International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 4 Issue 10, October-2015

Analytical Approach for Privacy Preserving of Medical Data

Vaibhav Lawand
SIT,Lawale
Dept of Computer engg
Pune ,India

Prathik Sargar
SIT,Lawale
Dept of Computer engg
Pune,India

Anand Bhalerao
SIT,Lawale
Dept of Computer engg
Pune,India

Pradip Jadhav
SIT,Lawale
Dept of Computer engg
Pune,India

***Abstract*— E-Science is getting more collaborative and distributed hence the data privacy of these digitalized big data is a important task, especially when the data is personal, confidential and contains sensitive information like patient's mental health records, psychotherapy notes and human behavior records etc. Medical health records are integral to self managed healthcare are applicable but when it comes to patient access it become a serious concern that requires a better balance for personalization, information care, security controls and privacy protection of individual data.**

For privacy management, existing policies are too restrictive and there is not much privacy aware data analysis research, for supporting of big data analysis. Thus there is need of analyzing the large variety, variation, velocity and volume of data. However, increased accessibility of highly sensitive mental records threatens the privacy and confidentiality of patient's records. In this paper we are describing the data analysis processes as workflows for medical data. We describe the workflow for privacy aware data analysis in mental health research and develop a analytical approach for privacy preserving of medical data to address these concerns.

Keywords— Privacy; Security; Mental Health Records; Data Analysis; Digital data; Data Mining

I. INTRODUCTION

Data privacy is important in e-science, especially when it's come to the analysis of huge amount of data collected from various sources including health care providers, pharmacies, insurance companies, government agencies, and research institutions [1]. In Health care information is considered as sensitive in nature and its privacy is a major concern when patient's personal health care data is analyzed for research purpose.

Existing data access policies offer a very basic privacy mechanism, where a user can access a data source if his/her certificates and credentials satisfy the access policies defined for that data source [2]. One of the novel approach is to check out with privacy preserving technique, and also certain credentials of privacy preserving techniques should be applied to the data before releasing it, where sensitive information like personal identity or medical background are properly hidden through anonymization and partitioning, but this is less restrictive. Data science aims to extract new insights from large amounts of structured and unstructured data and solve challenging task, such as patient assessment, early mental disease diagnosis, early interventions, and assessment of drug efficacy and safety.

In US the confidentiality of medical records is protected by the Health Insurance Portability and Accountability Act of 1996 (HIPAA) and Data Protection Act is applied in UK [3]. India now has a privacy law which includes reasonable security practices and procedures for sensitive personal data or information.

A. Security vs Privacy

A standard definition of Security, security is defined as the mechanism for protecting the entire HealthCare data including the ability to control access to patient information, safeguard from unauthorized disclosure, alteration, loss or destruction of patient information [4].

Privacy is also one of the most important keyword when we talk about medical data, since medical data contains sensitive information of individuals like type of disease. In research a live data or a real time data is always crucial especially for doing medical data analysis the original data should be available for making accurate predictions else it lead to impractical solutions. Any kind of disclosure related to the person specific information leads to many problems including ethical issues. Therefore it is important to take extra care to protect privacy of individuals before giving access or publishing such data. This crucial term can also be interpreted as preventing unwanted disclosure of information. For entire database both privacy and security measures are needed.

The privacy issue varies person to person according to the data in use and the context it is used is also based on different management level. In our case it is data related to medical research. But, the most important issue is how to provide privacy to data without loss of information. The methods like attribute removal, data hiding, and data compression can be applied on the data set to provide privacy, but will lead to information loss [4].Privacy and security are different but overlapping area of concern. Proper securities plans can be design to preserve privacy like certain limitation for collecting personal data.

This paper is drafted in systematical manner, Section I is an introductory part and it clearly describe the difference between privacy and security. Section II is about existing approaches like De-identification, Notice and consent Data Anonymization, Micro aggregation for privacy in medical data. In section III security of mental health data is discussed. Privacy aware data analysis workflow for medical data is also described in this section. At last we conclude our work with references.

II. EXISTING APPROACHES FOR PRIVACY IN MEDICAL DATA

Different approach for privacy preserving in medical data is enlisted as:

A. De-identification
B. Notice and consent
C. Data Anonymization
D. Micro aggregation

A. De-identification

One of the ways to preventing disclosure of confidential information is to remove the identical information from medical records.

One method that comes under de-identification is safe-harbor method; it is a statistical method, where identical personal information is deducted from medical records. The personal information can be names, address, telephone numbers, fax numbers, e-mail addresses, social security numbers, medical record numbers, finger and voice prints, full-face photographic images, health plan beneficiary numbers, account numbers, certificate/license numbers etc [5][6]. In the process of de-identifying information, a code is assign to the de-identified information by the covered entity so that it may re-identify it. The code may not be derived from information related to the individual. The covered entity is not allowed to disclose the key to the code to anyone else.

Under the statistical method, verification is done by a statistician or person with appropriate training, that enough identifiers have been removed so the risk of identification of the individual is very small [2].

B. Notice and consent

The notice and consent model means before release of clinical data for secondary uses, giving patients options to control how their data is used, patients should be given notice of it [7][8].

Before any personal information is collected from patients, a notice of an entity's information practices is given to them including identification of the uses to which the data will be put and identification of any potential recipients of the data.

Obtaining the individual informed consent for each specific secondary use is impossible or impractical for most current records-based research. Many studies depend on information collected years or decades previously, when no secondary use was anticipated. In these cases the cost of contacting patients and the response rates are disappointing and limiting research to those contacted will lead to biased conclusions. For these reasons, records based research depends on having access to identifiable health data without obtaining specific informed consent.

C. Data Anonymization

Data anonymization [9] is a process of information sanitization whose intent is privacy protection. It is the process of either removing or encrypting personally identifiable information from medical data, so that the people whom the data describe remain anonymous. With respect to medical data, anonymized data refers to data from which the patient cannot identify by the recipient of the information. Data Anonymization can be performed using methods like anonymization via anatomy and dynamic anonymization, these methods can be efficiency use to increase the accuracy of privacy preservation,

1) Anonymization via anatomy
2) Dynamic anonymization

D. Micro aggregation

Micro aggregation is a method of aggregating the records into groups. Instead of actually releasing the sensitive information for individual record, the mean of the group to which the observation belongs is released. This method includes various sub processes like, calculation of means with standard deviations and frequencies for all variables including hospital utilization, emergency services, use of outpatient services, and survival estimates for days to first medication visit post discharge by patients enrolled in medication management. This method is important to minimize the information loss by grouping similar records together.

In general, these privacy-preserving standards and procedures need to implement and enforce these technologies are available as shown in Table1. The goal of these techniques is to investigate a new kind of privacy protection policies that constrain the type of processing on the data, rather than the access to the data.

Table 1: Privacy Challenges and privacy preserving techniques

Privacy Challenges	Privacy-preserving Techniques
Re-Identification of records data	Suppressing both personal identifiers and quasi-identifiers, data aggregation
Notice and consent	Institution review boards, Transparency, Informed consent.
Data Anonymization	Information sanitization, encryption
Micro aggregation	Data aggregation, standard deviations

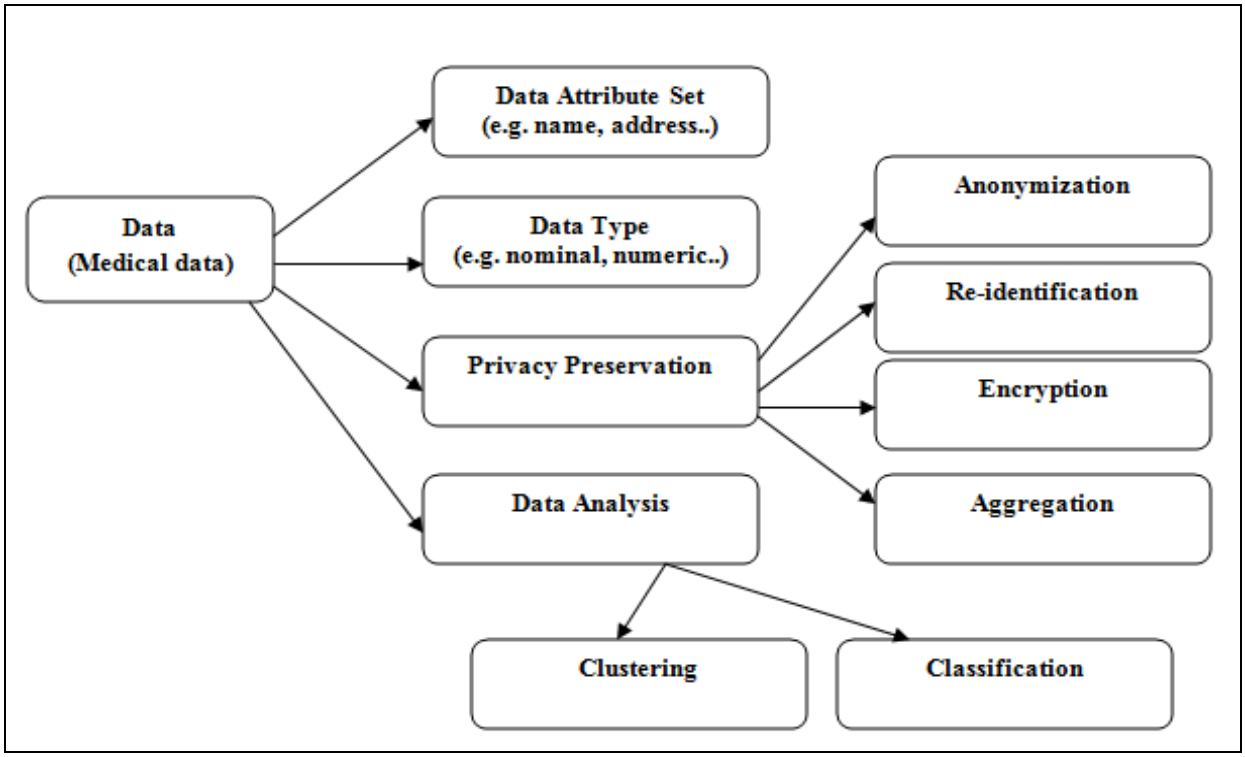

Figure1. Privacy aware data analysis workflow for medical data

Figure1 depicts the important concepts needed for describing privacy aware data analysis workflow for medical data.

We propose a systematic workflow for data privacy of medical health records. In our case the data is medical data as we are considering medical health record. The data is classified using different attribute sets like patients name, address, and telephone number; email address, medical record number etc. This data can be of any type like numeric data or nominal data.

Different privacy preservation techniques are applied on medical records. Some of them are anonymization, re-identification, encryption and aggregation. For data analysis process we can consider clustering and classification approach. For clustering purpose k-means algorithm is applied and for classification purpose SVM concept is used. In this way, the data analysis workflow is carried out for medical data.

III. SECURITY OF MENTAL HEALTH DATA

A high standard of data security is critical to the overall success of any scientific research or the provision of health care services. In the context of mental health records, security means the appropriate collection and handling of patient's data, providing protection to patient's data from unauthorized access or modification and safe storage of data.

Procedural security considers the internal processes and mechanism to handle the medical records. It includes who handles which data records in which situation, what they do with this data records and appropriate procedure for handling medical data [5].

Once granted access to medical records data, researchers must take any and all reasonable steps to prevent inappropriate access, misuse, and disclosure of identifiable clinical information.

Data encryption and security technologies such as firewall and Secure Socket Layer (SSL) can be applicable to provide the security to medical data [3]. These technologies enable medical data records to be safely stored, exchanged and analyzed.

IV. CONCLUSION

Health care information often consist of personal data involve analysis of big data collected from various sources including health care providers, pharmacies, insurance companies, government agencies, and research institutions etc with varying volume, velocity and with certain variation. Privacy is a major concern for information sharing in the healthcare domain. Privacy preserving data analysis procedures should be done to prevent inappropriate use of or disclosure of mental health records.

In this paper we discussed the existing privacy policies for medical health records along with the security for such information. Medical health information is more sensitive than any of the types of personal information. Medical information like mental disorders and treatments for mental disorders are generally more sensitive than are other health conditions. Inappropriate use and disclosure of any medical information may lead to discrimination in health insurance.

REFERENCES

[1] Adam, N., White, T., Shafiq, B., Vaidya, J., & He, X. (2007). Privacy preserving integration of health care data. In *AMIA Annual Symposium proceedings* (Vol. 2007, p. 1). American Medical Informatics Association.

[2] Cheung, W., & Gil, Y. (2007, July). Towards privacy aware data analysis workflows for e-science. In *AAAI Workshop on Semantic e-Science* (pp. 22-26).

[3] Bulgurcu, B., Cavusoglu, H., & Benbasat, I. (2010). Information security policy compliance: an empirical study of rationality-based beliefs and information security awareness. *MIS quarterly*, *34*(3), 523-548.

[4] Panackal, J. J., & Pillai, A. S. (2013). Privacy Preserving Data Mining: An Extensive Survey. In *ACEEE. International Conference on Multimedia Processing, communication and Information Technology*.

[5] Breeding, M. (2009). Privacy and Confidentiality. *Library Technology Reports*,*38*(3), 15-16.

[6] Curran, W. J., Kaplan, H., Laska, E. M., & Bank, R. (1973). Protection of Privacy and Confidentiality Unique law protects patient records in a multistate psychiatric information system. *Science*, *182*(4114), 797-802.

[7] Berg, J. W., Appelbaum, P. S., Lidz, C. W., & Parker, L. S. (2001). Informed consent: legal theory and clinical practice.

[8] Appelbaum, P. S., Lidz, C. W., & Meisel, A. (1987). Informed consent: legal theory and clinical practice.

[9] Byun, J. W. (2007). Toward privacy-preserving database management systems---Access control and data anonymization. ProQuest.

[10] Bennett, Kylie, Anthony James Bennett, and Kathleen Margaret Griffiths. "Security considerations for e-mental health interventions." *Journal of medical Internet research* 12, no. 5 (2010).

[11] Roos, L. L., Brownell, M., Lix, L., Roos, N. P., Walld, R., & MacWilliam, L. (2008). From health research to social research: Privacy, methods, approaches.*Social science & medicine*, *66*(1), 117-129.