Tourism & Management Studies, 14(2), 2018, 72-79 DOI: 10.18089/tms.2018.14208

Privacy preservation in data intensive environment

Preservação de privacidade em ambiente intensivo de dados

Jyotir Moy Chatterjee
Department of Computer Science & Engineering, GD-RCET, Bhilai, India, jyotirm4@gmail.com

Raghvendra Kumar
Department of Computer Science and Engineering, LNCT College, Jabalpur, India, raghvendraagrawal7@gmail.com

Prasant Kumar Pattnaik
School of Computer Engineering, KIIT University, Bhubaneswar, India, patnaikprasantfcs@kiit.ac.in

Vijender Kumar Solanki
CMR Institute of Technology (Autonomous), Hyderabad, TS, India, spesinfo@yahoo.com

Noor Zaman
College of Computer Sciences & IT, King Faisal University, Saudi Arabia, nzaman@kfu.edu.sa

Abstract

Healthcare data frameworks have enormously expanded accessibility of medicinal reports and profited human services administration and research work. In many cases, there are developing worries about protection in sharing restorative files. Protection procedures for unstructured restorative content spotlight on recognition and expulsion of patient identifiers from the content, which might be lacking for safeguarding privacy and information utility. For medicinal services, maybe related exploration thinks about the therapeutic records of patients ought to be recovered from various destinations with various regulations on the divulgence of healthcare data. Considering delicate social insurance data, privacy protection is a significant concern, when patients' mediclinical services information is utilized for exploration purposes. In this article we have used feature selection for getting the best feature set to be selected for privacy preservation by using PCA (Principle Component Analysis). After that we have used two methods K-anonymity and fuzzy system for providing the privacy on medical databases in data intensive enviroments. The results affirm that the proposed method has better performance than those of the related works with respect to factors such as highly sensitive data preservation with k-anonymity.

Keywords: Healthcare, healthcare data frameworks, unstructured restoration, fuzzy systems.

Resumo

As estruturas de dados de assistência médica expandiram enormemente a acessibilidade de relatórios médicos e o trabalho de administração e pesquisa de serviços humanitários. Em muitos casos, há preocupações crescentes sobre a proteção no compartilhamento de arquivos. Os procedimentos de proteção para conteúdo recuperado não estruturado são o reconhecimento e a exclusão de identificadores de pacientes do conteúdo, o que pode estar faltando para salvaguardar a privacidade e a utilidade da informação. Para os serviços de medicina, talvez a exploração relacionada pense que os registros terapêuticos dos pacientes devam ser recuperados de vários destinos com várias regulamentações sobre a divulgação dos dados de saúde. Considerando os dados do seguro social, a proteção da privacidade é uma preocupação significativa, quando as informações dos serviços médicos dos pacientes são utilizadas para fins de exploração. Neste artigo usamos a seleção de recursos para obter o melhor conjunto de recursos a ser selecionados para preservação da privacidade usando a ACP (Análise de Componentes Principais). Depois disso, usamos dois métodos K-anonimato e sistema *fuzzy* para fornecer privacidade em bancos de dados médicos em ambientes intensivos em dados. Os resultados afirmam que o método proposto tem melhor desempenho do que o dos trabalhos relacionados a fatores como preservação de dados altamente sensíveis com k-anonimato.

Palavras-chave: Cuidados de saúde, estruturas de dados de assistência médica, restauração não estruturada, sistemas *fuzzy*.

1. Introduction

Healthcare association has underutilized innovation as contrasted with different associations. A large portion of the medical associations is as yet depending on paper-based therapeutic records and written by hand solutions for analytic. Data digitized by social insurance association is commonly not compact; in this manner, there is little probability of sharing this data among various social insurance elements. Since data sharing is uncommon there is the absence of correspondence and coordination between patients, doctors, and another restorative group.

Cloud Computing advances permit the healthcare administration suppliers to enhance their administrations with the utilization of SaaS (Software as a Service) and DaaS (Database as a Service) model. With the utilization of such innovation empowers the healthcare administration supplier to development in the appropriation of PHR (Personal Health Records), EMR (Electronic Medical Records) and EHR (Electronic Health Records). Cloud figuring offers a few advantages in human services segment: social insurance association gives fast access to registering and vast storeroom at low cost. However, distributed computing likewise encourages sharing of medicinal services information crosswise over different offices and topographies. Electronic Medical Record/Electronic Health Record (EMR/EHR) systems are used to collect and store different type of patient data as well as their records (Dean, Lam, Natoli, Butler, Aguilar, &Nordyke, 2010; Lau, Mowat, Kelsh, Legg, Engel-Nitz, &Watson, 2011; Makoul, Curry, & Tang, 2001.

Privacy protecting information examination systems intend to anticipate exposure of delicate individual data while making the anonymized information usable for investigation. With electronic healthcare record information, minimization of such revelation dangers is of high need. Lately, there has been a great deal of work concentrating on the perception of wellbeing record information, both at the individual record level (Plaisant, Mushlin, Snyder, Li, Heller, &Shneiderman, 1998; Horn, Popow, &Unterasinger, 2001), furthermore at a total level to take a gander at companion examination and recognizing worldly patterns of medicines and patient arrangements (Perer, & Sun, 2012; Aigner, & Miksch, 2006). In both these cases, information's privacy protection is in danger as a result of the mind-boggling biological system of the human services industry, including both trusted and untrusted clients (Fung, Wang, Chen, &Yu, 2010). Here we used methods of feature selection for getting the best feature set to be selected for privacy preservation by using PCA (Principle Component Analysis). After that, we have used two methods K-anonymity and fuzzy system for providing the privacy on medical databases in data-intensive environments.

The rest of the article is organized as follows: Section 2 deals with the brief history of healthcare information system followed by the related work in this area. Section 3 discusses the leading healthcare challenges. In the Section 4 we discuss the motivation of our work followed by the dataset description. Section 5 deals with the proposed work with simulation results. Section 6 deals with a comparative study of the various techniques with their advantages and limitations and finally Section 7 gives the conclusion and future work.

2. Brief history

Accessing medicinal services information is an imperative prerequisite for medicinal experts and pharmaceutical specialists to study qualities of ailments. Lately, the expansion of distributed computing administrations empowers healing facilities and establishments to travel their medicinal services information to the cloud, which gives universal information access and on-interest high caliber administrations requiring little to no effort. Regardless of the advantages of medical services cloud benefits, the related privacy issues are generally worried by people, what's more, governments (Omnibus, 2013; Wang, & Zhang, 2015). Privacy dangers rise when outsourcing individual social insurance records to the cloud because of the delicate nature of wellbeing data and the social and legitimate ramifications for its revelation. A characteristic system is to encode social insurance information before traveling them to cloud (Cao, Wang, Li, Ren, & Lou, 2011; Yuan, & Yu, 2013; Wang, Ren, Yu, &Urs, 2012). On the other hand, preparing encoded information is not effective furthermore, it is constrained to particular operations, and along these lines, data with adaptable utilizations is not suitable for medicinal services. An option arrangement is applying existing privacy preserving data publishing (PPDP) systems, for example, segment-based anonymization (Liu, Wang, 2010; Mohammed, Fung, Hung, & Lee, 2009; Xiao, & Tao, 2006), and differential privacy preservation (Dwork, McSherry, Nissim, & Smith, 2006; Cormode, Procopiuc, Shen, Srivastava, & Yu, 2012; McSherry, & Mahajan, 2010; Lee, Wong, Goel, Dahlin, &Shmatikov, 2013), to the outsourced medicinal services information.

In 2007 Brickell, et al. proposed a secure assessment for the symptomatic system. The branching algorithms which can be spoken to as twofold choice tree is utilized to make wellbeing proposal to the client. The client might include their health-related information to the examining so as to branching algorithms then some edge values branching algorithms will make wellbeing proposals to clients. In parallel choice (Binary Decision) tree shaped by branching algorithms middle of the road hub contains predefined limit values, while leaf center point contains a course of action name to the customer. By applying the proposed method, the client assesses servers branching algorithms on clients nearby information without uncovering any information to server aside from arrangement name (Brickell, Porter, Shmatikov, &Witchel, 2007).

In 2007 Adam, et al. proposed a methodology that permits joining of heterogeneously conveyed information in a privacy protected and security saving way. Their proposed approach uses a cryptography based method, where the property estimations of all the qualifying datasets from every source are commutatively mixed by each one of the sources using their own keys. Commutative encryption ensures that the encoded keys from different data sets will be proportional if and just if their remarkable qualities are equal. The encryption keeps any source or addressing gathering from isolating the independently identifiable or sensitive information from the joined data set. A relative commutative disentangling ensures that simply the scrutinizing gathering can isolate the last result set. A future work is to address the noteworthy investigation test of planning data without a typical complete identifier in a privacy sparing manner (Adam White, Shafiq, Vaidya & He, 2007).

In 2008 Microsoft, along with Mohan et al. dispatched an effort which is intended for the people groups experiencing diabetes and cardiovascular disease in remote reaches in Caribbean countries. This effort can give redid human administrations for the patients having diabetes. They use wearable sensors like wearable biosensors for estimation of glucose level and circulatory strain this sensor data is sent to the cell phone through Bluetooth or USB-links. Later data is transmitted to the web server by the method for GPRS. At the web server, current sensor readings are united with past readings and submitted to thinking engine. By using these readings thinking engine produces altered access to calm (Mohan, Marin, Sultan, &Deen, 2008).

In 2009 Layouni et al. addressed one of the primary obstacle opposing the appropriation of therapeutic telemonitoring, is the worry among patients that their privacy may not be accurately ensured. This concern is tended and Layouni et al., suggested a privacy protecting telemonitoring method for human services. The method permits patients to specifically

reveal their personality data and ensures that no health-related information is sent to the observing focus without the patients' former endorsement. The endorsement procedure can be computerized and requires just a beginning setup by the patient. The convention proposed by Layouni et al., gives security against mimic violation, notwithstanding when the patient's PC is traded off. This is accomplished utilizing a smartcard-based two-element validation instrument (Layouni, Verslype, Sandikkaya, De Decker, &Vangheluwe, 2009).

In 2010 Danilatou and Ioannidisdiscussed biomedical exploration regularly depends on having entry to inconceivable measures of touchy data. Danilatou and Ioannidis suggested a building design that joins conveyed access control systems with protection saving cryptographic conventions to empower secure sharing and calculations on the cloud holding delicate biomedical information. Access rights might be assigned to different parties making coordinated efforts less demanding. At last, information can be worked on cryptographically to concentrate particular data without bargaining the whole information set. Danilatou and Ioannidis introduced a two-level structural engineering for security and privacy protection in the biomedical cloud. Danilatou and Ioannidis consolidated the force of decentralized administration and access control, gave by cryptographic certifications, with the capacity to perform protection safeguarding set operations on information (Danilatou, & Ioannidis, 2010).

In 2011 Barni et al. concentrated on the improvement of a privacy protection programmed finding framework whereby a remote server synthesizes biomedical signals gave by the client without getting any information about the signals itself and the last eventual outcome of the request. Specifically, here two techniques for the protected course of action of electrocardiogram (ECG) signals are presented and took a look at the past signals in light of direct (straight) stretching programs (a particular kind of decision tree) and the keep going relying upon neural frameworks (Barni, Failla, Lazzeretti, Sadeghi, & Schneider, 2011). This paper deals with each one of the requirements and difficulties related to working with data that ought to stay encoded in the midst of all the computation steps, including the need of working with adjusted point computations with no truncation while guaranteeing the same execution of a drifting point use in the plain space. Their proposed structures exhibited that doing complex endeavors like ECG portrayal in the encoded space profitably is in actuality possible in the semi-honest to goodness model, planning to fascinating future applications wherein security of signals proprietors is guaranteed by applying high-security rules.

3. Leading healthcare challenges

The healthcare challenges are as follows:

- More Data Sharing: As health-care frameworks enhance their examination capacities and the nature of their medical services information, imparting it to clinicians will turn out to be progressively vital. Sharing information definitely affects responsibility, efficiency, care quality, and advancement. Important information can rouse, encourage joint effort, and serve as an extremely powerful specialized apparatus. In spite of the fact that it's enticing to sit tight for flawless information, organize continuous information sharing over flawlessness.
- Engaging Patients *through Automation:* Using innovation to draw in patients will be a point of convergence of patient engagement discussions in 2016. The social insurance Internet of Things showcase is relied upon to hit $117 billion by 2020. The selection of wearables gadgets expanded by 60 percent in 2015. Half of the patients hospitalized in the most recent year began utilizing wearables devices after their healing facility remain.

4. Method and dataset

We get motivated to carry out the work in Hunka, Dash, &Pattnaik, P. K. (2016) to a step further. We have tried to curb the privacy preservation issues of healthcare sector using a fuzzy approach. As Healing centers and mind suppliers are ordinarily the information proprietors in this biological community. Different partners incorporate investigators, insurance agencies, clinicians, pharmaceutical organizations, and so on. They can get to be information caretakers and utilize the information for their own investigation. Much of the time a percentage of the information are additionally freely accessible, promoting accessibility of the information to clients who are outside of the biological community.

4.1 Dataset

In this work, we have used the Breast Cancer Wisconsin Dataset, it is having 699 instances and 11 attributes and 2 classes, to recognize malignant (cancerous) from benign (non-cancerous) specimens. The dataset comprises of some classification patterns or instances with an arrangement of numerical highlights or qualities is indicated in Figure 1 and Figure 2.

Figure 1 - Principal Components Variances Vs Components

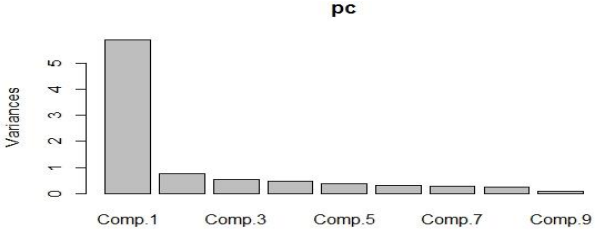


Figure 2 - Eigenvalue Vs Components

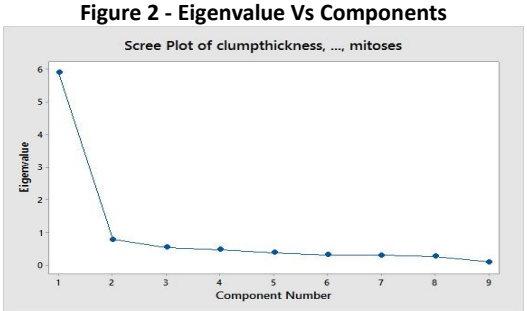


Fig 1. Shows the Principal components which are selected for privacy preservation Variance vs the various components (mitoses, clumthickness, etc.) of the dataset values used for getting PCA value for privacy preservation. In Fig 2. It shows the difference between the Eigenvalues with the various components of the dataset (i.e. clumthickness, mitoses, etc.). This Fig shows the screen plot graph.

5. Proposed work

Here using the method of feature selection for getting the best feature set to be selected for privacy preservation by using PCA (Principle Component Analysis) (Hasan, & Tahir, 2010). After getting the best feature set by using:

- K-anonymity
- Fuzzy logic

Principal component analysis (PCA) has been widely applied in the area of computer science and machine learning. It is well known that PCA is a popular transform method and the transform result is not directly related to a sole feature component of the original sample**.** There is no rule specified for choosing the features, one can choose features according to their requirements (Hasan, & Tahir, 2010). Table 1 indicates the CT-Clump Thickness, UCS_i-Uniformity of Cell Size, UCS_h-Uniformity of Cell Shape, MA-Marginal Adhesion, SEC_s-Single Epithelial Cell Size, BN- Bare Nuclei, BC-Bland Chromatin, and NN-Normal Nucleoli.

Table 1 - PCA analysis result

	CT	UCS_i	UCS_h	MA	SEC_s	BN	BC	NN	Mitoses
	-0.303	-0.381	-0.378	-0.333	-0.337	-0.333	-0.346	-0.336	-0.230
	-0.147	-	-	-	0.163	-0.246	-0.230	-	0.909
	0.862	-	-	-0.420	-0.112	-	-0.198	-0.134	-
	-	0.203	0.177	-0.468	0.363	-0.553	-	0.457	-0.237
	-	0.138	0.105	-	0.688	0.131	-0.257	-0.622	-0.133
	0.268	0.121	-	0.671	-	-0.605	-0.257	-	-0.129
	-	-0.205	-0.141	0.148	0.179	-0.257	-0.696	0.506	-0.140
	-0.243	0.439	0.585	-0.123	-0.450	-0.696	-0.410	-	-
	-	0.736	-0.665	-	-	-0.410	-	-	-
Support	0.0437	1.0501	-0.317	-0.525	0.494	-2.969	-2.394	-0.129	0.04

So, after applying PCA we are getting two kinds of features:

- Negative features (UCS_h, MA, BN, BC, NN)
- Positive features (CT, UCS_i, SEC_s, Mitoses)

By careful consideration of the feature set, we are selecting the positive features set for privacy preservation purpose by taking the negative feature set have to process for privacy preservation of more attributes taking a bit more complexity. Figure 3, Figure 4 and Figure 5 represent the various chart plots generated during PCA analysis.

Figure 3 - Outlier plot Mahalanobis Distance Vs Observations

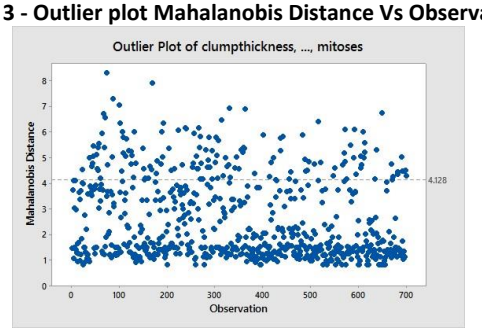


Figure 4 - Biplot Second Component Vs First Component

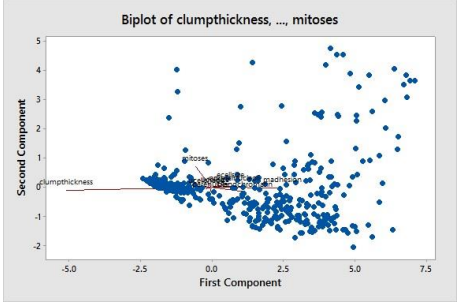
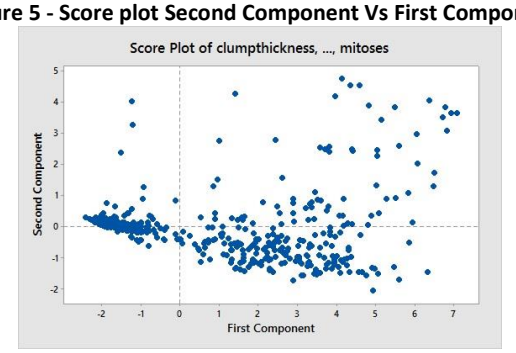


Figure 5 - Score plot Second Component Vs First Component



5.1 Privacy Preservation using K-Anonymity

Here we have used the K-anonymity method. Basically, we are working with the breast cancer dataset using k-anonymity to check if it provides the best privacy to dataset while sharing it over the cloud using "ARX tool". The "ARX tool" is having a wide range of algorithms in-built like K-anonymity, t-closeness, l-diversity, ð-privacy etc. A discharge gives k-anonymity insurance if the data for every individual contained in the discharge can't be recognized from in any event (k-1) people whose data likewise shows up in the discharge. The K-anonymization procedure is finished by applying speculation and concealment systems to the microdata set. The primary goals which should be accomplished: To discharge most extreme measure of information with the goal that it can be utilized for business or examination related work by different associations.

To guarantee that security of no individual is being placed in threat because of the discharged information by ensuring discharged data against surmising and connecting assaults are indicated in Algorithm 1.

Algorithm 1: Privacy Preservation using K-anonymity Method

Input:

(1) Source database DB,

(2) Anonymized parameter k,

Output:

A converted dataset DB'.

Algorithm:

1. Import the dataset into the workspace.
2. Categorize the Quasi-identifier values.
3. Apply the necessary conditions (hierarchies) required for the simulation.
4. Analyse the classification accuracy of the input and the anonymized dataset.
5. Set the value of k=3.
6. Set the suppression limit to 100%.
7. Set all the QI values to 0.5 as attribute weights.
8. Set the generalization and suppression balanced.
9. Finally, in analyze risk check for the re-identification risk of the input and the anonymized dataset.

By using K-anonymization for privacy preservation of data it provides the following results, which are indicated in Table 2, in Figure 6 and in Figure 7:

Table 2 - ARX simulation result for Quasi-Identifier attributes

	Risk (%)	Highest (%)	Average (%)
Input dataset	0.35211	100	36.90987
Anonymized output dataset	0.26525	33.33333	4.84653

Figure 6 - Result analysis of Input vs 3-Anonymous dataset

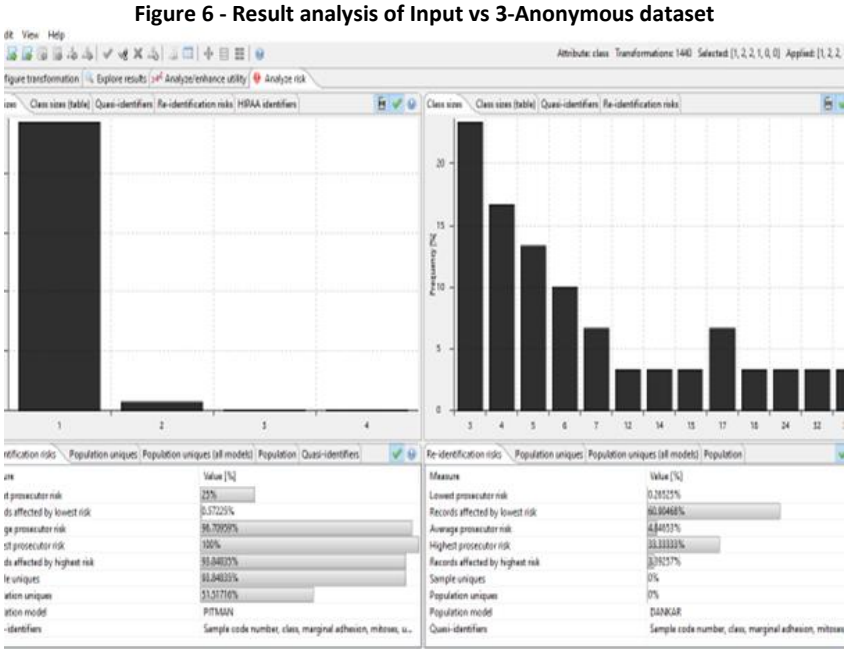


Figure 7 - Re-identification Risk analysis

K-anonymization is not **100%** safe for privacy preservation as it can be attacked by homogeneity and background knowledge attack. For better privacy preservation, we are using fuzzy logic.

5.2 Privacy Preservation using Fuzzy Method

In classical sets or crisp sets, the items in sets are called components or individuals from the set. A component x having a place with a set A is characterized asx∈A. A characteristic

function or membership function $\mu_A(x)$ is defined as an element in the universe U having a crisp value of 1 or 0. For every $x \in U$,

$$\mu_A(x) = \begin{cases} 1 \; for \; x \in A \\ 0 \; for \; x \notin A \end{cases}$$

The membership functions for the crisp set can take a value of 1or 0, the membership functions for fuzzy sets can take values in the interval [0,1]. The range between 0 and 1 is referred to as the membership grade or degree of membership (Hunka, Dash, &Pattnaik, 2016). A fuzzy set *A* is defined below:

$$A = \{(x, \mu_A(x) | x \in A, \mu_A(x) \in [0,1]\}$$

Where $\mu_A(x)$ is a membership function belonging to the interval [0,1].

Parameters that are to be applied in our proposed method are given as follows:

D: Raw dataset with *n* transactions information

C: Cleanse dataset with *n* transactions information

F: Fuzzified dataset.

Representations of the proposed work are given herein details step-by-step by using the Algorithm 2:

Algorithm 2: Privacy Preservation using Fuzzy System

Input:

(1) Source database DB,

(2) Minimum support-value (M_support),

Output:

A converted dataset DB' so that no one can deduce useful fuzzy rules.

Algorithm:

1. Begin
2. Cleansing of the dataset, DB→C.
3. The fuzzification of the cleansed dataset, C → F;
4. Calculations of every item's support values where the f∈ F, in fuzzified database F.
5. if all the f (Support) <M_support then
6. exit;
7. Find the large 2 item sets from the F;
8. Change the upgraded database F to DB' and yield redesigned DB';
9. End

Stage 1: Cleaning

The database is scrubbed by substituting the missing qualities by zero and disposing of the excess qualities. For our situation, no missing worth accessible so no adjustment in Table 3 and Figure 8. The cleansed dataset in Table 4 is fuzzified utilizing trapezoidal membership function given as a part of condition (1) into 4 areas namely a, b, c, and d as appeared in Figure 8.

Table 3 - Sample data with four attributes

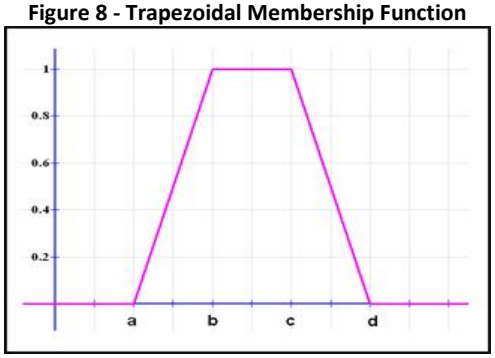
CN	CT	UCSh	Ma	Mi
1000000	0.5	0.1	0.1	0.1
1002900	0.5	0.4	0.5	0.1
1015400	0.3	0.1	0.1	0.1
1016300	0.6	0.8	0.1	0.1
1017000	0.4	0.1	0.3	0.1

Stage 2: Fuzzification

Table 4 - Cleaned data

CN	CT	UCSh	Ma	Mi
1000000	0.5	0.1	0.1	0.1
1002900	0.5	0.4	0.5	0.1
1015400	0.3	0.1	0.1	0.1
1016300	0.6	0.8	0.1	0.1
1017000	0.4	0.1	0.3	0.1

Figure 8 - Trapezoidal Membership Function



$$\mu_{trapezoidal} = \text{Max}\left(\min\left(\frac{x-a}{b-a}, 1, \frac{d-x}{d-c}\right), 0\right) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Where **a** is the lower limit, **d** is an upper limit, **b**an lower support limit, and **c** an upper support limit, where **a < b < c < d.**

Stage 3: Now find the bolster tally of every trait locale, R on the exchanges information by summing up the fuzzy estimations of the considerable number of exchanges in the fuzzified exchange information as in Table 5.

Table 5 - Fuzzified transaction data

N	CTa	CTb	CTc	CTd	UCSa	UCSb	UCSc	UCSd	ECa	ECb	ECc	EDd	Ma	Mb	Mc	Md
1000000	0.5	1	0.5	0	0.5	0	0	0	1	0	0	0	0.5	0	0	0
1002900	0	1	0.5	0	1	1	0	0	0	0.5	1	0.5	0.5	0	0	0
1015400	1	0.5	0	0	0.5	0	0	0	1	0	0	0	0.5	0	0	0
1016300	0	1	1	0	0	0	1	1	1	0.5	0	0	0.5	0	0	0
1017000	0	1	0	0	0.5	0	0	0	1	0	0	0	0.5	0	0	0
Count	1.5	4.5	2	0	2.5	1	1	1	4	1	1	0.5	2.5	0	0	0

Stage 4: Inspect whether the number of every quality is more noteworthy than or equivalent to the predefined least bolster esteem. In the event that a characteristic fulfills the above condition, place it in the arrangement of substantial 2 itemsets (L2). Consider the base backing here is set as 2.5.

Stage 5: Mark the important rules. Extract the items occurring in the important rules into a new table. In the example, if UCSa →CTb, UCSa →SECa, CTb →Ma, CTb→SECa, UCSa→Ma,CTb→ Ma are marked as sensitive then the items occurring in the sensitive rules are extracted as indicated in Table 6.

Table 6 - Items in the important rule

Sample code	CTb	UCSa	SECa	Ma
1000000	1	0.5	1	0.5
1002900	1	1	0	0.5
1015400	0.5	0.5	1	0.5
1016300	1	0	1	0.5
1017000	1	0.5	1	0.5
Support	4.5	2.5	4	2.5

Stage 6: Defuzzification using centroid technique is done on the changing qualities to get back quantitative qualities. The updated table D' is indicated Table 7.

Table 7 - Defuzzified Table

CN	CT	UCSh	Ma	Mi
1000000	0.5	0.1	0.1	0.1
1002900	0.5	0.4	0.5	0.1
1015400	0.3	0.1	0.1	0.1
1016300	0.6	0.8	0.1	0.1
1017000	0.4	0.1	0.3	0.1

6. comparative study

Table 8 provides a comparative analysis of the various previous techniques used for privacy preservation of data with their various advantages and limitations.

Table 8 - Comparative Study

Author	Technique or Parameter	Advantages	Limitations
Brickell et al., 2007	The binary decision tree or Branching program, Secure multi-party computation, Cryptographic techniques and software fault diagnosis	Privacy-preservation protocol is used for evaluation of diagnostic programs,represented as binary decision trees or branching programs.	The Computation and Communication complexity is very high
Adam et al., 2007	Integration and query of healthcare data from multiple sources, privacy preserving association rule mining, commutative encryption, commutatively encrypted by all the sources using their own keys, commutative decryption	A methodology that permits questioning and coordination of information from different sources, proposed approach utilizes a cryptography-based arrangement, whereby all the touchy qualities in the inquiry result are encoded by all the information sources utilizing their own keys.	One of the vast difficulties in consolidating information is the absence of a typical identifier crosswise over information frameworks, diverse sources gather distinctive components of data for the same arrangement of information. Rather than this, with homogeneous information dissemination, distinctive sources gather the same bits of data about various substances.
Mohan et al., 2008	Mobile Healthcare system, self-care process, reasoning engine	Suggested a personalized recommendation topatients suffering from diabetes and high blood pressure in a mobile environment.	There must be a way how social connection can be thought about amid personalization, distributive nature of the provincial MediNets is liable to produce a variety of exploration issues including information proprietorship and direction controls.
Layouni et al., 2009	Remote delivery of healthcare, medical telemonitoring, privacy-preserving telemonitoring protocol	This convention permits patients to specifically uncover their personality data and certifications that no wellbeing information is sent to the checking focus without the patients' earlier endorsement.	The issue of risk likewise merits further examination and the HMC (Health Monitoring Centre) ought to keep a record of the considerable number of endeavors it made to help the patients.
Danilatou& Ioannidis, 2010	Biomedical research, electronic data, bio-repositories and databases, data migration in the cloud, distributed access control mechanism, cryptographic techniques, security policies	An engineering that consolidates dispersed access control mechanism with privacy safeguarding cryptographic conventions to empower secure sharing and calculations on mists holding delicate biomedical information. The information imparted is labeled to security arrangements that characterize who has entry to it and how they ought to be utilized.	It is insufficient on its own when we would like to avoid revealing information unnecessarily
Barni et al., 2011	Biomedical signal processing applications, secure multiparty computation, cryptographic techniques, automatic diagnosis system, secure classification of ECG signals using linear branching program and neural network.	It concentrates on the advancement of a protection safeguarding programmed determination framework whereby a remote server orders a biomedical sign gave by the customer without getting any data about the sign itself and the last aftereffect of the arrangement, a profoundly effective adaptation of the fundamental cryptographic primitives is utilized here.	The outcomes got for the specific instance of ECG order ought to be reached out to more broad setups with the objective of inferring some broad decisions about the reasonableness of the QDF and the NN ways to deal with characterization in an SSP structure.

7. Conclusion and future work

In this article, we discussed the review on different privacy preservation healthcare frameworks and deduced that rather than by using k-anonymity method, the fuzzy method provides the highest privacy. The necessity for privacy preservation is reliably extending in our overall population due to the wide range of online passed on organizations offered by non-trusted gatherings having potential access to private data, ex. customer's data or other individual data. This need is extensively furthermore crushing in settings where the information to be guaranteed is related to the quality of the customers: with the vicinity of more online restorative storage facilities, it is anything but difficult to imagine that in two or three years the best approach to managing social protection will be absolutely not the same as the real one and it is of the utmost noteworthiness that control of sensible data does not deal the assurance of customers. Here in this article, we observed that the fuzzy method provides much better privacy rather than the K-anonymity method and in future work, it can increase the privacy by using different cryptographic techniques and protocols in medical databases with zero percentage of data leakage.

References

Aigner, W.& Miksch, S. (2006). Carevis: integrated visualization of computerized protocols and temporal patient data. Artificial intelligence in medicine, 37(3), 203–218.

Adam, N., White, T., Shafiq, B., Vaidya J., & He X., (2007). Privacy preserving integration of health care data. AMIA Annual Symposium proceedings, (pp.1–5), AMIA.

Brickell, J., Porter, D., Shmatikov, V. & Witchel, E. (2007). Privacy-preserving remote diagnostics. *In Proc. 14th ACM Conf. Computer andCommunications Security*, (pp. 498–507), ACM.

Breast cancer (Wisconsin) original dataset: http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Original%29

ARX tool: http://arx.deidentifier.org/api/. Accessed: 2016-03-20.

Barni, M., Failla, P., Lazzeretti, R., Sadeghi, A. & Schneider, T. (2011). Privacy-preserving ECG classification with branching programs and neural networks. *IEEE Trans. Inf. Forensics Security*, 6(2), 452–468.

Cao, N., Wang, C., Li, Ren, M.K., &Lou, W. (2011). Privacy-preserving multikeyword ranked search over encrypted cloud data, in: Proceeding of the IEEE INFOCOM (pp. 121-132).

Cormode, G., Procopiuc, M., Shen, E., Srivastava, D. & Yu, T. (2012). Differentially private spatial decompositions. In: Proceedings of the IEEE ICDE, (pp. 154-165), IEEE.

Dwork, C., McSherry, F., Nissim, K. & Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In: Theory of Cryptography, Springer, 265–284.

Danilatou, V. & Ioannidis, S. (2010). Security and Privacy Architectures for Biomedical Cloud Computing. I*n 10th Int. Conf. IEEE Information Technology and Applications in Biomedicine (ITAB),* (pp.1-4), IEEE.

Dean, B.B., Lam, J., Natoli, J.L., Butler, Q., Aguilar, D., &Nordyke, R.J. (2010). Use of electronic medical records for health outcomes research: a literature review. Medical Care Research Review, 66(6), 11–38.

Horn, W., Popow, C. & Unterasinger, L. (2001). Support for fast comprehension of icu data: Visualization using metaphor graphics. Methods of information in medicine, 40(5), 421–424.

Hunka, T., Dash, S., &Pattnaik, P. K. (2016). Web based Privacy Disclosure Threats and Control Techniques. In *Design Solutions for Improving Website Quality and Effectiveness* (pp. 334-341). IGI Global.

Hasan, H. & Tahir, N.M. (2010). Feature selection of breast cancer based on principal component analysis. In Signal Processing and Its Applications (CSPA), 2010 6th International Colloquium in IEEE, (pp. 1-4), IEEE.

Layouni, M., Verslype, K., Sandikkaya, M., De Decker, B. & Vangheluwe, H. (2009). Privacy-preserving telemonitoring for ehealth. I*n Dataand Applications Security XXIII*, 95–110.

Lau, E.C., Mowat, F.S., Kelsh, M.A., Legg, J.C., Engel-Nitz, N.M., &Watson, H.N. (2011). Use of electronic medical records (EMR) for oncology outcomes research: assessing the comparability of EMR information to patient registry and health claims data. Clin. Epidemiol, 3(1), 259–272.

Liu, J. & Wang, K. (2010). On optimal anonymization for l+-diversity. In: Proceedings of the IEEE ICDE, (pp. 23-32), IEEE.

Mohammed, N., Fung, B., Hung, P. & Lee, C. (2009). Anonymizing healthcare data: A case study on the blood transfusion service. In: Proceedings of the ACM SIGKDD, (pp. 32-41), ACM.

McSherry, F. & Mahajan, R. (2010). Differentially-private network trace analysis. In: Proceedings of the ACM SIGCOMM, (pp.24-31), ACM.

Mohan, P., Marin, D., Sultan, S. & Deen, A. (2008). Medinet: Personalizing the self-care process for patients with diabetes and cardiovascular disease using mobile telephony. I*n Proc. 30th Ann. Int. Conf. IEEE Engineering in Medicine and Biology Society, 2008 (EMBS 2008), (*pp. 755–758), IEEE.

Omnibus, Hipaa rule in the Federal Register, (2013). Retrieved 12 April, 2016 fromhttp://www.gpo.gov/fdsys/pkg/FR-2013-01-25/pdf/2013-01073.pdf.

Plaisant, C., Mushlin, R., Snyder, A., Li, J., Heller, D.& Shneiderman, B. (1998). Lifelines: using visualization to enhance navigation and analysis of patient records. In Proceedings of the AMIA Symposium, American Medical Informatics Association, (pp. 76-85), IEEE.

Perer A. & Sun, J. (2012). Matrixflow: temporal network visual analytics to track symptom evolution during disease progression. In AMIA annual symposium proceedings, American Medical Informatics Association, (pp. 716-725), AMIA.

Wang, W., &Zhang, Q. (2015). Towards long-term privacy preservation: A context aware perspective. IEEE Wireless Communication, 24(2), 142-159.

Wang, C., Ren, K., Yu, S., & Urs, K.M.R. (2012). Achieving usable and privacy-assured similarity search over outsourced cloud data. In: Proceedings of the IEEE INFOCOM (pp. 185-196).

Xiao, X. & Tao, Y. (2006). Personalized privacy preservation. In: Proceedings of the ACM SIGMOD, (pp.45-56), ACM.

Yuan, J. & Yu, S. (2013). Efficient privacy-preserving biometric identification in cloud computing. In: Proceedings of the IEEE INFOCOM, (pp. 178-186), IEEE.

Received: 13.12.2017

Revisions required: 10.02.2018

Accepted: 15.04.2018