IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Special Issue-9, April 2019 E-ISSN: 2347-2693

Big Healthcare Data Privacy Preservation –A Technological Perspective

Suneetha V[1*], Srivatsala V[2], Kumara Swamy Y S[3]

[1]Rayalaseema University, Kurnool, Andhra Pradesh, India
[2]Visvesvaraya Technological University, Belgaum, Karnataka, India
[3]Rayalaseema University, Kurnool, Andhra Pradesh, India

[*]*Corresponding Author: hod-mcabu@dayanandasagar.edu, Tel.: +91-9632831002*

DOI: https://doi.org/10.26438/ijcse/v7si9.6975 | Available online at: www.ijcseonline.org

Abstract— In this digital world, by virtue of highly diversified data generating technologies – huge amount of data is being churned out by organizations like hospices, banks, e-commerce, retail and supply chain, etc.,. Heaps and loads of big data is being generated every minute, by humans and machines. Because of onset of big data the industries have fundamentally changes their way of handling data. The volume and velocity big data generated from the various sources can be managed and analyzed to take appropriate decisions to benefit the organization. One of the most promising fields where big data analytics can be applied is healthcare. Big healthcare data and its analytics has considerable potential to improve quality of patients' life, gain valuable insights, prevent diseases, make healthcare more affordable. Securing data of patients and ensuring its security is major concern of data analytics. Unless the privacy and security issues of Big Data are addressed in healthcare industry it cannot be too useful. Invasion of patient privacy is a growing concern in big data analytics as emerging threats and vulnerabilities continue to grow. It is necessary to ensure a secure and sound environment for big data for better future in research by repairing the available solutions. In this paper, we present the security and privacy issues in big data applicable to healthcare industry. Also, we discuss the various Anonymization and Encryption techniques to preserve the privacy of the data, comparing their strengths and limitations.

Keywords— Healthcare; Healthcare privacy; Big data security; Big data Privacy; Data Anonymization, K-anonymity, T-closeness, L-diversity;

I. INTRODUCTION

Big Data refers to the humongous amount of unstructured data generated through the various digital channels. Big data is often recognized by the 4 Vs – Volume, Velocity, Veracity and Variety. This big data is making inroads into the healthcare industry too. The big data in healthcare is often referred to as Healthcare big data. It refers to the vast quantities of data that is now available to healthcare providers. Due to digitization of health care information and increase in the care system based on values, huge data is created, also vast amounts of data come from other sources, such as wearables, mobile applications, digital marketing efforts, social media, and more. All of the above ads up to an unbelievable amount of information, inspiring health systems to adopt big data techniques and technologies to effectively collect, analyze, and take advantage of this information. How is big data benefitting the healthcare? Why is it important to healthcare?

In this context while stating the importance of benefits of big data analytics in healthcare this paper discusses the various challenges faced by healthcare industry in adopting big data and highlights the importance of privacy and security of the big healthcare data. This paper also discusses the various privacy techniques for healthcare big data. A survey of the techniques and how they can be further improved is presented in this paper. Finally, it concludes by highlighting the importance of the privacy of big data in healthcare and comparing the various techniques. Section II focuses on big data and its role in healthcare industry, Section III deal with privacy issues in big healthcare data Section IV focuses on Privacy preservation techniques in big healthcare data and their comparisons. Section V finally concludes by reiterating the importance of protecting the big health care data.

II. BIG DATA ANALYTICS IN HEALTHCARE

Big data are two words that describe large volume of data that may be structured or unstructured, that floods the businesses every day. The data itself and its volume is not as important as what does the business do with this data. How do they mine this data and get the insights is what matters. This Big data is cleaned, stored and analyzed to gain insights to help the business managers make better strategies and decisions. Big data is characterized by 4 Vs Volume,

Velocity, Veracity and Variety. Volume refers to the large amount of data generated through various channels. Velocity refers to large speed at which the data is created. Variety refers to the variety of data created i.e. text, numbers, videos, comments etc. The strategy of analyzing large volumes of data, or big data is called Big Data analytics. This big data is gathered from a wide variety of sources, including social networks, videos, digital images, sensors, and sales transaction records. The analysis of this data is aimed at uncovering hidden patterns and connections that help in providing useful insights about the users and people who created this data. Thus, the businesses try to be one step ahead of their rivals and also make some useful decisions.

Big healthcare data too has all the characteristics of big data i.e. 4 Vs. Source of healthcare big data includes Electronic healthcare records, wearbles, health monitors, scan reports, surveys, etc; Health care industry is adopting big data in major way. The reasons behind it are a) Volume of data generated by the various channels in healthcare industry b) Keeping in mind the improvement of healthcare systems and the Government regulations the healthcare industry requires the data for analyzing and designing better healthcare systems. c) Customers are most valuable and the healthcare industry needs to create customized healthcare systems. Various insights are required into the customers' needs and requirements to create specialized packages. There are various benefits in adopting big data analytics by healthcare industry. Big data analytics gives a holistic view of the all the people involved- consumers, patients and doctors. Providing the patients with better and improved healthcare schemes, find the best marketing efforts to reach consumer and patients with best information, build predictive models to provide preventive car strategies, improve health care research and overall optimize the growth care, efficiency, effectiveness and personalization.

Challenges faced by the healthcare big data.

Huge amount of data leads to various challenges like sorting and prioritizing the data. Right kind of data must reach right kind of people to ensure correct and accurate analysis. Another challenge is the data being inconsistent, incomplete or noisy. Unclean data is big challenge as it is not useful to anyone.

Particular challenge faced by big healthcare data is privacy and security of data. Ensuring patient privacy is an enormous challenge.

What role does big data analytics play is healthcare in particular? Just like any industry the healthcare industry too can adopt the big data analytics. With already available large amount of clinical data and advanced tools and technologies the industry can now use big data analytics to gain useful insights and design better medical solutions and make revolutionary advances in medical research. It is imperative that the healthcare industry must incorporate big data analytics in its latest technologies for a better and bright future.

III. SECURITY AND PRIVACY ISSUES IN HEALTHCARE BIG DATA ANALYTICS

Big data analytics and medical research having real time access to patient record helps doctors to take decisions. Electronic Health Records (EHR) helped a lot in digitizing the health care system and various incentive programs motivate hospitals to create an accurate and complete EHR. At the same time, EHR having personal information of patient may lead to breach of privacy. Hence, techniques for preserving privacy are required and data need to be anonymized or encrypted before data analysis.

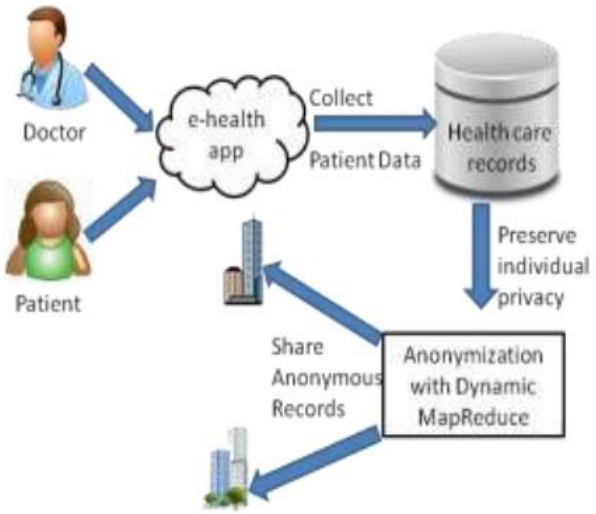

Fig-1 Privacy Preserved E-health app

Privacy vs. Security

Ability to protect the sensitive information about the patient is called privacy of data. Data privacy can be assured by ensuring there is governance of personal data that means proper means are employed and rigid policies are adhered while collecting sharing and utilizing data.

Data collected meticulously may be sometimes prone to malicious attacks and also may lead to misuse for making a profit. And protecting the data from threats like mentioned above is ensuring Security of the data. Though ensuring security is very important it is not a compromise on ensuring privacy of the patient data. Fig-2 Focuses on extra distinction between privacy and security.

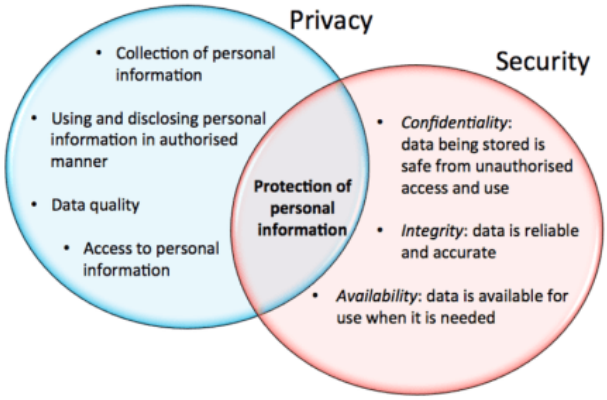

Fig-2: Privacy VS Security

Security of big healthcare data

Healthcare companies have mechanisms in place to ensure proper storage, maintenance of data and also have the ability to deal with release of voluminous amounts of data to support deliverance and care, but on flip side they lack technical support and minimal security. Among all the industries the most vulnerable industry is healthcare industry in data breaches. The malicious attackers can breach the private healthcare data and release it to public. Ensuring security is major concern and complex process. There are risks to be undertaken as the security measures become more and more complex. But it is essential and imperative that organizations ensure security and privacy of the healthcare data by designing security solutions to fulfil the healthcare goals.

Privacy of big healthcare data

The biggest fear in the big data analytics and healthcare industry is invasion of patient privacy. At the advent of new and advanced techniques to attack and steal data from information systems, this fear is real. As a result, organizations are challenged to address these diverse and critical issues. The incident where in a particular company has sent some baby related coupons to a teenage girl without her parents' knowledge has food for concern. Incidents reported like above mentioned forces everybody involved in analytics to big data privacy. Applications must strongly have in place security and privacy preservation techniques. They must also ensure that the key identifiers of private data remain private irrespective of the up gradations of the applications.

Ensuring Privacy of medical data is a vital and is critical.

Some light on the worldwide privacy protection laws in the next paragraph.

Data protection laws

To take care of the growing threat to breach of data and applicable data protection legislation, it is ever crucial that healthcare organizations are able to manage and guard personal and private information and also take care of their risks and legal related problems. Regulations concerning data protection and laws of data are different in different countries. The OECD Health Care Quality Indicators (HCQI project is accountable for measuring and comparing the quality of health services. HCQI is the quality indicators to study certain factors on the health services [20].

The General Data Protection Regulation (GDPR), decided upon by the European Parliament and Council in April 2016, will replace the Data Protection Directive 95/46/ec in Spring 2018 as the primary law controlling over how companies protect EU citizens' personal data.

Companies that are already in compliance with the directive should guarantee compliance with new needs of the GDPR [21].

Data Breaches are becoming too common

All of the regulatory necessities around data privacy, security, and preventing data breach of personally identifiable health data, are not sufficient and have become all too general across the industry. There were data breaches and thefts of medical data in the years 2015 and 2016 of around 80 million people and 11 million people respectively at Anthem and Premera. One analysis of the US Health and Human Services data breach database found a raise from 268 data breaches in 2015 to 328 separate breaches in 2016, with more than 16 million health records of American citizens being affected.

IV. PRIVACY PRESERVATION TECHNIQUES

An individual can decide which information can be shared or restrict the access to ensure his privacy which is basic requirement. If the key information is publicised then it is very vulnerable as the data is at the control of information holder. Here information holder will be websites, mobile apps, social networking application, e-commerce website, banks, hospitals etc. Guaranteeing privacy is crucial in present computing atmosphere. While not this, users feel uneasy to use and sleep in the UC atmosphere. The implementation of privacy safeguard or privacy enhancing technologies goes to be a protracted road. Recognizing the privacy preservation and possible protection techniques are useful to develop and implement the privacy preserved system. It's the responsibility of the information holder to make sure privacy of the user's data.

Various Privacy protective methods in which majority are based on Anonymization. Privacy preservation methods are as below.

1. Anonymization Technique
   - K anonymity
   - L diversity
   - T closeness

2. Randomization Technique
3. Cryptographic Technique
4. Data distribution Technique
5. Multidimensional Sensitivity Based Anonymization

1. Anonymization

Data gathering for analytics causes massive privacy issues. Person identifiable information (PII) is very tough because the information is shared too quickly. Urge to eliminate privacy issues, the conformity between the info holder and therefore the individual should be determined by policies. Users information ought to be anonymized (de-identified) and then be released to data analyser via protected channels. Before publication, the first table is altered consistent with the actual privacy needs. Anonymization is that the irreversible elimination of knowledge that would guide to a private being recognized, either on the premise of the removed info or together with different information.

The flustered information should be stripped of distinctive data, making it unworkable to realize insights on a discreet individual, even by the party that's in responsible of the anonymization. Anonymization operations used to preserve data privacy are listed below.

a. *Generalization*: It is replacing the specific Quasi-identifier (QID) values with reduced specific description with added generalization. Here we swap data values with a general value in the classification of a data attribute. An example where generalizing a designation data attribute with Employee instead of Developer or Tester. The types of generalization techniques include full domain generalization, sub tree generalization, multidimensional generalization, sibling generalization, and cell generalization.

2. ***Suppression***: Data values are swapped with some special characters like "*" in suppression operation. This way we can avoid disclosing of real data. Suppression includes record suppression, value suppression, and cell suppression.

3. ***Anatomization***: In anatomization, we dissociate the relationship between the quasi-identifiers and sensitive attributes. The data on QID and SA are released in two different data tables. One table contains quasi-identifier and the other table contains sensitive attributes which contain one attribute that is commonly referred to as GroupID. The GroupID will have same group value for the same group linked to the sensitive values.

4. ***Permutation*****:** The relationship between quasi-identifier and numerically sensitive attribute is disassociated by partitioning a set of records into groups and shuffling their sensitive values within each group in his data permutation,

5. ***Perturbation*****:** Perturbation is the way of swapping the original data values with some synthetic values, so that the statistical information computed from changed data does not differ considerably from the statistical data computed from the original data. Some examples like, adding noise, swapping data, and generating synthetic data values. The problem with perturbation is it is insignificant to the receivers as the published record is synthetic and does not have any meaning in the real world.

K-anonymity**:**

It is an interesting anonymization technique for big data privacy preservation. Here we deal with quasi identifier attributes. It is a unique approach in applying k-anonymity for QID attributes. It's a new algorithm called "*k*-anonymity without prior value of the threshold *k*". But, other many proposed *k*-anonymity algorithms the threshold *k* of *k*-anonymity has to be known before anonymizing the data set. K-anonymity is applied on the healthcare information as shown in Table 1. The table shows data before anonymization.

Table1: Original Data.

Slno	Pin	Age	Disease
1	515001	29	Diabetic
2	515275	22	Diabetic
3	500094	27	Diabetic
4	524369	43	Dermatology
5	524362	52	Diabetic
6	524333	47	Kidney
7	522365	30	Diabetic
8	522466	36	Kidney
9	522236	32	Kidney

K-anonymity technique is applied with k as 3 to ensure three identical data. K-anonymity is applied on the two attributes viz. age and Zip and is shown in Table 2.

Table2: Anonymization on age and Pin.

Slno	Pin	Age	Disease
1	515**	2*	Diabetic
2	515**	2*	Diabetic
3	500**	2*	Diabetic
4	5243*	>40	Dermatology
5	5243*	>40	Diabetic
6	5243*	>40	Kidney
7	522**	3*	Diabetic
8	522**	3*	Kidney
9	522**	3*	Kidney

In the above table, we have provided Anonymization using generalization or suppression operations. But still, if it is

known that Ram is 29 aged and residing at 515001 pin then easily we can make out that Ram is a Diabetic yet later data anonymization as shown in Table 2. This data leak is known as Homogeneity attack. In other case, if Ram is aged 36 and with little back ground knowledge that Ram does not have Kidney problem, we can easily find out that Ram must be having Diabetic problem. This is known as background knowledge attack. K-anonymity suffers with these limitations.

L-diversity is proposed to overcome the limitations of k-anonymity. This novel method is introduced to ensure privacy preservation by avoiding data attribute leak even with background knowledge. This operation "well-represent" the sensitive attributes in each group. Sensitive data attributes are varied among every quasi-identifier equivalence class. This is a K-anonymity's modification operation.

Table3: L-diversity privacy preservation technique.

Slno	pin	Salary	Age	Disease
1	515**	5k	2*	Diabetic
2	515**	6k	2*	Diabetic
3	515**	7k	2*	Diabetic
4	5243*	20k	>40	Dermatology
5	5243*	22k	>40	Diabetic
6	522*	24k	>40	Kidney

Data attribute leak cannot be ensured even when overall distribution of information is twisted into equivalence categories. For instance, if all the data rows are dispersed into 3 equality groups then semantic proximity of these data attributes could result data attribute leakage. L diversity suffers from homogeneity attack. In Table 3 it is found that if Ram is aged 29 and resides at 515001 zip is known, then definitely Ram falls in low income category as wage of all three persons in 515** pin are less compared to other data values. This is known as homogeneity attack.

T-closeness: Refinement of l-diversity by decreasing the granularity of the interpreted data is t-closeness operation. The analyser's scope of knowledge on a specific data is partial while the facts are not limited to the overall table containing the datasets. Therefore, this lessens the association between the quasi-identifier attributes and the sensitive attributes. In Table 4 it is found that Ram is aged 29, yet it is hard to guess whether Ram is Diabetic and falls under less salaried category or not. Attribute disclosure is ensured by T-closeness but sometimes, it might not provide correct allocation of data attributes.

Table 4: T-closeness technique

Slno	Pin	Salary	Age	Disease
1	515**	5k	2*	Diabetic
2	515**	16k	2*	Kidney
3	515**	9k	2*	Dermatology
4	5243*	20k	>40	Dermatology
5	5243*	42k	>40	Diabetic
6	5243*	8k	>40	Flu

2. Randomization technique:

The modification process of data values by adding noise using probability distribution is known as Randomization. It is widely implemented in surveys, sentiment analysis etc. It doesn't other records knowledge and can be implemented during data compilation and pre-processing time. Randomization technique is not applicable on large datasets because of time complication and data usefulness. For an example, we have considered and taken 20k records from a Hospital DB into Hadoop's HDFS and executed by Map Reduce.

• As data volume increased, more and more Mappers and Reducers are used.
• Significantly dissimilar results are found before and after randomization.
• Outlier records remain unchanged by randomization and are more unprotected and vulnerable for attacks.

Hence it is not an appropriate operation for the big data.

Fig-3 Randomization Technique

3. Data distribution technique

Data distribution technique is distributing the data across many sites. It is in two ways:

i. Horizontal distribution of data
ii. Vertical distribution of data

Horizontal distribution of data: In Horizontal distribution, data values are scattered among numerous sites with the equal valued identifiers. This technique can be allied on the information without really sharing the data. Here, the data values are scattered at diverse sites of various organizations. The parties help each other in noticing the data values at other ones. Here, truthful is expected among all participating sites.

Vertical distribution: Vertical distribution distributes the user's specific data in diverse nodes under control of various organizations. Here, every site stores partial set of identifiers of an individual. Data values will to be pooled from all these sites for data analytics and then, there is a possibility of privacy breach vulnerability. Ensuring privacy during data analytics of vertically distributed data is difficult as the data

values are scattered among various sites under control of various organizations.

4. Cryptographic techniques

This technique ensures privacy by encrypting sensitive. It provides proper toolset for algorithms of cryptography. We describe here results of cryptanalytic analysis that shows how different parties will together work out any function of their inputs, while not revealing the other information. Maximal privacy is achieved which hides all vital data except for the required output of the function. This technique attempts to model the world which is both realistic and common. Still there are few "real world" aspects which are not modelled; the privacy preservation and the generality of the results are quite notable. The data analyser may encode the data before discharging the equivalent for analysis. The crucial problem is encoding vast scale data using conventional encryption method is very difficult. So, it does not hold good for large databases. Also, this approach is hard to scale when more parties are involved.

5. **Multidimensional Sensitivity-Based Anonymization framework (MDSBA):**

There are three stake holders involved in Big Data analytics; they are the data owner, the service provider, and the user analyzer. MDSBA implements a bottom-up technique of k-anonymity. It is an enhanced Anonymization method implemented on huge data sets with lessen data attributes loss and predefined Qid's. This technique adapts a distinguished multi-access level for users. MAP REDUCE framework is used to handle Big data sets. The framework aims to apply a complete solution for MapReduce operations in big data. The solution basis mimics the parallel distributed processes over MapReduce nodes. This divides the single rigorous anonymization process into multi-tasks that can be distributed more than one node. Accessing data for analytics is conducted by many users with multi-level access in the big data environment. This compels a steady level of the data access and view. Users with a low-level permission are less trusted by data owners. Therefore, more restrictions are applied to a data view. Apache Pig scripting language filters are used to split data values into different bags upon likelihood allocation of the Qid's.

Apache Pig with four quasi identifiers is used in Data Anonymization. It can be safe guarded from background knowledge attack as the data is vertically distributed into various groups and when the bag holds less data identifiers. Data mapping with exterior sources to reveal any individual sensitive data is very hard in this technique.

Various features of above discussed data privacy preservation techniques are analysed and enlisted in the table as below.

Table 5: Comparison of privacy preservation techniques

Features	**Privacy preservation techniques**				
	Anonymization techniques	Cryptographic techniques	Data distribution	Randomization	MDSBA
Suitability of unstructured data	N	N	N	N	Y
Attribute preservation	N	N	N	Y	Y
Damage to data utility	N	N	Y	N	Y
Very complex to apply	N	Y	Y	Y	Y
Accuracy of data analytics results	N	Y	N	N	N

Because of Social media and IOT, major part of the data generated is unstructured in nature. But, based on outcomes and analysis of our well organized literature survey; most of the present privacy preservation techniques can handle optimally structured data. We need to think about the listed issues.

i. Design a robust technique to preserve privacy in structured as well as unstructured information.

ii. To develop Reliable, Scalable and sturdy techniques to hold huge diversified information.

iii. Health care Big data analytics can be carried out by ensuring privacy preservation by keeping the data in its native form and without data alteration.

iv. New methodologies should be designed to assure privacy upon countering key privacy threats that exemplify personal information revelation, police investigation etc.

v. Increasing information utilization for analytics by guaranteeing information privacy.

V. CONCLUSION

Abundant chances and prospects are open for health care big data to steer health care data analysis, predictions, decision making, planning strategies, data sighting, medical carefulness, and health care management. But, there are vast barriers and challenges that hamper its true possible outcome in the data healthcare analytics field, like practical issues,

privacy and security challenges and etc. These Big data safety and secrecy issues are critical obstructions for investigators in this area. Here in this paper, we have conversed how health care industry is making use of the big data analytics. Security and Privacy issues of Health care big data are discussed. Various Privacy preservation

techniques in the context of big healthcare data are discussed. But the problem ever exists. As our future scope we focus on designing optimized techniques for the privacy preservation of dynamically scaled big data healthcare privacy and security. Further we have tried to solve the problem by optimizing Anonymization using MapReduce framework.

We have analyzed and compared various Privacy preserving techniques. The knowledge of various available privacy preservations techniques will be useful to design Privacy preserved Healthcare environment.

It is very much needed to generate consciousness among the individuals about the different ways of protecting personal information against privacy and security breaches apart of these technological solutions. Lot of private data like images, contacts, mails, messages, chats and data files are accessed by lots of apps in our smart phones without our awareness. Many times, people install the apps without reading the privacy agreement statements. Therefore, here is a crucial need and necessity to instruct public to be aware of the variety of vulnerabilities which may allow private data breaches.

REFERENCES

[1]. Karim Abouelmehdi, Abderrahim Beni-Hssane, Hayat Khaloufi, Mostafa Saadi. "Big data security and privacy in healthcare: A Review", Procedia Computer Science, 2017

[2]. Aggarwal CC, Philip SY. A general survey of privacy-preserving data mining models and algorithms. Privacy-preserving data mining. Springer: US; 2008. p. 11–52.

[3]. Jiang R, Lu R, Choo KK. Achieving high performance and privacy-preserving query over encrypted multidimensional big metering data. Future Gen Comput Syst. 2018;78:392–401.

[4]. Hettig M, Kiss E, Kassel J-F, Weber S, Harbach M. Visualizing risk by example: demonstrating threats arising from android apps. In: Smith M, editor. Symposium on usable privacy and security (SOUPS), Newcastle, UK, July 24–26, 2013.

[5]. Bayardo RJ, Agrawal A. Data privacy through optimal k-anonymization. In: Proceedings 21st international conference on data engineering, 2005 (ICDE 2005). Piscataway: IEEE; 2005.

[6]. Iyengar S. Transforming data to satisfy privacy constraints. In: Proceedings of the eighth ACM SIGKDD international conference on knowledge discovery and data mining. New York: ACM; 2002.

[7]. LeFevre K, DeWitt DJ, Ramakrishnan R. Incognito: efficient full-domain k-anonymity. In: Proceedings of the 2005 ACM SIGMOD international conference on management of data. New York: ACM; 2005.

[8]. LeFevre K, DeWitt DJ, Ramakrishnan R. Mondrian multidimensional k-anonymity. In: Proceedings of the 22nd international conference (ICDE'06) on data engineering, 2006. New York: ACM; 2006.

[9]. Samarati, Pierangela, and Latanya Sweeney. In: Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. Technical report, SRI International, 1998.

[10]. Sweeney Latanya. Achieving k-anonymity privacy protection using generalization and suppression. In J Uncertain Fuzziness Knowl Based Syst. 2002;10(05):571–88.

[11]. Sweeney Latanya. k-Anonymity: a model for protecting privacy. Int J Uncertain, Fuzziness Knowl Based Syst. 2002;10(05):557–70.

[12]. Williams R. On the complexity of optimal k-anonymity. In: Proc. 23rd ACM SIGMOD-SIGACT-SIGART symp. principles of database systems (PODS). New York: ACM; 2004.

[13]. Machanavajjhala A et al. L-diversity: privacy beyond k-anonymity. In: Proceedings of the 22nd international conference on data engineering (ICDE'06), 2006. Piscataway: IEEE; 2006.

[14]. Xiao X, Yufei T. Personalized privacy preservation. In: Proceedings of the 2006 ACM SIGMOD international conference on Management of data. New York: ACM; 2006.

[15]. Rubner Y, Tomasi T, Guibas LJ. The earth mover's distance as a metric for image retrieval. Int J Comput Vision. 2000;40(2):99–121.

[16]. Priyank Jain* , Manasi Gyanchandani and Nilay Khare; Big data privacy: a technological perspective and review; Journal of Big data.

[17]. Karim Abouelmehdi*, Abderrahim Beni‑Hessane and Hayat Khaloufi; Big healthcare data: preserving security and privacy; Journal of Big data.

[18]. P. Ram Mohan Rao, S. Murali Krishna, A. P.Siva Kumar. "Privacy preservation techniques in big data analytics: a survey", Journal of Big Data, 2018 journalofbigdata.springeropen.com

[19]. Kajol Patel, G. B. Jethava. "Privacy Preserving Techniques for Big Data: A Survey", 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), 2018.

[20]. OECD Health Care Quality Framework (OECD Health Working Paper No. 23, March 2006). https://eugdprcompliant.com/

Authors Profile

Mrs Suneetha V is HOD-MCA Department, Dayananda Sagar College of Arts, Science and Commerce. She has overall experince of 19 years. She is pursiung her research at Rayalaseema University, Kurnool. Her area of research is Privacy Preservation in Big Data.

Mrs. Srivatsala V is working as Asst Professor in Department of MCA , Dayananda Sagar College of Arts, Science & Commerce. She has overall teaching experience of 14 years.Her area of research is Big Data in Healthcare and also Digital Marketing.

