solutions
[14]	CryptoNets	Allows encrypted predictions on cloud	Complex and potential information leakage	Enhanced security in neural network models
[15]	HE-based Framework for Big Data	Safeguards sensitive data	Response time issues, interaction complexity	More efficient and user-friendly solutions
[16]	BGN HE Techniques	Privacy-preserving for IoT applications	High complexity, potential data leakage	Improved IoT data security
[17]	Secure Cloud Computing Platform	Utilizes HE to protect user data privacy	Requires access to decryption keys	Secure data processing methods
[18]	Multi-modal Data Classification	Special treatment for higher variability of data	Performance issues with regular models	Enhanced classification models

Recent developments [12] have shown a notable rise in the field of secure data outsourcing to untrusted cloud servers. By implementing encrypted data, HE improves the privacy of data consumers and cloud servers. It has been used in different computing applications that ensure anonymity. For instance, [13] devised a solution that guarantees privacy in Internet of Things applications by employing communication-efficient techniques rooted in BGN HE. In their study, presented a framework that leverages HE to safeguard sensitive data in the age of big data. Furthermore, [14] presented a secure cloud computing system that employs HE to safeguard the confidentiality of user data.

The sensitive unpredictability of multi-modal data, as opposed to unimodal data, poses a distinct learning challenge that demands focused attention. Many classification models struggle to accurately match multi-modal data. Consequently, the classifier's performance experienced a notable decline. To effectively handle the classification of multi-modal data, it is essential to improve the performance of the classification model [15]. Homomorphic encryption is a highly effective method that allows for secure computations to be conducted on encrypted data. Traditional encryption methods depend on the sharing of keys between parties to ensure the secure transfer of encoded communications. However, there are valid concerns about maintaining confidentiality when using these methods. Individuals and organizations with the necessary authorization are given restricted data access. Ensuring the security of sensitive data can be a daunting task when depending on widely-used cloud service providers.

The owner desires to limit access to the information, however, if the key is not kept confidential, unauthorized individuals will be able to obtain access to the encrypted data. Furthermore, even if users opt to stop using unreliable cloud services, individuals such as employees, suppliers, and workers will still have prolonged access to the selected user items [16-17].

The inefficiency of the FHE scheme is due to its dependence on identifiable matrix problems. For instance, if an integer message undergoes encryption and encounters a minor error during decryption, the resulting message could contain disturbing content. This is done to improve security during the encryption process. Mastering this problem becomes easy when employing simple encryption and decryption techniques. Make sure the message is aligned with the correct element [18].

The model ensures a secure transfer of input to the nonlinear transformation, decrypts the owner's data, performs the conversion, encrypts the output, and transmits it again. Unfortunately, the individual responsible for the data encounters extra obstacles and endures extended periods of waiting as a result of these interactions. Furthermore, the model provides extensive information. To resolve this problem, security measures have been put in place that utilize arbitrary execution commands. Nevertheless, the process being described does not necessitate intricate connection diagrams. After the encryption process is finished, the data is transmitted securely to the authorized recipient. After the calculations are finished, the model sends the prediction (coding) again [19]. Securing data stored in the cloud is crucial, and this can be accomplished by implementing proven encryption methods. In cryptographic systems, the data can only be decrypted by the recipient or a second party who possesses the private key of the sender. When a user submits a request to the cloud's virtual environment for secure and efficient computing on their data, they must provide a private key. Processing is considered finished once the data has been successfully decoded. As computations become more complex, the risk of the key being exposed also increases [20]. In this situation, the user will have to make changes to the key. In addition, it is crucial for both parties to possess a compatible secret key when employing symmetric encryption. In addition, it is important to make a copy of the confidential key in case there is a potential security breach. Security vulnerabilities can result in extended processing times and heightened computational requirements, irrespective of overall performance. Privacy protection mechanisms are capable of effectively addressing this concern. Encryption is a robust mechanism that guarantees the highest level of security and privacy for data [21].

The presented research encompasses a diverse array of methodologies within the realm of big data security and analytics. Papers [22-25] collectively propose an integrated methodology for classifying and securing big data before mobility, duplication, and analysis, underscoring the importance of securing data mobility through risk-based classification. In [26], recent works are summarized in tables, providing valuable insights into cybersecurity trends and open challenges. Papers [27-29] introduce a security-by-design framework for deploying big data frameworks over cloud computing, employing systematic security analysis and an automated assessment framework. [30] Reviews security implementations in leading database models, with a specific focus on security and privacy attributes. [31] Presents a novel approach to big data storage security, leveraging blockchain technology and a flexible finality condition-based highway protocol. Papers [32-35] propose a secure data protection method in the cloud, emphasizing efficient partitioning, partial decryption, and analysis. [36] Evaluates classification techniques on changing data using phishing datasets, considering both unbalanced and balanced schemes. [37] Suggests a method for picking random data chunks for classification, utilizing weak KNN (K- Nearest Neighbour) classifiers and majority voting. Papers [38-40] introduce a deep learning-based model for classifying encrypted mobile traffic data. Notable trends in security threats and defensive techniques in machine learning are discussed, and a blockchain-based auditable privacy-preserving data classification scheme for IoT is proposed. The paragraph further reiterates an integrated methodology for big data classification and security. A comprehensive study on adversarial attacks and defenses for fault classifiers in data-driven FDC (Fault Detection & Classification) systems is presented [41-44]. Additionally, an MDSA (Machine learning based Digital Signature Algorithm) algorithm for real-time measurement data classification is proposed, incorporating feature extraction and machine learning. Various methods are analyzed with concrete examples, elucidating different usage scenarios. Another paper introduces an identity-based dynamic data auditing scheme supporting dynamic operations, analyzing the risk of security and privacy leakage in medical big data, and establishing a risk indicator system [45-46]. Finally this delves into early economic security analysis through big data, exploring risk early-warning methods and achieving scientific economic decision-making. Table 2 shows the data classification table.

Table 2 data classification survey table

Ref	Method	Advantages	Disadvantages	Research Gap
[42]	Automatic Document Classification	Efficient handling of large volumes of data	Potential misclassification issues	Accuracy and reliability in diverse data sets
[43]	Big Data Security Mechanisms	Protects fixed data against threats	Insufficient for dynamic nature of big data	Scalable and adaptable security for big data
[44]	Phishing Detection Techniques	Targets email-based threats	Limited to specific types of threats	Broader applicability and detection capabilities
[45]	Cloud Computing Security Solutions	Enhances organizations' reputation and trust	May not be fully effective against all threats	Comprehensive security solutions for cloud systems
[46]	Data Storage and Management	Ensures data confidentiality, integrity, availability	Concentration of data increases risk of attacks	Secure decentralized data storage methods

[47]	Intelligent-driven Security Model	Monitors users for abnormal behaviors	No protection against data loss and leakage	Improved data loss and leakage prevention
[48]	NP-Hard Data Analysis	Addresses crucial issues in big data analysis	Computationally intensive	More efficient computational methods
[49]	Confidentiality Techniques	Protects big data from unauthorized access	Implementation difficulties for big data	Feasible and efficient big data confidentiality
[50]	Risk Metrics-Based Assessment	Promotes risk management	May not cover all potential risks	Comprehensive risk assessment for big data

Ensuring security and data privacy is crucial when utilizing machine learning for tasks that involve sensitive, financial, or medical data. There may be restrictions on the use of cloud-based machine learning technology for certain tasks due to ethical and legal considerations [47]. Proposed a technique for converting trained neural networks into CryptoNets. This allows the data owner to securely transmit encrypted data to network hosts, particularly cloud services. With the implementation of encryption protocols, the data remains secure and inaccessible to the cloud without the correct keys. In addition, the cloud service has the capability to deliver the data in an encoded format. This is achieved by using a neural network to generate encoded predictions for the encoded data. Accessing the encrypted forecasts and deciphering them requires possession of the secret key. Appropriately, it is not possible to access the forecast generated by the cloud facility or the raw data itself [48].

There are three methods that can be used to identify large data packets: weighted Euclidean distance, radial integral kernel function SVM (Support Vector Machine), and dimensionality reduction. The SVM lacks the ability to handle multiple classifications, and constructing the model can be a time-consuming task. The algorithm effectively resolved these concerns [49]. An advanced framework for cloud platforms that utilizes cutting-edge reinforcement learning techniques, with a strong focus on prioritizing privacy as a key concern. The proposed approach highlights the importance of gaining knowledge from errors and utilizes a cryptosystem for Fully Homomorphic Encryption (FHE). When analyzing and assessing the performance of the proposed architecture for privacy-preserving reinforcement learning, various intelligent service scenarios in the cloud are considered [50].

2.2 Research gap

- **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 Fully Homomorphic Encryption (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.

- **Efficient Privacy-Preserving Solutions**: Designing more efficient privacy-preserving mechanisms that reduce computational overhead and processing time while maintaining high levels of data privacy.
- **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.
- **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.
- **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.
- **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.

3 RESULT

The survey paper provides a thorough examination of recent advancements in data classification techniques within the domain of big data and cloud computing, emphasizing the critical importance of robust mechanisms for managing data security and privacy. From traditional encryption methods to cutting-edge Homomorphic Encryption (HE), various methodologies were explored to ensure data confidentiality, integrity, and availability. Key findings underscored the necessity for innovative solutions capable of adapting to the dynamic nature of big data, while also addressing challenges such as improving cloud server reliability, enhancing the efficiency of Fully Homomorphic Encryption (FHE), and designing user-friendly encryption techniques. Ultimately, the paper identifies significant research gaps and challenges, offering clear directions for future exploration to develop more effective data classification systems and bolster data security and privacy in the digital era.

4 CONCLUSION

In conclusion, this paper highlights the pivotal role of advanced data classification techniques in managing the ever-growing challenges of data security and privacy. Through an extensive review of various methodologies, from traditional encryption to cutting-edge Homomorphic Encryption (HE), it becomes evident that while significant progress has been made, substantial gaps remain in achieving optimal efficiency, scalability, and user-friendliness. The paper underscores the need for innovative solutions that can adapt to the complex and dynamic nature of huge data, ensuring not only the confidentiality and integrity of data. The identified research gaps and challenges set a clear agenda for future exploration in this field, guiding researchers towards developing more robust, efficient, and comprehensive data classification systems. This paper ultimately serves as a resource for driving forward the crucial task of enhancing data security and privacy in the digital world.

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] 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.

[8] 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.

[9] 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.

[10] H. Kwon, "Friend-Guard Textfooler Attack on Text Classification System," in IEEE Access, doi: 10.1109/ACCESS.2021.3080680.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

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

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] 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.

[34] 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.

[35] 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.

[36] 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.

[37] 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.

[38] 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.

[39] 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.

[40] 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.

[41] 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.

[42] 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.

[43] 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.

[44] 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.

[45] 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.

[46] 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.

[47] 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.

[48] M. Ali, M.-R. Sadeghi, X. Liu, Y. Miao, and A. V. Vasilakos, “Verifiable online/offline multi-keyword 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.

[49] Munjal, K., Bhatia, R. A systematic review of homomorphic encryption and its contributions in healthcare industry. Complex Intell. Syst. 9, 3759–3786 (2023). https://doi.org/10.1007/s40747-022-00756-z

[50] 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.