ISSN 1533-9211

DOI 10.17605/OSF.IO/YDGR4

A SHORT EVALUATION ON DIFFERENTIAL PRIVACY FOR PRIVACY PRESERVATION OF BIG DATA WITH FUTURE RESEARCH PERSPECTIVES

ANNIE SHERYL S [1], Dr. S.KEVIN ANDREWS[2] and Dr. RAJAVARMAN V.N[3]

[1]Research Scholar, Department of Computer Science and Engineering, Dr. MGR Educational and Research Institute, Chennai, Tamilnadu, India. Email: anniesherls17@gmail.com
[2]Professor, Department of Computer Science and Engineering, Dr. MGR Educational and Research Institute, Chennai, Tamilnadu, India. Email: kevin.mca@drmgrdu.ac.in
[3]Professor, Department of Computer Science and Engineering, Dr. MGR Educational and Research Institute, Chennai, Tamilnadu, India. Email: rajavarman.vn@drmgrdu.ac.in

Abstract

Owing to many communication technology devices, processes, and electronic information, such as the cloud, corporate data, social networks, internet activities, personal archives, and sensors, the information age has exponentially grown these types of data. Properly managing this enormous and diverse data is the most difficult problem. One of these significant and varied sorts of data is big data. The major big data problems are privacy and security because of its huge size. There is a risk because privacy could be revealed at any level in managing big data. The characteristics of unsuited methods of big data, such as anonymization and encryption, have been established to prevent the privacy of enormous datasets. Thus, this research discusses various literature works on differential privacy preservation for big data. It also reveals the intelligent protocols' methodologies, including machine learning and deep learning. Moreover, the survey part is also analyzed by exploring the chronological review of privacy preservation for big data, utilized big data, performance metrics for evaluation, and implementation platforms. At last, the challenges observed in traditional research works for privacy preservation using differential privacy for big data are discussed, and the future perspectives on promoting the utilization of big data are also explained.

Keywords: Privacy Preservation of Big Data; Differential Privacy; Methodologies of Big Data; Performance Metrics; Implementation Tools; Dataset Details; Algorithmic Classification

I. INTRODUCTION

Analytics in big data healthcare offers several advantages, and tremendous changes can alter healthcare. But, it also poses several obstacles and difficulties in differential privacy. Indeed, as information and communication technology advance quickly, society is raising worries about the privacy and security of huge amounts of healthcare data [1]. Daily, a massive amount of raw data is quickly created from several sources, including online stores, transportation, and social media websites [2]. Differential privacy offers a useful solution to protect the attackers and provides a quantitative privacy analysis of big data risk.

The major issues impeding the expansion of big data are privacy preservation. Privacy-Preserving Data Publishing (PPDP) is one of the many uses of big data processing. It has drawn much interest from businesses and academics [3]. To guarantee privacy against the leaking of sensitive information, data is cleaned up using privacy preservation techniques before it is available to the general public. Differential privacy assures the addition or deletion of a single record cannot impact the results of any study [4]. Thus, a rigorous mathematical definition of privacy needs has emerged among the current privacy mechanisms. While these privacy-preserving techniques stop consumers' information from being compromised, they always result in data loss for the utility. The privacy-utility trade-off dilemma tries to maximize the data utility constrained by privacy limitations. It is undoubtedly one of the most significant difficulties. Numerous research is conducted about the trade-off issue, which is not unique to big data. It is often separated into two groups [5]. The first step is to observe the pros and cons of various privacy-preserving technologies, and the second is to select the best privacy setting to maximize utility. It is usually assumed that these data sets problems are independent of one another in terms of privacy. Hence, this survey examined the challenges of differential privacy preservation of big data for future researchers. This survey over differential privacy preservation of big data has major contributions like:

a) To establish differential privacy preservation of big data with a collection of information from the prior privacy preservation techniques of big data to secure the data and to predict the patterns accurately.

b) To conduct a survey over differential privacy preservation of big data over analyzing implementation tools, utilized protocols, chronological review, and performance metrics.

c) To tackle the limitations over the previous differential privacy preservation of big data to develop improved privacy preservation of big data.

This survey on differential privacy preservation of big data consists of the remaining sections, as expressed below. The reviews of the current works are explained in Part II. The tools and algorithmic classification techniques are discussed in Part III. The research gaps and challenges are mentioned in Part IV. At last, the conclusion is summarized in Part V.

II. LITERATURE REVIEW ON DIFFERENTIAL PRIVACY PRESERVATION METHODOLOGIES OF BIG DATA

a) Related Works

In 2022, Keshk et al. [6] have created a purpose for big data component analysis in preventing unauthorized access to sensitive data. The Independent Component Analysis (ICA) method was utilized for modifying raw cyber-physical system (CPS) data while retaining its usefulness as data. The process was tested using the dataset of power CPS, and the findings showed that the method achieved a greater level of privacy protection than four other privacy-preservation methods. In 2022, Wu et al. [7] have analyzed medical data from data transfer and data sharing data collecting to address the main issues with privacy protection and developed the MNSSp3 privacy protection sharing platform. It accomplished the separation of data and users to ensure the security of medical data. To provide user mining techniques, this platform focuses on sharing security and transmitting medical big data. In 2019, Chamikara et al. [8] have presented the nonreversible perturbation method PABIDOT, which was effective and scalable for protecting huge data privacy through optimum geometric transformations. When compared to

ISSN 1533-9211

DOI 10.17605/OSF.IO/YDGR4

related privacy-preserving algorithms, experiments revealed that PABIDOT outperformed them following accuracy, attack resistance, execution speed, and scalability in large-scale privacy-preserving data categorization. In 2021, Zhao et al. [9] have analyzed a federated learning strategy that was anonymous and protected privacy for the mining of industrial big data. Comparing our results to other schemes provided the security analysis and performance assessments. By exchanging fewer parameters between each participant and the server, the suggested technique lowered leakage in privacy.

In 2018, Yang et al. [10] have suggested a framework for crowdsensing approaches that fulfilled differential privacy and offered stringent worker location protection. Extensive testing demonstrated that the suggested solution offered a rigorous guarantee of anonymity and vastly enhanced the performance. In 2022, Zhang [11] have examined a scalable method based on Map Reduce, a cutting-edge data processing paradigm, for big data multidimensional anonymization. Recursive computing was achieved using a tree indexing structure. Experimental findings using actual data showed that this method could scale multidimensional anonymization much more easily than other approaches. In 2019, Chamikara et al. [12] have suggested a novel data perturbation technique based on Chebyshev interpolation and Laplacian noise, called Secure and Efficient Data Perturbation Algorithm Employing Local Differential Privacy (SEAL), which offered a fair balance between privacy and usefulness with great efficiency and scalability. One example of this flexibility was the quantity of additional noise. In 2020, Chen et al. [13] have suggested a new approach, Recurrent Neural Network for Dynamic Trajectory Privacy Protection (RNN-DP), for differential privacy based on the protection of dynamic trajectory privacy. The experimental findings showed great performance in data availability and privacy protection for dynamic trajectory data compared to the present methods. In 2019, Yan et al. [14] have created an adaptive density grid structure to cluster the massive location that not only minimized the uniform assumption mistakes but also prevented noise errors by a high number of vacant nodes. Comparative studies demonstrated that the hierarchical differential privacy hybrid decomposition approach significantly improved the precision of regional counting inquiries. In 2017, Phan et al. [15] have concentrated on creating a differentially Private Convolutional Deep Belief Network (PCDBN), also known as a Convolutional Deep Belief Network (CDBN). The theoretical studies demonstrated that it also obtained the approximate polynomial representation's sensitivity and error boundaries. Therefore, it was possible for maintaining differential privacy in CDBNs. In 2021, Yang et al. [16] have provided two techniques with differential privacy for calculating the persistent point traffic volume and the persistent point-to-point traffic volume. The efficiency of the suggested systems was demonstrated by the experimental findings using data from actual transportation traffic locations. In 2022, Shailaja et al. [17] have suggested a new PPDM paradigm with data sanitization and restoration as its two steps. The creation of the ideal key, which was utilized for separating the actual data from the sanitized data, was the first and most important strategy in the suggested privacy protection approach. The research difficulties were reduced, such as erroneous rule generation, information preservation rate, degree of change, and concealment failure rate. The suggested TU-WPA model performance was then compared to other widely utilized models. In 2020, Dagher et al. [18] have suggested a system in which the data provider uploaded an encrypted index of anonymized data to a DaaS service provider. Experiments with actual data showed that the suggested structure retained the usefulness of the data, effectively responded to range queries, and scaled with growing data volumes. In 2016, Lin et al. [19] have recommended a special privacy security system for BSN's sensitive large data. Histograms were transformed into an entire binary tree using the transformation technique called Haar Wavelet. According to experimental findings, the structure of the tree significantly lowered the overhead calculation while maintaining the user's differential privacy.

In 2015, Zhang et al. [20] have analyzed a proximity privacy model that allowed for the semantic closeness of sensitive values and numerous sensitive characteristics and have modeled the issue of local recoding as a proximity-aware clustering problem. For achieving greater scalability using data-parallel processing in the cloud, it constructed the algorithms utilizing Map Reduce. Numerous tests using actual data sets showed that the technique considerably increased the capacity for resisting near privacy breaches. In 2014, Soria-Comas et al. [21] have demonstrated that there was a synergy between differential privacy and k-anonymity: k-anonymity could enhanced the usefulness of differentially private replies to random queries. The anonymized output's overall analytical usefulness was boosted due to noise reduction. The theoretical advantages were demonstrated by the actual analysis of three datasets. In 2016, Zhang et al. [22] have suggested a Deep Computation Model (DCM) that protected privacy by shifting the costly processed cloud. The suggested model encrypted private information using the BGV encryption technique to safeguard it and then leveraged the cloud server to effectively run the high-order back-propagation algorithm on the encrypted information for the training of DCM. Additionally, by using more cloud servers, it was very scalable and ideal for massive data. In 2016, Lin et al. [23] have investigated a differentiated privacy security system for body sensor network large data. This plan offered privacy protection with more availability and dependability. Experimental findings showed that the suggested technique sufficiently interferes with large sensitive targets. In 2018, Zhang et al. [24] have presented a Double-Projection Deep Computation Model (DPDCM) for learning big data feature interactions. More significantly, it showed that promise for large data feature learning by efficiently increasing the efficiency of training parameters. This model replaced the hidden layers of the traditional DCM with layers of double-projection, which projected the raw input into subspaces of two distinct hidden layers. In 2022, Gosain and Chugh [25] have examined huge data privacy preservation techniques and discovered that differential privacy was a superior option. The three key privacy preservation techniques were consent, notice, and data anonymization in differential privacy. The field of data analytics has undergone a revolution because of big data. Big data's volume, diversity, and size make it impossible for securing traditional approaches.

b) Year-Wise Chronological Review

The differential privacy preservation techniques of big data based on year-wise implementation are depicted in Fig. 1. Several analyses are made chronologically over various privacy preservation techniques of big data. The period analyzed from 2014 to 2022 is covered by the evaluation of various privacy preservation techniques of big data. For this study, over 20 research papers are examined for evaluation. The year 2022 has made the greatest contributions to the survey of various privacy preservation techniques of big data. For 2015 and 2017, the contribution level is 5%. The level of contribution for the years 2014, 2018, 2020, and 2021 is 10%. For the years 2016 and 2019, the contribution level is 15%. At last, compared to the other levels on this survey, the contribution level for 2022 is 20% greater.

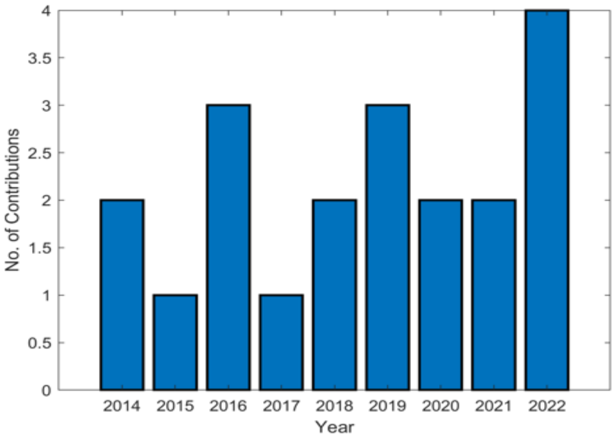

Fig. 1: Chronological analysis of various privacy preservation techniques of big data

c) Big Data Analysis of various privacy preservation techniques

Various analyses are done with the support of many datasets in various privacy preservation techniques of big data, which is presented in Table I.

Table I: Various Privacy Preservation Techniques of Big Data over Analysis of Datasets

Citation	Datasets Used in Big Data
[6]	CPS
[7]	Medical data
[9]	MNIST
[10]	Yelp and Simple Geo Places Dataset
[11]	Anonymization data
[12]	UCI data repository
[13]	real-world datasets
[15]	Location data
[16]	Health data
[17]	real transportation traffic dataset
[18]	MIRACL
[19]	Health data
[20]	Cloud dataset
[21]	Cloud dataset
[22]	Cloud dataset
[23]	Shimmer data
[24]	NUS-WIDE-14 and Animal-20
[25]	Anonymization data

The differential privacy preservation techniques of big data use datasets like medical data, cloud datasets, anonymization data, and real-world data, which are the most common datasets used for this survey. Hence, the various privacy preservation techniques of big data are evaluated with diverse data for protecting efficient information.

III. SHORT EVALUATION ON DIFFERENTIAL PRIVACY PRESERVATION OF BIG DATA TECHNIQUES

a) Methods of Big Data

In differential privacy preservation techniques of big data, the big data uses several protocol methods and techniques. Many researchers have analyzed various protocol methods and techniques. This survey includes machine learning and deep learning techniques and methods. Some big data techniques like machine learning, deep learning, and big data methods are stated below.

Differential privacy preservation of big data based on Machine learning: Machine learning methods evaluate the preservation techniques to protect the data. Some of the algorithms are analyzed with machine learning techniques like Supervisory Control and Data Acquisition (SCADA) [6], Federated Learning (FL) [9], Data-as-a-service (Daas) [18], Body Sensor Networks (BSN) [19].

Differential privacy preservation of big data based on Deep learning: Some of the deep learning techniques are used in differential privacy for privacy preservation of big data, such as DNN (Deep Neural Network) [9], Recurrent Neural Network-Dynamic Trajectory Privacy Protection (RNN-DP) [13], spatial decomposition [14], Convolutional Deep Belief Network (CDBN) [15], back-propagation algorithm [22], Double-Projection Deep Computation Model (DPDCM) [24].

Miscellaneous privacy preservation techniques in big data: The differential privacy for privacy preservation of big data is analyzed with some techniques and methods like the Internet of Things (IoT) [7], Data mining [8], Cryptography [10], [25], MRMondrian driver [11], Secure And Efficient Data Perturbation Algorithm Utilizing Local Differential Privacy (SEAL) [12], Traffic volume estimation [16], Privacy-Preserving Data Mining (PPDM) [17], Proximity Privacy [20], k-Anonymity [21], Body Area Network (BAN) [23].

The overall analysis of the differential privacy preservation of big data over different algorithms and techniques is shown in Fig 2.

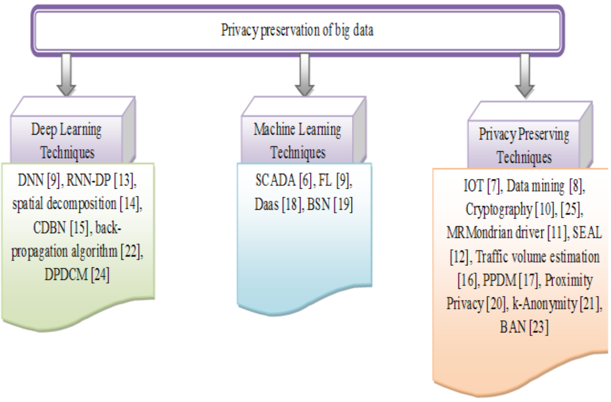
Privacy preservation of big data
Deep Learning Techniques
DNN [9], RNN-DP [13], spatial decomposition [14], CDBN [15], back-propagation algorithm [22], DPDCM [24]
Machine Learning Techniques
SCADA [6], FL [9], Daas [18], BSN [19]
Privacy Preserving Techniques
IOT [7], Data mining [8], Cryptography [10], [25], MRMondrian driver [11], SEAL [12], Traffic volume estimation [16], PPDM [17], Proximity Privacy [20], k-Anonymity [21], BAN [23]

Fig. 2: Differential privacy preservation techniques of big data over algorithmic classification approaches

b) Evaluating the Implementation Tools

Many Researchers utilized various implementation techniques for differential privacy preservation of big data, which are presented in Fig. 3. Some implementation tools include cloud platform, cross-validation, MATLAB, and Java. The commonly used tools for differential privacy preservation of big data are cross-validation and cloud.

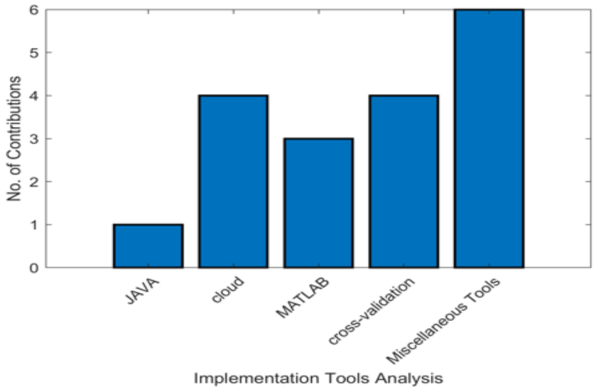

Fig.3: Differential privacy preservation of big data using different implementation tools

c) Various privacy preservation techniques of Big Data over diverse performance measures

Standard performance measures are used to assess differential privacy preservation of big data that are displayed in Table II. Different metrics are used to examine the performance analysis, including I loss, execution time, efficiency, average travel distance, cost, privacy, accuracy, and scalability.

ISSN 1533-9211

DOI 10.17605/OSF.IO/YDGR4

Table II: Performance measures over differential privacy preservation techniques of big data

Citation	Accuracy	I loss	Scalability	Privacy	Average Travel Distance	Cost	Execution Time	Efficiency	Others
[6]	✓	-	-	-	-	-	-	-	Detection rate and False Positive Rate
[7]	✓	-	✓	-	-	-	-	-	-
[8]	✓	-	✓	-	-	✓	✓	✓	Memory overhead
[9]	✓	-	-	-	-	-	-	-	-
[10]	-	-	-	-	✓	-	✓	-	Average Notified Workers
[11]	-	✓	-	-	-	-	✓	-	-
[12]	✓	-	-	✓	-	-	-	-	-
[13]	-	-	-	✓	-	-	✓	-	-
[14]	✓	-	-	-	-	-	✓	-	-
[15]	✓	-	-	-	-	-	-	-	-
[16]	✓	-	-	✓	-	-	-	-	-
[17]	-	-	-	-	-	-	-	-	T10, Chess, Retail, and T40
[18]	✓	-	✓	-	-	-	-	✓	-
[19]	-	-	-	-	-	-	-	-	Number of Shimmer wearable sensors
[20]	✓	✓	✓	✓	-	-		-	-
[21]	-	✓	-	-	✓	-	-	-	Relative cumulative frequency
[22]	✓	-	-	-	-	-	✓	-	-
[23]	-	-	-	-	-	✓	-	-	Number of interferences
[24]	✓	-	-	-	-	-	✓	-	-
[25]	✓	-	-	✓	-	-	✓	✓	-

IV. RESEARCH GAPS AND CHALLENGES ON CONVENTIONAL DIFFERENTIAL PRIVACY PRESERVATION TECHNIQUES OF BIG DATA

Big Data is used to describe the vast amounts of digital data that are gathered by various businesses and governmental agencies. The variety, volume, and velocity of big data include privacy vulnerabilities, high volume inter-cloud movement, streaming nature of data capture, large-scale cloud infrastructures formats, magnified security, and diversity of data sources. The big data system is additionally increased by the usage of cloud-based, large-scale infrastructure, with various platforms of software distributed over huge computer networks. The following are the primary reasons for privacy and security issues in big data, (1) Data source. (2) Secure communication and access control that are enforced via cryptography. (3) Distributed program frameworks and fast calculations. (4) Endpoint input filtering and validation. (5) Access control with granularity. (6) Continuous security and compliance oversight. (7) Safe data sources and records of transitions. (8) Endpoint input filtering and validation. (9) Security recommendations for sources of non-relational data. (10) Granular audits and (11) Flexible and modular analytics and data mining for privacy. A significant difficulty in large data is maintaining privacy.

The following list includes the numerous elements that affect large privacy and some difficulties, such as aggregated data sets and co-related data sets, context-based privacy, threat modeling, Budgeting for privacy, and policy and legal implications. Aggregated and co-related data sets: The big data are interconnected in data sets, so if one data set is disclosed in privacy, it also causes the data set privacy of another to be disclosed. Context-based privacy is the fundamental obstacle in big data privacy. Applying such privacy becomes trickier since it is impossible to determine the required privacy for numerous data sets in several contexts. Threat modeling is an organized method of creating a solution of privacy-preserving techniques by identifying the privacy objectives and potential assaults. Big data makes it difficult to identify threats and resolve them. Budgeting for privacy: this is another difficult problem that is solved to protect privacy in large data. It cannot select a computationally expensive strategy since the cost calculation is done regarding computing needs. Therefore, delivering excellent utility and privacy at a reduced cost is achievable if it has efficient processing, which is even harder to achieve. Policy and Legal Implications: The function of policy and legality is crucial. For data and its privacy, various countries have varied laws and practices. Maintaining privacy to all of these legal requirements is a significant challenge. Therefore, this helps the researchers to fill the research gaps in future research works and to improve the various privacy techniques in big data using differential privacy.

V. CONCLUSION

To preserve privacy for huge amounts of data, this research has observed numerous previously developed works on differential privacy preservation techniques of big data. It also analyzed the intelligent protocols' approaches, including machine learning and deep learning techniques. Additionally, this survey was examined by reviewing the protocols utilized in big data, performance metrics, and implementation platforms utilized in big data. Finally, the difficulties with various privacy preservation techniques of big data utilized in conventional research studies were highlighted to protect the information accurately. Therefore, it provided future insights on advancing the use of big data and encouraged the researchers to implement a novel technique to overcome the current research gaps in differential privacy for privacy preservation of big data.

References

1. Jain, P., Gyanchandani, M.and Khare, N. "Enhanced Secured Map Reduce layer for Big Data privacy and security," Journal of Big Data, vol. 6, pp.30, 2019.
2. Karim Abouelmehdi, Abderrahim Beni‑Hessane, and Hayat Khalouf " Big healthcare data: preserving security and privacy," Journal of bid data, 2018.
3. Lei Xu, Chunxiao Jiang, Jian Wang, Jian Yuan, and Yong Ren "Information Security in Big Data:Privacy and Data Mining," IEEE Access, vol.2, 2014.
4. Priyank Jain, Manasi Gyanchandani, and Nilay Khare " Differential privacy: its technological prescriptive using big data," Journal of Big Data, vol.5:15, 2018.
5. Xiaotong Wu, Taotao Wu, Maqbool Khan, Qiang Ni and Wanchun Dou " Game Theory Based Correlated Privacy-Preserving Analysis in Big Data," pp.2332-7790, 2016.
6. Keshk, M., Moustafa, N.and Sitnikova, E. "Privacy-preserving big data analytics for cyber-physical systems," Wireless Networks, vol.28, pp.1241–1249, 2022.
7. Wu, X., Zhang, Y.and Wang, A. "MNSSp3: Medical big data privacy protection platform based on Internet of things," Neural Computing & Applications, vol.34, pp.11491–11505, 2022.
8. M.A.P. Chamikara, P. Bertok and Seyit A Camtepe " Efficient Privacy Preservation of Big Data for Accurate Data Mining," Information Sciences, vol.527, May 2019.

9. B. Zhao, K. Fan, K. Yang, Z. Wang, H. Li and Y. Yang, "Anonymous and Privacy-Preserving Federated Learning With Industrial Big Data," IEEE Transactions on Industrial Informatics, vol. 17, no. 9, pp. 6314-6323, Sept. 2021.

10. M. Yang, T. Zhu, Y. Xiang, and W. Zhou, "Density-Based Location Preservation for Mobile Crowdsensing With Differential Privacy," IEEE Access, vol. 6, pp. 14779-14789, 2018.

11. X. Zhang, "MRMondrian: Scalable Multidimensional Anonymisation for Big Data Privacy Preservation," IEEE Transactions on Big Data, vol. 8, no. 1, pp. 125-139, 1 Feb. 2022.

12. Chamikara, M., Bertok, P., Liu, D., Camtepe, S., and Khalil, I. " An efficient and scalable privacy-preserving algorithm for big data and data streams," Computers & Security, vol.87, pp.101570, 2019.

13. Chen, S., Fu, A., Shen, J., Yu, S., Wang, H., & Sun, H. " RNN-DP: A new differential privacy scheme base on Recurrent Neural Network for Dynamic trajectory privacy protection," Journal of Network and Computer Applications, vol.168, pp.102736, 2020.

14. Yan, Y., Hao, X. and Zhang, L. "Hierarchical differential privacy hybrid decomposition algorithm for location big data," Cluster Computing, vol.22, pp.9269–9280,2019.

15. Phan, N., Wu, X. and Dou, D. Preserving differential privacy in convolutional deep belief networks," Machine Learning, vol.106, pp.1681–1704, 2017.

16. Yang, W., Sun, YE.and Huang, H. "Persistent transportation traffic volume estimation with differential privacy," Journal of Ambient Intelligent Human Computing, vol. 12, 213–231, 2021.

17. Shailaja, G.K.and Rao, C.V.G. "Robust and lossless data privacy preservation: optimal key-based data sanitization," Evolution of Intelligent, vol.15, pp.1123–1134, 2022.

18. Dagher, G.G., Fung, B.C.M.and Mohammed, N. "SecDM: privacy-preserving data outsourcing framework with differential privacy," Knowledge Information System, vol. 62, pp.1923–1960, 2020.

19. Lin, C., Wang, P. and Song, H. et al. "A differential privacy protection scheme for sensitive big data in body sensor networks," Ann. Telecommunications, vol.71, pp.465–475, 2016.

20. X. Zhang et al., "Proximity-Aware Local-Recoding Anonymization with MapReduce for Scalable Big Data Privacy Preservation in Cloud," IEEE Transactions on Computers, vol. 64, no. 8, pp. 2293-2307, 1 Aug. 2015.

21. Soria-Comas, J., Domingo-Ferrer, J.and Sánchez, D. "Enhancing data utility in differential privacy via micro aggregation-based k-anonymity," The VLDB Journal vol.23, pp.771–794, 2014.

22. Q. Zhang, L. T. Yang and Z. Chen, "Privacy-Preserving Deep Computation Model on Cloud for Big Data Feature Learning," IEEE Transactions on Computers, vol. 65, no. 5, pp. 1351-1362, 1 May 2016.

23. Chi Lin, Zihao Song, Houbing Herbert Song and Yanhong Zhou " Differential Privacy Preserving in Big Data Analytics for Connected Health," Journal of Medical Systems, vol.40, Feb 2016.

24. Q. Zhang, L. T. Yang, Z. Chen, P. Li and M. J. Deen, "Privacy-Preserving Double-Projection Deep Computation Model With Crowdsourcing on Cloud for Big Data Feature Learning," IEEE Internet of Things Journal, vol. 5, no. 4, pp. 2896-2903, Aug. 2018.

25. Anjana Gosain and Nikita Chugh "Privacy Preservation in Big Data," International Journal of Computer Applications, vol. 100, pp.17, August 2014.