

# Privacy Preserving Data Mining for Pharmacovigilance

Kamatchi Sankar, Latha Parthiban

**Abstract:** In this paper, visual cryptography concept is used for preserving the privacy of sensitive data used for pharmacovigilance. Overall analysis of adverse events of a specific drug helps in finding the potential danger of using the a specific drug. Preserving data owners privacy is done by using visual cryptography technique. Tetracycline drugs adverse effect present in FAERS dataset is extracted, encrypted and decrypted by the novel methodology proposed.

**Index Terms:** Visual Cryptography, Pharmacovigilance, Adverse event.

## I. INTRODUCTION

Preserving the privacy is very important for data owner security[1]. Large volume of data is useful for doing research. Sensitive data must be preserved without attacks as they are used by researchers continuously[2]. According to [3], “In India, we do not have robust structure which could in turn compromise the safety and lead to adverse outcome like in Rofecoxib story of devastation. Adverse effects could be avoided by vigilant monitoring and reporting. In developing countries like India, the pharmaco-vigilance programme was initiated by Central Drugs Standard Control Organization (CDSCO) in Nov 2004 under the aegis of ministry of health and family welfare based on the recommendations made in the WHO document entitled ‘Safety monitoring of medicinal products-guidelines for setting up and running a pharmacovigilance centre’ with the objective to monitor ADRs and report through hierarchy of pharmacovigilance network and disseminate the information with global healthcare community through WHO-Uppsala Monitoring Centre[4]. In India, there is a severe resistance to adopt the new technology. Due to lack of sufficient information and under reporting, the programme has been modified as Pharmacovigilance Programme of India (PvPI) and reinitiated in June 2010 with the aim to expand the existing structure and proactively report ADRs[5]. The purpose of PvPI is to collect, collate and analyze data to recommend regulatory interventions and communicating risks to healthcare professionals and consumers”[6]. The paper is organized as follows: Section 2 deals with literature survey; Section 3 with proposed

methodology; Section 4 with experimental results and section 5 with conclusions.

## II. LITERATURE REVIEW

Literature revealed that adverse reactions in drugs released to public has “deaths and hospitalizations numbering in millions (up to 5% hospital admissions, 28% emergency visits, and 5% hospital deaths), and associated costs of about seventy-five billion dollars annually [7]. Pharmacovigilance becomes necessary to identify these ADRs” [8]. Hence, present research focus is for privacy preservation of this sensitive data.

The data for this research is obtained from social media, hospitals and databases like FAERS by enforcing some constraints which is filtered. After filtering, learning algorithms are trained and used to find ADRs[9]. To improve the speed of algorithm parallel computing hardware can be used[10]. As adverse reaction can result in death, there is a need for designing parallel computing algorithms for analyzing the signal in big data generated.

## III. PROPOSED METHODOLOGY

The algorithm of the process contains 3 steps:

1. Message are broken into small blocks.
2. Encryption of the message creates ciphertext C.
3. Decryption results in original message at receiving end

The four rounds involved for encryption are as follows:

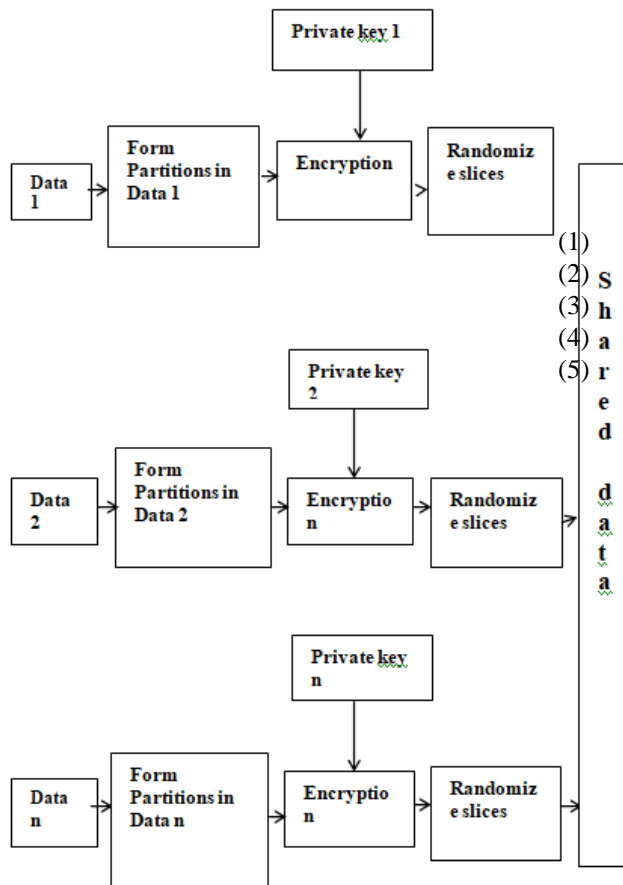
1. Frames creation: Frames are created from images.
2. Slices Formation: In this step, the frames are further divided into number of slices.
3. Randomization of the slice positions: In this step four bytes in each column are taken as input and generated as output.
4. Exploiting Concurrency with hardware: In this process the slices are processing concurrently in multiple partitions.

**Revised Manuscript Received on October 18, 2019.**

**Kamatchi Sankar**, Research Scholar, Dept. of Computer Science & Engg, Bharath Institute of Higher Education and Research, Tamilnadu, India, Email : sankam1309@gmail.com

**Latha Parthiban**, Department Computer Science and Engineering, Pondicherry University Community College, Puducherry, India. Email : lathaparthiban@yahoo.com

# Privacy Preserving Data Mining for Pharmacovigilance



1(a) Privacy preservation in sender side

The sliced data is combined in the receiver using destegano methodology as shown in the figure 2.

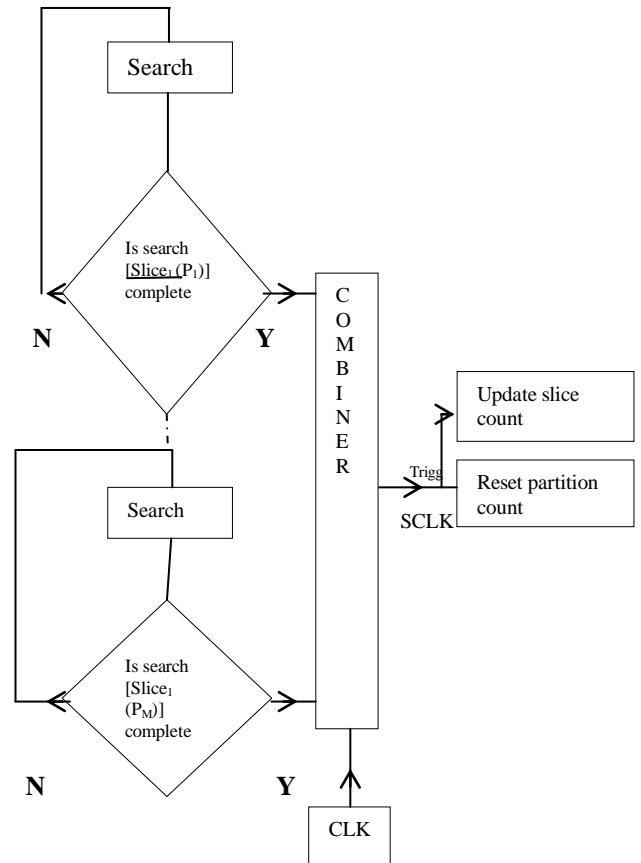


Figure 2 The destegano process

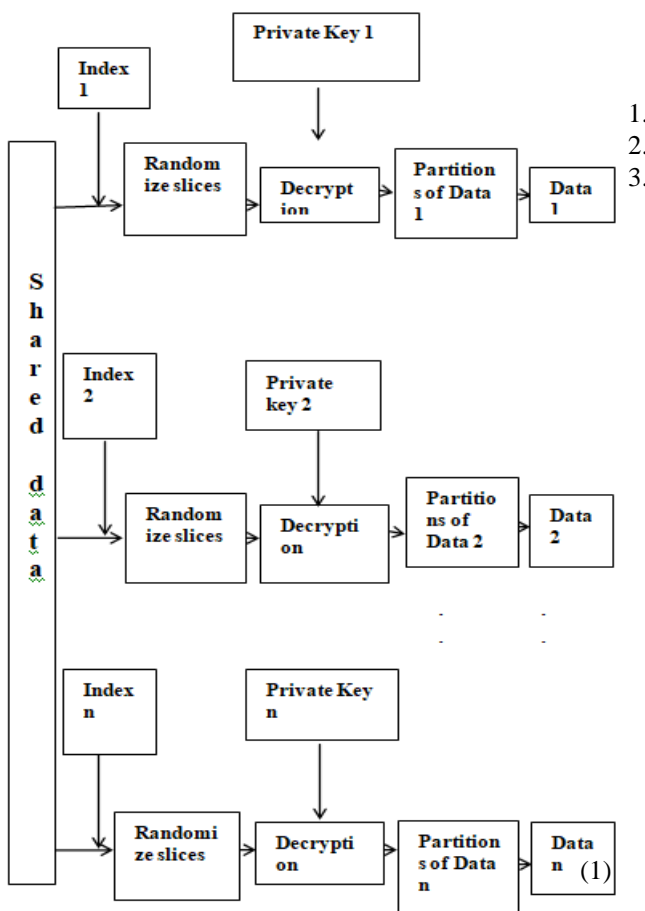


Figure 1b preserving Privacy in Receiver side

The steps involved in decryption are as follows:

1. Reverse shift slice positions
2. Combine frames
3. Reconstruct image

Privacy preserving data mining is shown in figure 1a and 1b.

## IV. RESULTS AND DISCUSSIONS

The verilog implementation results are shown in figure 3. Tetracycline ADRs from FAERS dataset is shown in Table 1.

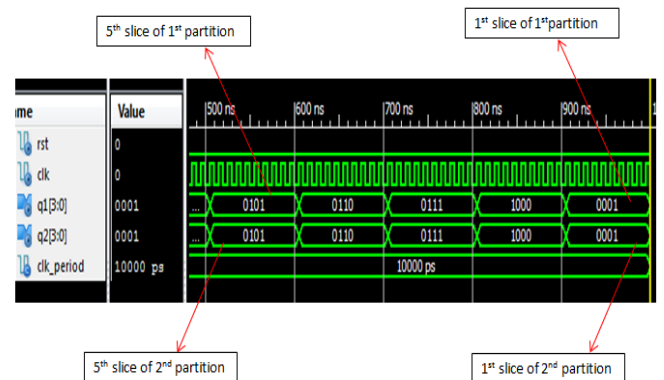


Figure 3 Verilog Implementation of privacy preservation

Table 1 Tetracycline ADR from FAERS dataset

Last database update	2019-06-08
Number of <b>matching records for this selection/group</b>	<b>2135</b>
Gender distribution	1541 (72.18%) female 415 (19.44%) male 35 (1.64%) unknown
Age distribution Age (Number of reports)	62 (42) 57 (37) 58 (37) 63 (37)
Top drugs Generic Name (Number of reports)	<a href="#">TETRACYCLINE HCL</a> (1449) <a href="#">BISMUTH SUBCITRATE</a> <a href="#">POTASSIUM,</a> <a href="#">METRONIDAZOLE,</a> <a href="#">TETRACYCLINE</a> <a href="#">HYDROCHLORIDE</a> (484) <a href="#">TETRACYCLINE</a> <a href="#">HYDROCHLORIDE</a> (203) <a href="#">ERYTHROMYCIN</a> (194)
Top medicinal products Product name (Number of reports)	TETRACYCLINE HCL (1443) PYLERA (484) CODEINE (167) ERYTHROMYCIN. (145)
Top drug classes MoA Mechanism of Action (Number of reports)	Cyclooxygenase Inhibitors [MoA] (251) Angiotensin-converting Enzyme Inhibitors [MoA] (11)
Top adverse events event (Number of reports)	DRUG HYPERSENSITIVITY (1205) RASH (126) TOOTH DISCOLOURATION (113) NAUSEA (100)
Top indications Indication (Number of reports)	HELICOBACTER INFECTION (257) PRODUCT USED FOR UNKNOWN INDICATION (230) HELICOBACTER GASTRITIS (62) HYPERTENSION (44)

**V. CONCLUSIONS**

In this paper, the authors addressed the data privacy problem in published data. The proposed method is designed for static micro data containing numeric quasi identifiers and classified quasi identifiers only. Further research is in progress to decrease the information loss with new attributes and incrementally update data records. The future work also focus on analyzing the adverse events due to multiple drugs

**REFERENCES**

1. Ching-Nung Yang, Li-Zhe Sun, Song-Ruei Cai (2016) Extended color visual cryptography for black and white secret image. Theor. Comput. Sci.609:143-161.
2. Ching-Nung Yang, Che-Yu Lin (2015) Almost-aspect-ratio-invariant visual cryptography without adding extra subpixels. Information Sciences.312: 131-151
3. Benton A, Ungar L, Hill S, Hennessy S, Mao J, Chung A, et al. Identifying potential adverse effects using the web: a new approach to medical hypothesis generation. J Biomed Inform 2011;44:989–96.
4. Hadzi-Puric J, Grmusa J. Automatic drug adverse reaction discovery from parenting websites using disproportionality methods. In: IEEE/ACM conference on advances in social networks analysis and mining; 2012. p. 792–7.
5. Yang CC, Yang H, Jiang L, Zhang M. Social media mining for drug safety signal detection. In: Proceedings of the 2012 international workshop on smart health and wellbeing; 2012. p. 33–44.
6. Sriramakrishnan and Latha Parthiban “ Data mining techniques for finding serious Adverse Events Journal of Chemical and Pharmaceutical Sciences, Volume 9 Issue 4, 2016 ISSN: 0974-2115
7. Shetty KD, Dalal SR. Using information mining of the medical literature to improve drug safety. J Am Med Inform Assoc 2011;18:668–74.
8. Wang, X., Chase, H., Markatou, M., Hripsak, G. and Friedman, C. Selecting information in electronic health records for knowledge acquisition. J Biomed Inform, 43, 4, 595-601.
9. Dr. K V Prasad et al, “Patient Self Controllable and Multi Level Privacy Preserving Co-Operative Authentication in Distributed m – Healthcare Computing System”, IJMETMR(2017), Vol.4, P.P.824-829.
10. Harshit Kumar, “Review paper on Big Data in healthcare informatics”, International Research Journal of Engineering and Technology (IRJET), IEEE (2017), Vol.04, P.P.197-201.

**AUTHOR PROFILE**



**Kamatchi Sankar** profile which contains their education details, their publications, research work, membership, achievements, with photo that will be maximum 200-400 words.



**Latha Parthiban** profile which contains their education details, their publications, research work, membership, achievements, with photo that will be maximum 200-400 words.