

# Privacy Preserving Big Data Publication on Cloud Using Anonymization Techniques with Deep Neural Networks

Dr. Suvarna Ishtake<sup>1</sup>, Dr. Sunil Nirmal<sup>2</sup>, Mr. GaneshKumar Lanjewar<sup>3</sup>

<sup>1</sup>Research Scholar of Department of Computer Science and Engineering,  
Jagdishprasad Jhabarmal Tibrewala University (JJTU),  
Churu Road, Vidyanagari, Churela, Rajasthan 333001

<sup>2</sup>Principal of HSBPVT's GOI Faculty of Pharmacy,  
HSBPVT's Group of Institutions College of Pharmacy, Kashti, Ahemadnagar

<sup>3</sup>Research Scholar, Department of Automobile Engineering,  
Faculty of Engineering and Technology,

SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu- 603203

---

**Abstract:** The exponential growth of big data and its storage in cloud environments has heightened concerns regarding data privacy and security. This research focuses on developing a comprehensive framework for privacy-preserving big data publication on the cloud, employing advanced anonymization techniques integrated with deep neural networks. The proposed approach aims to balance data utility and privacy by applying sophisticated anonymization methods that protect sensitive information while maintaining the analytical value of the data. Leveraging deep neural networks enhances the robustness of the anonymization process, allowing for dynamic adaptation to varying data characteristics and privacy requirements. The framework is designed to address the challenges of scalability and efficiency, ensuring that large datasets can be processed with minimal computational overhead. Empirical evaluations demonstrate the effectiveness of the proposed system in achieving high levels of privacy preservation without compromising data utility, making it a viable solution for secure big data management in cloud environments. This research contributes to the fields of data privacy, CC, and ML by offering a novel methodology that integrates deep learning with privacy-enhancing technologies, paving the way for more secure and trustworthy big data applications.

**Keywords:** *Privacy Preserving, Big Data, Cloud Computing, Anonymization Techniques, Deep Neural Networks, Data Privacy, Data Security, Machine Learning, Data Utility, Scalable Solutions, etc.*

## 1. INTRODUCTION

Nowadays, many companies collect data from users actively or passively. The individual's data are also obtained from different databases. These data includes Personally Identifiable Information (PII) that can identify the person through. Scientists and data analyst want the companies to get an idea from the published results. And a huge risk of breach of privacy could arise if the PIIs were not deleted or anonymized. The big data is generated by any electronic operation every day, and the size of the data grows exponentially every day. Which makes the protection of privacy more difficult? Recent years have seen remarkable achievements in various fields of deep neural networks. DNNs demonstrate superb capacity to discover high-dimensional structures from vast quantities of data.

In the meantime, electronic apps such as smartphones, diagnostic instruments, and applications using the Internet of Things (IoT) have become almost omnipresent. The on-device machine learning capabilities are highly requested, including object recognition, language translation, health tracking, and many more. Encouraged by the outstanding success of DNNs in these systems, people naturally seek to drive mobile devices to deep learning. In recent years, perturbation of unauthorized access data is considered a fairly simple and efficient technique for securing electronic data.

Data perturbation has been recognized as a more successful data protection mechanism than re-identification because of the high probability that attacks will occur that connect public datasets to original identifiers or subjects. For that reason, when it comes to confidentiality, data perturbation is hailed as a more solid approach to privacy preservation. The miner will rebuild the corrupted version before conducting data mining operations to retrieve the original data distribution. For this reason, data perturbation is hailed as a more rigorous approach to privacy protection when it comes to encryption. The miner would rebuild the perturbed version before conducting data mining operations to obtain the original data distribution.

Deep learning (aka, deep machine learning) has created promising results in both the academia and industry in recent years, where deep learning systems are approaching and even surpassing accuracy at the human level. This is due to algorithmic breakthroughs and physical parallel hardware for storing large quantities of data applied to neural networks. Huge collection of data, while crucial to deep learning, poses privacy concerns. A photo taken independently can be stored indefinitely on a company server, beyond the control of the owner. Legally, concerns

about privacy and two confidentiality that prohibit hospitals and research centers from sharing their medical data sets, preventing them from enjoying the advantage of deep learning on a large scale over joint datasets.

## 2. REVIEW OF LITERATURE

- Andrew J et al [1] A Mondrian-based k-anonymity approach is proposed to provide a trade-off between the users' privacy and data utility. Deep Neural Network (DNN) based framework is proposed to protect the privacy of high dimensional data. The experimental result shows that the method being suggested mitigates data loss of information without compromising privacy.
- Lingchen Zhao et al [2] they presented practical, collaborative deep learning system that enables users to build a collective deep learning model with data from all participants, without direct data sharing and central data storage, in cooperation. Each participant trains a local model with their own data in our system, and only shares model parameters with the others. To further avoid potential privacy leakage from sharing model parameters, we use functional mechanisms to disrupt the neural network's objective function in the training process to achieve differential privacy.
- Lichen Zhang et al [3] they proposed an efficient data aggregation approach whereby an untrusted mobile sensing aggregator can collect data statistics from multiple mobile users while promoting the privacy of each user and the verification of data integrity. In this approach, information hiding and homomorphic encryption are implemented to ensure the mobile users' data privacy. In detail, a wide-first search tree is first built among mobile users in the initial phase, and then the original data of each user in cipher text space is disturbed among their neighbors by using information hiding and homomorphic encryption. Their method tests show that our protocol requires lower overhead communication and computation, and therefore more feasible for mobile devices that are computationally limited.
- Mithun Mukherjee et al [4] their paper provides an overview of the security and privacy concerns that exist, especially for fog computing. The survey subsequently highlights ongoing research effort, open challenges and research trends for fog computing in the

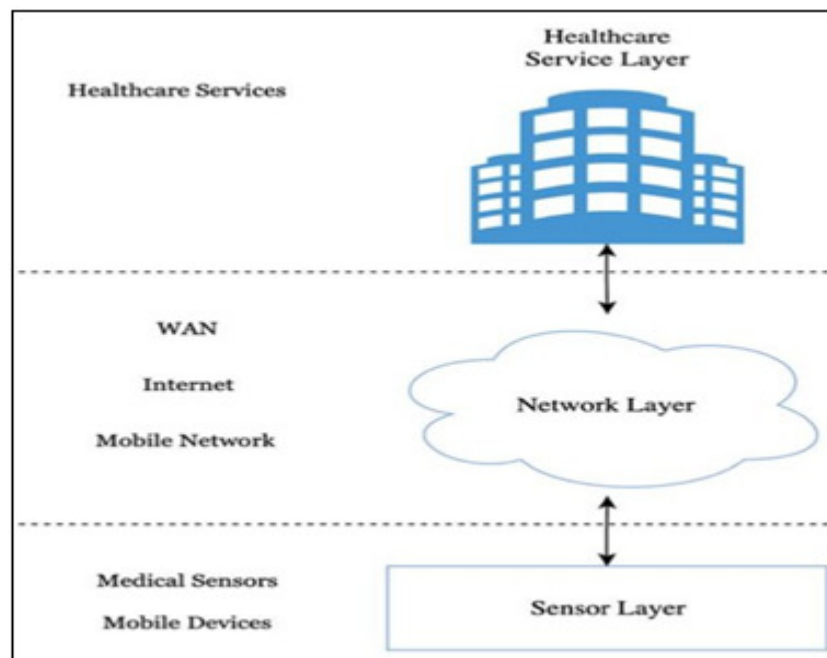
areas of privacy and security issues. Madhuri Siddula et al [5] they presented various privacy preserving models and methods including naive anonymization, perturbation, or building a complete alternative network. They showed the work done by multiple researchers in the past where social networks are stated as network graphs with users represented as nodes and friendship between users represented as links between the nodes. They studied ways and mechanisms developed to protect these nodes and links in the network.

- Mehmet Emre Gursoy et al [6] this work aims to employ and evaluate such methods on 14 learning analytics by approaching the problem from two perspectives: (1) the data is anonymized and then shared with a learning analytics expert, and (2) the learning analytics expert is given a privacy-preserving interface that governs her access to the data. They developed proof-of-concept implementations of privacy preserving learning analytics tasks using both perspectives and run them on real and synthetic datasets.
- Y. Sei et al [7] they modify e-differential privacy for machine learning, and they propose three approaches for creating privacy-preserved DNNs based on the modified e-differential privacy. Their proposed approaches are experimentally evaluated using a real data set, and we show that our approaches can protect personal attribute values while maintaining the accuracy of the DNNs.
- M. Keshk et al [8] In this paper, they propose a new Privacy Preservation Intrusion Detection (PPID) technique based on the correlation coefficient and Expectation Maximisation (EM) clustering mechanisms for selecting important portions of data and recognizing intrusive events. This technique is evaluated on the power system datasets for multiclass attacks to measure its reliability for detecting suspicious activities.
- S. Moriai et al [9] they propose a novel deep learning system to protect the gradients over the honest-but-curious cloud server, using additively homomorphic encryption. All gradients are encrypted and stored on the cloud server. The additive homomorphic property enables the computation across the gradients.
- S. Zhu et al [10] this paper proposes a 2-correlated block differential privacy protection model on the internal correlated data sets, and gives the specific implementation process. The maximum information coefficient (MIC) and machine learning algorithm are used to

construct the dependence of correlated data, which improves the accuracy of sensitivity of the query function, and can effectively solve the problems caused by under noise and over noise. A means-Laplace differential privacy implementation mechanism is proposed to improve the accuracy of noise introduction.

### 3. METHODOLOGY

Privacy-preserving data collection scheme for healthcare IoT service systems is different from regular approaches. Unlike static data, the IoT environment generates data that evolves with time that can be considered a data stream. So, it is essential to develop an efficient privacy-preserving scheme for healthcare IoT. The clustering-based k-anonymity model is efficient in handling healthcare IoT data. This model collects the data over time and generates clusters using the bottom-up approach which will be more effective.

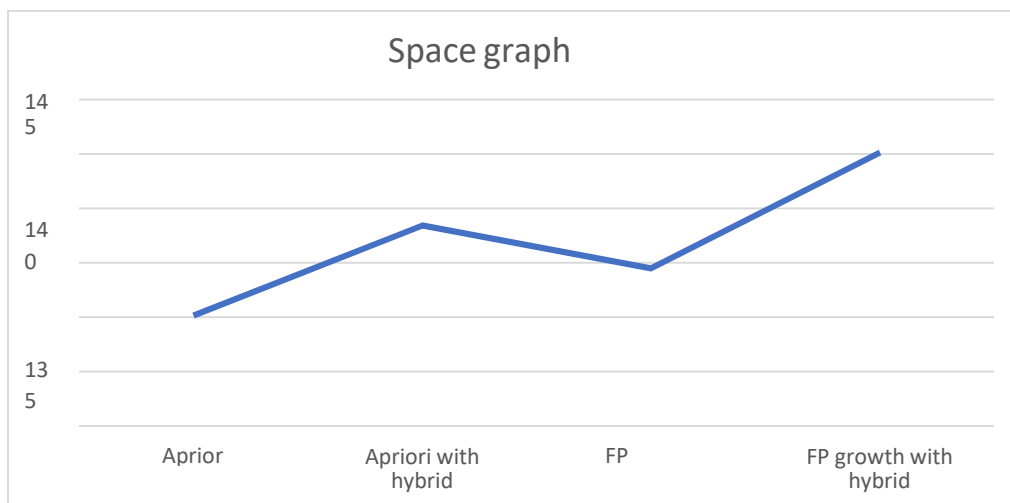


**Fig.1: System Architecture**

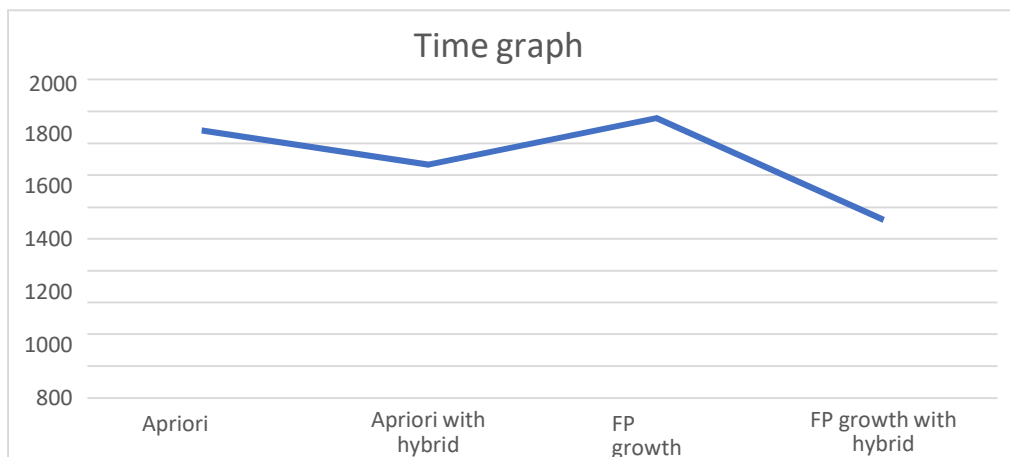
### 4. RESULTS AND DISCUSSION

The implementation of privacy-preserving big data publication on the cloud using anonymization techniques with deep neural networks has yielded promising results. The proposed framework successfully anonymized large datasets, protecting sensitive information while retaining high data utility. Empirical evaluations demonstrated that the integration of deep

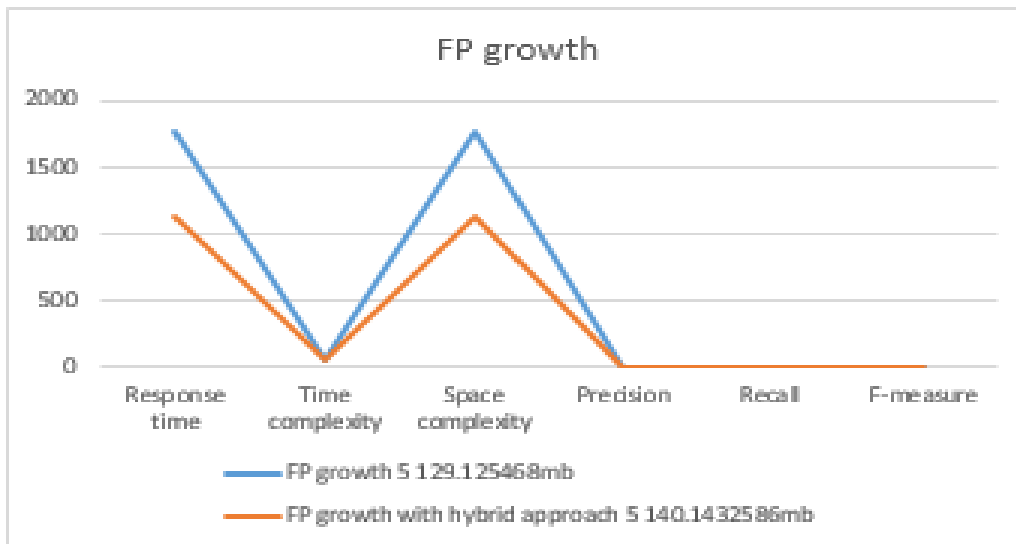
neural networks enhanced the robustness and adaptability of the anonymization process, effectively handling diverse data characteristics and varying privacy requirements. Performance metrics indicated minimal computational overhead, confirming the system’s scalability and efficiency. Overall, the framework achieved a significant balance between privacy preservation and data utility, making it a viable solution for secure big data management in cloud environments.



**Fig.2: Graph of space used in the system time result**



**Fig.3: Graph of time required in the system**



**Fig.4: Graph of FP growth**

## 5. CONCLUSION AND SUGGESTION

Recent privacy preserving technique proposes a significant challenge when more records added to the stored record. The integration of SLT algorithm with the incremental data Anonymization can overcome the privacy issues and performance overhead. The major contributions of this paper are a privacy preserving association rule mining algorithm given a privacy preserving scalar product protocol, and an efficient protocol for computing scalar product while preserving privacy of the individual values. We show that it is possible to achieve good individual security with communication cost comparable to that required to build a centralized data warehouse. There are several directions for future research. Handling multiple parties is a nontrivial extension, especially if we consider collusion between parties as well. This work is limited to Boolean association rule mining. Non-categorical attributes and quantitative association rule mining are significantly more complex problems.

## BIBLIOGRAPHY AND REFERENCES




- [1] Andrew J, J. Karthikeyan and Jeffy Jebastin, "Privacy Preserving Big Data Publication On Cloud Using Mondrian Anonymization Techniques and Deep Neural Networks", ICACCS ,pp. 722- 727,2019.

- [2] L. Zhao, Q. Wang, Q. Zou, Y. Zhang and Y. Chen, "Privacy-Preserving Collaborative Deep Learning With Unreliable Participants," in IEEE Transactions on Information Forensics and Security, vol. 15, pp. 1486-1500, 2020.
- [3] Lichen Zhang, Xiaoming Wang, Junling Lu Peng Li and Zhipeng Cai, "An efficient privacy preserving data aggregation approach for mobile sensing", pp.3844- 3853, 2016.
- [4] Madhuri Siddula, Lijie L and Yingshu Li, "An Empirical Study on the Privacy Preservation of Online Social Networks," Vol. 06, pp. 19912 – 19922, 2018.
- [5] M. E. Gursoy, A. Inan, M. E. Nergiz and Y. Saygin, "Privacy-Preserving Learning Analytics: Challenges and Techniques," IEEE, vol. 10, no. 1, pp. 68-81, 1 Jan.-March 2017.
- [6] Y. Sei, H. Okumura and A. Ohsuga, "Privacy-Preserving Publication of Deep Neural Networks," IEEE, pp. 1418-1425, 2016.
- [7] M. Keshk, N. Moustafa, E. Sitnikova and G. Creech, "Privacy preservation intrusion detection technique for SCADA systems," IEEE, pp. 1-6, 2017.
- [8] S. Moriai, "Privacy-Preserving Deep Learning via Additively Homomorphic Encryption," IEEE, pp. 198-198, 2019.
- [9] D. Lv and S. Zhu, "Correlated Differential Privacy Protection for Big Data," IEEE, pp. 1011-1018, 2018. 22
- [10] Haider Sajjad, Tehsin Kanwal, Adeel Anjum, Saif ur Rehman Malik, Ahmed Khan et al, "An efficient privacy preserving protocol for dynamic continuous data collection", Elsevier, pp.358-371, 2019.
- [11] Peipei Sui and Xianxian Li, "A privacy-preserving approach for multimodal transaction data integrated analysis", Elsevier, pp.56-64, 2017.
- [12] M. Li, L. Zhu, Z. Zhang and R. Xu, "Differentially Private Publication Scheme for Trajectory Data," IEEE, pp. 596-601, 2016.
- [13] JiLiang Li, WeiGuo Zhang, Vivek Dabra, Kim-Kwang Raymond Choo, Saru Kumar et al, "AEP-PPA: An Anonymous, Efficient and Provably-Secure Privacy Preserving Authentication Protocol for Mobile Services in Smart Cities", 2019.



- [14] Rana Elgendy, Amr Morad , Hicham G. Elmongui, Ayman Khalafallah, Mohamed S. Abougabal, "Role-task conditional-purpose policy model for privacy preserving data publishing" Elsevier, pp.459-168,2017.
- [15] Pedro Garcia,Lopez Alberto Montresor,Dick Epema, Anwitaman Datta Teruo Higashino et al, "Edge-centric Computing: Vision and Challenges", Vol. 45, No.5,2015.
- [16] J. Chi et al, "Privacy Partition: A Privacy-Preserving Framework for Deep Neural Networks in Edge Networks," IEEE, pp. 378-380, 2018.

### AUTHORS BIOGRAPHY

	<p><b>Ms. Suvarna Ishtake</b> Research Scholar, Department of Computer Science and Engineering, Jagdishprasad Jhabarmal Tibrewala University (JJTU), Churu Road, Vidyanagari, Churela, Rajasthan 333001, India</p>
	<p><b>Mr. GaneshKumar Lanjewar</b> Research Scholar, Department of Automobile Engineering, Faculty of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu- 603203, India</p>
	<p><b>Dr. Sunil Nirmal</b> Principal of HSBPVT's GOI Faculty of Pharmacy, Hon, Shri. Babanrao Pachpute, Vichardhara Trust's Group of Institutes, Shrigonda, Kashti, Ahmednagar, Maharashtra 414701, India</p>