,L,

by

MARIOS G. VARDALACHAKIS

Informatics Engineer, Technological Educational Institute of Crete, 2020

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

SCHOOL OF ENGINEERING

HELLENIC MEDITERRANEAN UNIVERSITY

2024

Approved by:

Major Professor
Dr. Nikos Papadakis

Copyright

MARIOS G. VARDALACHAKIS

2024

If you choose not to include the Copyright page, delete this entire page. If you do include the Copyright page, delete these two paragraphs, but retain the Copyright heading, your name, and graduation year.

The Copyright page is not required unless you plan to register for copyright.

Abstract

In the digital era, health organization’s gather millions of volumes of useful data regarding a patient. It is very important to sufficiently hide patient data, while at the same time allowing valid medical research. Since data privacy has become the most paramount concern, mechanisms should be developed which protect private data but allow useful querying nonetheless. ShinyAnonymizer was a package designed for anonymizing health data and had solved the issue. It does this by making various anonymization techniques easily accessible to its end-users. This paper introduces the re-designed version of ShinyAnonymizer that is far more performant. We explain how data analysis and visualization, along with privacy-oriented statistical paradigms, are coupled with data anonymization, hashing, and encryption to offer a wide range of tools for researchers and data analysts to manage data in a secure way.

Keywords: healthcare data; data privacy; data anonymization; hashing; data encryption; statistics; visualization; privacy-preserving data analysis; ShinyAnonymizer.

Περίληψη

Τα ινστιτούτα υγειονομικής περίθαλψης παράγουν τεράστιες ποσότητες πολύτιμων δεδομένων των ασθενών στην σύγχρονη ψηφιακή εποχή. Είναι σημαντικό να βρεθεί η σωστή ισορροπία μεταξύ του απορρήτου των ασθενών και της ζήτησης για ιατρικές βελτιώσεις που βασίζονται σε αυτά τα δεδομένα. Δεδομένου ότι το απόρρητο των δεδομένων γίνεται όλο και πιο σημαντικό, πρέπει να αναπτυχθούν ισχυρές τεχνολογίες για την προστασία των ιδιωτικών δεδομένων και την ουσιαστική εξερεύνηση. Αυτό το ζήτημα αντιμετωπίστηκε από την εφαρμογή "ShinyAnonymizer", το οποίο δημιουργήθηκε αρχικά για την ανωνυμοποίηση (anonymization) δεδομένων υγείας. Αυτό το επιτυγχάνει καθιστώντας τις μεθόδους ανωνυμοποίησης εύκολα διαθέσιμες στους χρήστες. Η βελτιωμένη έκδοση του "ShinyAnonymizer", με σημαντική βελτίωση στην απόδοση, παρέχεται σε αυτή τη μελέτη. Εξηγούμε τη συγχώνευση των παραδειγμάτων ανάλυσης δεδομένων, οπτικοποίησης και στατιστικών που εστιάζουν στο απόρρητο με την ανωνυμοποίηση δεδομένων, τον κατακερματισμό και την κρυπτογράφηση, προσφέροντας μια εκτενή συλλογή εργαλείων για αξιόπιστη διαχείριση δεδομένων σε ερευνητές και αναλυτές δεδομένων.

Λέξεις-Κλειδιά: Ιατρικά Δεδομένα, Ιδιωτικότητα Δεδομένων, Ανωνυμοποίηση Δεδομένων, Αλγόριθμοι Κατακερματισμού, Κρυπτογράφηση Δεδομένων, Στατιστικά, Οπτικοποίηση Δεδομένων, Διατήρηση Ιδιωτικότητας Δεδομένων, ShinyAnonymizer.

List of Figures

Figure 1. Data anonymization process. 1
Figure 2. Hashing Process 2
Figure 3. Encryption Process 31
Figure 4. System Workflow 34
Figure 5 System Architecture 37
Figure 6 ShinyAnonymizer Interface 38
Figure 7 Graphical representation of patients 45
Figure 8 Descriptive Statistics of Patients 46

List of Tables

Table 1.An example of the Initial Dataset(Before Anonymization ……...7
Table 2.Anonymized Dataset……. ……...7
Table 3.An example of the Initial Dataset(Before Hashing…..……...7
Table 4 Hashed Dataset…….27
Table 5 An example of the Initial Dataset(Before Encryption…….32
Table 6 Encrypted Dataset……32
Table 7 Comparison with Other Works……47

Acknowledgments

My life has been surrounded by the love, strength, and the unrelenting belief in me that my late mother instilled in me. All my work, I do first and foremost for her. Her spirit, though no longer with me physically, still inspires me in everything I do. During this journey, it was her memory that stimulated me continuously, and this achievement is just as much hers as it is mine.

More than that, I would like to express my deep appreciation to my advisors and professors for the crucial support and guidance given.I am truly grateful that they believed in me and this kept me driven toward perfection.

Lastly, I would like to thank my sister, my dad, and also my friends who have stood by me in bad times when I lost my mother and shared the good ones with me. Due to your continuous support, motivating words, and company, this journey was more bearable and worth it.I would also like to extend my heartfelt thanks to all those who supported me in this. We could not have done it without them.

Preface

My Master's thesis represents the culmination of my research and development efforts in the subject of Data Privacy and Security, with a special emphasis on ShinyAnonymizer, a tool developed to protect sensitive data while at the same time keeping modern data protection requirements. Nowadays, the protection of private and sensitive data remains one of the biggest challenges that enterprises face around the world, since data breaches and cyber-attacks are becoming increasingly common. This thesis discusses some of the advanced approaches in anonymization, which can assure much better security for data while maintaining the usability and integrity of information. The focus relies on anonymization hashing and encryption methodologies.

This demand is driven by the growing need for secure data management applications that are in compliance with laws such as the Health Insurance Portability and Accountability Act and the General Data Protection Regulation. These two regulatory frameworks have placed tight measures on data privacy upon companies operating with private data, including financial information, medical records, and personal identifiers. To this end, ShinyAnonymizer seeks to meet such challenges head-on with a powerful, flexible platform that engineers security and compliance.

This thesis looks into the advantages and disadvantages of the current anonymization methods with the view to proposing improvements based on the use of sophisticated hashing algorithms and encryption algorithms. This study focuses on technical aspects that would be considered for eliminating hash collisions-like vulnerabilities while sustaining performance effectiveness. It has satisfied several firms in balancing the value of security against usage and other research contexts.

I am also very grateful to my academic advisors, colleagues, and all those people who helped me at every stage of this project. Their guidance and input were extremely useful to shape this research toward an appropriate direction with adequate depth. I hope this work will add to the continuous discussion of data security and foster further development in anonymization technologies to build safer and more secure digital environments.

Table of Contents

Copyright ..... **ii**

Abstract ..... **iii**

Περίληψη ..... **iv**

List of Figures ..... **vi**

List of Tables ..... **vii**

Acknowledgments ..... **viii**

Preface ..... **ix**

1.Introduction **3**

1.1 Background and Context **3**

1.2. The Significance of Privacy in Health Data Privacy 4

1.3. Introduction to ShinyAnonymizer and Its Role 5

1.4. Research Objectives 5

1.5. Overview of the Study 5

2. Literature review **6**

2.1. The Growing Awareness of Privacy Issues in Health Data 6

2.2. The Present State of Privacy Laws in Healthcare 6

2.3. Current Privacy-Preserving Tools 7

2.4. Conclusion of Literature Review 7

3. ShinyAnonymizer: Key features and functionality **7**

3.1. Key Modifications and Updates 7

3.2. Description of the Features of the Enhanced Version 8

3.3. Critique Concerning the Original Version 8

3.4. Technical Further Details 8

3.5. Overview of the Research Design 9

3.6. Description of the Research Methodology 10

3.7. Selection Criteria for Enhancements in Methodologies of ShinyAnonymizer 10
3.8. Data Sources Apis 10
3.9. Background of Data Anonymization Techniques 12
3.10. Technical Details of Anonymization Techniques 12
4 Enhancements to ShinyAnonymizer Hashing Techniques 16
4.1 Implementation of Advanced Techniques such as Hashing 16
4.1.1 MD5 16
4.1.2 Sha512 17
4.1.3 CRC-32 17
4.1.4 XHASH64 17
5. Enhancements to ShinyAnonymizer Data Encryption Techniques 21
5.1 Technical Details of Encryption Techniques 22
5.1.1. DES 22
5.1.2. XDES 22
5.1.3 BLOWFISH 22
5.1.4. AES512 23
5.2. Complete Data Security: Combining Hashing, Encryption, and Anonymization 26
5.3.Integration of Data Analysis Visualization and Statistics Paradigms 28
6.The Proposed Enhanced ShinyAnonymizer Tool 29
6.1.Workflow and System Architecture 29
6.2.The GUI 30
7.Real World-Use Cases 32
7.1. Clinical Research 32
7.2. Healthcare Analytics 32
7.3. Research on Public Health 32
7.4 Collaborative Research Efforts 33
7.5. Educational Environments 33

8. Discussion and Implications in the Contents of Privacy Regulations 34
9. Results and Experiments 35
9.1. Case Study: Clinical Study on Patient Attributes 36
9.2. Comparison with Other Works 40
10. Beyond ShinyAnonymizer: Future Directions 42
10.1. Artificial Intelligence and Learning 42
10.2: Blockchain Technology 42
10.3. Differential Privacy 43
10.4: Federated Education 43
10.5. Homomorphic Encryption 44
11. Conclusions 44
12. Bibliography of the Chapter 45

1.Introduction

1.1 Background and Context

Health data management is becoming increasingly complex, and robust privacy protection is being demanded. Health professionals can develop comprehensive patient records that make the management of patients' health easier, can immediately access records, and possibly reduce costs[1].In the context of increasing use of electronic medical records and continued development of health-related technologies, protection of sensitive information of patients has

turned out to be an issue of immediate concern. Various ramifications of data breaches in the healthcare industry include loss of trust in the whole system and confidentiality broken for individuals[2.3]. The development of privacy-preserving technologies and their improvement is paramount as we navigate through this complicated environment[4]. The paper explores the area of health data management, putting particular emphasis on privacy and discloses recent improvements in ShinyAnonymizer an interface intended for solving such a hard problem as striking a good balance between provisions on data availability and strict standards of privacy. More robust preservative approaches are needed with the severe data protection imposed by legislation, such as GDPR and HIPAA [5,6]. With the current status of technology, it may be a challenging task to strike the right balance between the value of data and privacy. If anonymization methods based on the past are applied, then data will become less useful for study and analysis [7]. While modern technologies, such as hashing and encryption, increasingly possess exceptional privacy guarantees, implementation in many cases is both physically demanding and extremely time-consuming. Considering such a complicated scenario, where the focus is on privacy, development and enhancement remain fundamental in those methods which guarantee the existence of privacy. This paper talks about recent advancements in ShinyAnonymizer, a tool created to help with one of the toughest tasks nowadays: balancing strict privacy laws with data availability. ShinyAnonymizer was created for health data anonymization and it was intended to easily enable users to access a variety of techniques of anonymization. In such a manner, the data analysts and researchers would be in a position to apply a wide range of effective methods of data management

1.2. The Significance of Privacy in Health Data Privacy

Privacy in health data management is more than a legal or moral requirement; it is a critical component that helps in maintaining the trust of individuals in health systems [8]. Health records are very important, and their processing should be one of sensitivity and precision. Patients must have assurance that their personal medical information remains secure, and researchers/clinicians

require methods that facilitate information sharing for both research and clinical purposes while preserving the highest possible degree of confidentiality. ShinyAnonymizer stands out as one such unique response initially targeted at anonymizing health data, but later expanded to cater to wider needs of healthcare sector privacy [9].

1.3. Introduction to ShinyAnonymizer and Its Role

ShinyAnonymizer was originally designed to anonymize health data with the intent of allowing medical practitioners and academics alike to interact with datasets while maintaining individual privacy. In walking through the process, the need for an improved and better version of this tool becomes demonstrably apparent. This ShinyAnonymizer Enhanced Version described on this page takes a step beyond basic anonymization by including additional capabilities and addressing concerns found in its predecessor. It needs to be positioned as an all-inclusive, innovative approach to the management of health data with privacy preservation.

1.4. Research Objectives

This paper has two key objectives: first, identification and rectification of the limitations of existing solutions in the healthcare business, and second, it provides an improved ShinyAnonymizer, showing the improvement made to minimize privacy issues. Accomplishing these objectives indicates that this work contributes to the ongoing discussion about privacy in health data management.Also, it offers a low-cost alternative for healthcare providers and investigators who aim at significant trade-offs between data availability and individual privacy.

1.5. Overview of the Study

The following sections provide an in-depth literature review, presenting historic events and emerging trends of the protection problems of health data, while growing panorama analyzes underlined benefits but also obvious drawbacks of preserving technology. In the methodology section, this work discusses the approach pursued by this study and sets aims for the

improvement of ShinyAnonymizer. Afterwards, the technical details of the new variant are discussed: the changes and improvements included, ending with a comparison with the original version. The results section compares how well the improved tool works with past solutions and practical scenarios. The discussion covers the implications of these changes and how they might combine with advanced privacy-enhancing technologies.

2. Literature review

2.1. The Growing Awareness of Privacy Issues in Health Data

The reason for such increase in growth about privacy health data concerns is essentially related to the history of healthcare processes and its increasing dependence on digital technologies. While there has been a shift from study-based healthcare records to digital health records or EHRs, data accessibility has increased along with increasing the danger of violations of privacy. Digitalization of health data brought tremendous development in the form of easy sharing of data and simultaneously created a demand for strong security to protect sensitive information about people. Understanding the historical backdrop to privacy issues is critical to interpreting current problems in health data management today [10].

2.2. The Present State of Privacy Laws in Healthcare

With the rise in privacy concerns, multiple legislation has been put forward that bound health data processing. The two most impactful legislations that shape how healthcare organizations process patient data are the United States' HIPAA and the European Union's General Data Protection Regulation (GDPR) [11]. However, medical technology and activity continue to develop unabated, and the utility of these constraints is increasingly called into question.

2.3. Current Privacy-Preserving Tools

An in-depth analysis of the current status of today's privacy-preserving tools will represent the merits and downsides of the current landscape. A variety of anonymization and protection health programs have been developed. Each has varying merits and capabilities, including k-anonymity algorithms, differential privacy methods, and encryption techniques. These technologies, despite their advantageous features, are increasingly faced with narrowing constraints, difficult implementation, and low responsiveness toward changing privacy needs[12].

2.4. Conclusion of Literature Review

Again, the literature review identifies that the nature of health data privacy issues is evolving in nature. It traces the evolution from the early techniques to the present difficulties. This suggests a regulatory landscape significant in the making of norms on privacy, but present techniques have their shortcomings and ensure further inquiry. This gives the ground for the follow-up years of this study to point out the ShinyAnonymizer Enhanced Version as a solution that tries to fill the gaps and cover the improvements in the needs of the solutions for privacy preservation in healthcare data management.

3. ShinyAnonymizer: Key features and functionality

3.1. Key Modifications and Updates

The limitations present in the first usage of ShinyAnonymizer have been updated and corrected several times. This enhanced version now contains state-of-the-art capabilities, providing an even more holistic solution to handle health data with privacy considerations beyond common anonymization. Concretely, there are enhancements in the anonymization approach, a perfect

enhancement of the hash and data encryption algorithms, and increased overall speed and capacity.

3.2. Description of the Features of the Enhanced Version

Various features have been incorporated into the updates of ShinyAnonymizer, making it even more functional. This has been done to extend its functionality in support of a range of data formats, such as organized and unstructured information as contained in EHRs. The application is now capable of handling a wide variety of healthcare data settings with assurance on sensitive data security while maintaining its usefulness in research and analysis. The enhanced version has dynamic anonymization algorithms that may allow for changes in real-time, considering changing privacy requirements and newly emerging security dangers.

3.3. Critique Concerning the Original Version

Basically, it is important to compare the improvements made to ShinyAnonymizer's initial version. It compares the evident improvements in efficiency, usefulness, and protection of privacy by explaining what has been accomplished. The improved version efficiency in handling complicated situations in healthcare data is compared to the limitations of the original software and provides an overview of the progress developed to enhance privacy security.

3.4. Technical Further Details

The technical details of the changes made in ShinyAnonymizer are very informative. It also covers an in-depth study regarding the adjustments made on the anonymization algorithm and the approaches made towards providing a better level of privacy while sustaining data significance. New security protocols to avoid the disclosure of confidential medical information are demonstrated, user permission, and authentication techniques. More aspects of its flexibility

and efficacy, describing the improved version in its capability to handle several big and unique data sets.

3.5. Overview of the Research Design

The design of the study aims to analyze the improved ShinyAnonymizer and address the deficiencies in the privacy-preserving options that have been identified. One of the first components of the research involves a detailed analysis of the available techniques and tools for the promotion of data privacy. This is in response to growing healthcare data management concerns over privacy; the review analyzes their effectiveness, drawbacks, and adaptability. The purpose of this introduction to the steps of improvement is to point out some limitations and challenges that, and how, the improved version of ShinyAnonymizer needs to face.

After following both qualitative and quantitative research methodologies, an improved version of ShinyAnonymizer was developed. Quantitative analyses are given about the performance of the improved tool, with comparisons to former versions and alternative solutions that ensure data privacy. These metrics are organized in terms of the overall efficacy of the system, efficiency, and anonymization efficiency. Qualitative methods are sought through data analysts and healthcare professionals who interact with the application to receive further depth on feedback about the usefulness of the tool, and what areas need more development. Conceptually and practically, these improvements will be stronger with three methods. Further, a careful evaluation phase also forms part of the study design after the improved ShinyAnonymizer is real-life tested. This includes distribution in various healthcare environments and reviews on usability as well as potential impacts on the field of patient privacy. Results obtained from these deployments are analyzed to determine the effectiveness of the tool in integrating security concerns and data usefulness. These findings from the study, therefore, go on to improve the tool in the manner it might effectively integrate with the current privacy-preserving technologies employed in the healthcare sector.

3.6. Description of the Research Methodology

The technical details of the changes made in ShinyAnonymizer are very informative. It also covers an in-depth study regarding the adjustments made on the anonymization algorithm and the approaches made towards providing a better level of privacy while sustaining data significance. New security protocols to avoid the disclosure of confidential medical information are demonstrated, user permission, and authentication techniques. More aspects of its flexibility and efficacy, describing the improved version in its capability to handle several big and unique data sets.

3.7. Selection Criteria for Enhancements in Methodologies of ShinyAnonymizer

The improvements on ShinyAnonymizer are simply the result of a thoughtful review of methods, considering the weaknesses exhibited in previous solutions and also applying those methods to the growing demands of healthcare data management. Specifications such as handling different kinds of data, flexibility, utility, and adaptability to any kind of privacy threats. A research study team works in collaboration with experts in the field of privacy, clinical doctors, and technological experts in order to provide comprehensive analysis and choose an appropriate methodology for better involvement in the release.

3.8. Data Sources Apis

The tools employed in the collection of these data are manifold, with a combination of several sources to create a very expansive and exact dataset. The primary sources of this data will include all healthcare-related institutions and organizations that grant permission for access to de-identified health records and associated data. Collaboration with hospitals, clinics, and research institutions enables the gathering of real-world health information representative, at a minute level, of nuances in the conduct of the healthcare business. Publicly available health and

medical-related datasets have been researched in this study. Publicly available data from recognized sources such as academic bodies, medical organizations, and international databases contain voluminous data that can be researched and replicated. These databases can hold demographic data, medical diagnoses, therapeutic outcomes, and anonymized medical records. Since the used dataset in this study is publicly available, the results obtained are quite transparent and reliable.

These also represent some of the independent sources that provide insight into the practical perspective of the application of the privacy-preserving technology and a better understanding of the challenges that come with handling health data. The qualitative and quantitative data combined in this study from several sources comprehensively and perceptively analyze how well the improved ShinyAnonymizer addresses the privacy issues in handling health data. Our dataset includes a fully elaborated simulated medical database created to reduce the challenges inherent in accessing electronic medical records. EMRbots.org is an open system for simulation of many of the attributes of real medical databases. It contains patient statistics, economic data, admission information, and a host of test results. Its flexible capabilities let users create a wide range of simulated patient populations that vary by gender, background, ailments, and other facets. This allows studies to be conducted in educational institutions, facilitates easier analysis by machine learning algorithms, and at times even to show potential investors and purchasers how to develop new innovative EMR management platforms [13]. Finally, the patient database developed during the process has been set up with the most important data organized in tables with filled cores. The "PatientCorePopulatedTable" has a respective information bundle for each patient ID, attached to the gender, date of birth, race, marital status, language, and socioeconomic level. In this layer, there are more APIs that supply access to other sources of data, including relational databases, CSVs, and Excel files. Furthermore, external data can be fetched dynamically, and several hashing, encryption, and anonymization techniques can be applied by using the corresponding API and REST services.

3.9. Background of Data Anonymization Techniques

The first step in the deployment of our approach for the protection of private data and individual privacy is data anonymization. Data anonymization is intended to hide PII information from criminals, making it harder for them to connect a particular data point to a particular individual. First, names and social security numbers are removed, as they are identifiers that uniquely identify individuals in the dataset. Identities are, however, often not removed. Secondly, it involves the use of such multifaceted approaches that prevent re-identification, thus making it impossible for direct linkage of data to individuals. Methods of anonymization are combined with utility concepts and quantitative privacy models in risk assessment and mitigation[14].

3.10. Technical Details of Anonymization Techniques

- Generalization

a. Description: Generalization is a process of replacing individual data values by more general, higher-level categories but less specific. For instance, in place of the age of an individual, it could use an age range. The above-mentioned method preserves the essential information in a group but reduces the resolution of the data given and makes the process of deducing specific subjects tougher for an intruder.

b. Application: Generalization treats sensitive attributes such as age, income, or location fairly effectively.

Removing Information:

a. Description: This involves the removal of identifiable information from the source itself. For instance, in a database, it eradicates any direct identifiers, including social security numbers and all records of names.

b. Application: Indirect identification can still occur hence, even though the removal process is simple in practice, it may not always be enough. Because of this fact, it is often applied in combination with one or several other methods of anonymization

- Suppression.

a.Description: Suppression involves masking or hiding chosen data values. For example, sensitive variables can be replaced by a summary placeholder like an asterisk (*) for the patient's actual gender, or subsets of data can be exposed showing only the first three digits of a social security number.

b.Application: Suppression is generally used to protect data that is very sensitive and cannot be shared broadly without significantly reducing the utility of the data. When only a small amount must be held while keeping certain analysis of the data, it is useful.

-Bottom Coding

a.Description: Bottom coding enables the categorization of data in far more useful groups than usually available, especially for variables with a large range of values. Thus, earnings levels can be labeled "low," "medium," or "high" instead of showing exact values.

b.Application: Bottom coding reduces the risk of re-identification by aggregating individual data points into larger groupings and increases the probability that fewer individual data points will be lost. It is often applied along with other techniques for data protection with the purpose of enabling beneficial study.

Each method was chosen to obtain the best trade-off between information value and anonymity. Once anonymization techniques have been applied, an effective evaluation process is required. Evaluation processes define whether anonymization techniques have been successful in removing identifiable information from the data and whether the quality of data is still good enough for the purpose required. The process ensures that the privacy of every individual is accordingly protected, although the analytical value of data is not lost. Once verified, no risks are transmitted or kept within the anonymized dataset. The intended approach maximizes data security and privacy while allowing companies to use their data productively, adhering to ethical and regulatory requirements [15].

Also, unlike perturbation, which requires adding noise into data, generalization replaces larger units for individual ones [16]. Every method was chosen to achieve the best balance between the value of data and privacy. Anonymization techniques require an intensive review process when being used. The evaluation will check if anonymization techniques have correctly removed personal features or not, and if the data is still useful for the intended purpose with enough quality. Even though the anonymized data is useful for analysis, the process ensures that the privacy of each individual is well protected. Once confirmed, the anonymized dataset can be retained or sent without risk. It allows the organization to use its data in an effective manner while assuring maximum data security and privacy in a legally and ethically appropriate manner[17].

Figure 1 presents an overview of our anonymization process: The anonymization process starts right away when the database is identified. First, a business identifies which data set or sets need anonymization. This takes a simple understanding of the background structure of the data and type of data contained therein as the foundation upon which all the subsequent steps in the process rest. Second comes the risk assessment, almost tied with the identification of the database. In this process, the data is analyzed for sensitivity to identify any potential risk associated with disclosure. The assessment of risk also helps in discovering the possible uses of data and guides the choice of proper anonymization techniques to minimize such risks. This is followed by the selection of methods for anonymization when the evaluation of risk is complete. In the given scenario, efficient strategies are selected regarding data types and risks detected.

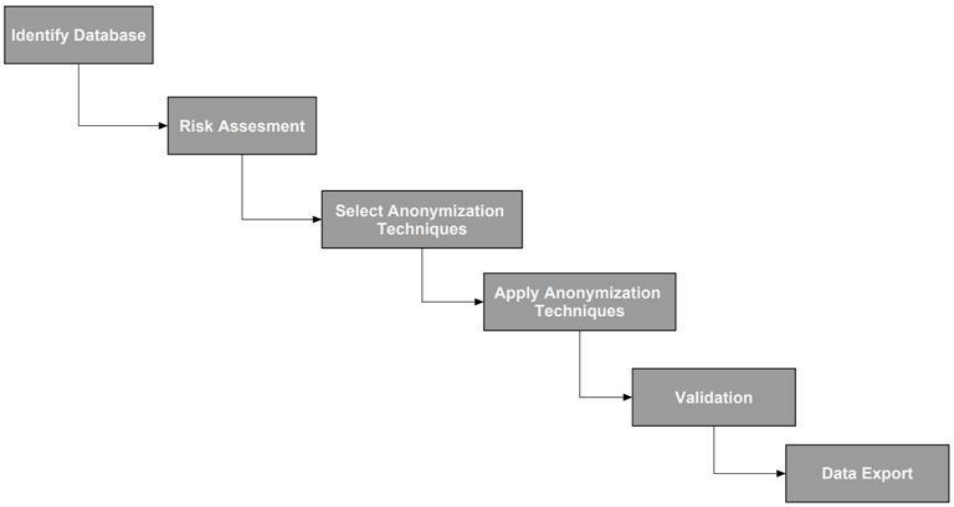


Figure 1. Data anonymization process.

Other techniques include generalization, masking, and encryption. In order to ensure the privacy of the anonymization process, the above techniques would be chosen and applied to the data transformation. The data validation occurs after the anonymization procedures have been applied to the data. This important step ensures that anonymization has been applied successfully and the data can remain usable for its intended purposes. This will protect unique identities, while the anonymized data retains its predictive capability. The final stage of the data export is for making the anonymized dataset available for distribution or archiving. The data must be made available for use for the purpose for which it was originally collected, under data security and privacy regulations.

Table 1 provides an example of the integration of several transformation models. The source dataset for the lead healthcare organization participating in this example is shown below within Table 1. There is a discrete dataset of patient information that includes gender, race, marital status, preferred language, and identifying values, or PatientId. An example would include an African American patient who is married and whose record number is FB2ABB23-C9D0-4D09-8464-49B70B982F0F0. Although such elaborated data are needed for internal procedures and analyses, there are some quite valid concerns related to privacy, considering possible personal identifications. For this reason, the company anonymizes the data as shown in Table 2. In this specific case, suppression and eliminating information techniques have deleted the columns of PatientGender and PatientRace and replaced them with general placeholders (*), which restricts monitoring on certain aspects. PatientMarital-Status has been generalized and several of the scenarios combined into bigger categories (e.g., 2, 2, 2), while PatientLanguage has been generalized into a specified range, which represents averaged language data, all by using bottom

coding. While keeping the dataset useful for study and analysis, this anonymized copy respects the privacy of sensitive individual data.

Table 1. Original Dataset.

PatientId	PatientGender	PatientRace	PatientMaritalStatus	PatientLanguage
FB2ABB23-C9D0-4D09-8464-49B70B982F0F0	Male	Unknown	Married	Icelandic
64182B95-EB72-4E2B-BE77-8050B71498CE	Male	African-American	Separated	English
DB22A4D9-7E4D-485C-916A-9CD1386507B	Female	Asian	Married	English

Table 2. Anonymized Dataset.

PatientId	PatientGender (Suppression)	PatientRace (Rem. Information)	PatientMaritalStatus (Bottom Coding)	PatientLanguage Generalization
FB2ABB23-C9D0-4D09-8464-49B70B982F0F0	*	127	2	45,000
64182B95-EB72-4E2B-BE77-8050B71498CE	*	228	2	45,000
DB22A4D9-7E4D-485C-916A-9CD1386507B	*	315	2	45,000

It is not possible, in the vast majority of cases, to perform anonymization methods allowing for such sophisticated transformation techniques by means of a search for an optimum solution over all possible output datasets; the accessible search spaces are, in the vast majority of cases, much too large. Techniques developed for this purpose include various hashing and elaborate data encryption methods. However, let us underline that one should bear in mind the conceptual representation of methods for anonymization of data using specific combinations of transformation, utility, and exposure models. For example, the actual application of previous algorithms is severely limited, often restricted to a specific set of pre-defined models. Thus, most of open-source technologies are freely accessible to whoever wants them.

It is well-recognized that context-especially the volume, difficulty, and statistical nature of data play a critical role in determining the success of different anonymization approaches. Key decisions are also needed regarding the type of analyses or applications on the data, whether or not the information will be subject to controls if the data are going to be made available to the public, and if the nature of the data is tabular, linear, or operational. Various algorithms and techniques for modifying the data and evaluating the change in utility must therefore be provided so that anonymization software can be applied in a wide range of application contexts.

In short, generalization, information removal, suppression, and bottom coding are helpful approaches to preserving utility while anonymizing data and ensuring the protection of privacy.Each of these techniques has strengths and weaknesses in order to achieve a balance between privacy and utility, they are combined in practice. The current study outlines different anonymization techniques and how to apply them to maximize the protection of sensitive data about privacy without compromising the utility of the data.

4 Enhancements to ShinyAnonymizer Hashing Techniques

4.1 Implementation of Advanced Techniques such as Hashing

In this respect, we will implement a cryptographic hash technique that can transform any format of data into a fixed-length sequence of characters called a digest or hash number. This will be the second approach for our methodology and system. Data integrity is ensured, and confidential data, like passwords, can safely be stored because of the irreversibility feature of this technique. While encryption does not allow the recovery of content to its original form, hashing is applied to confirm correctness without disclosure of the original data. Well-known hashing algorithms like XXHASH64, MD5, SHA512, CRC32, among others, are noted for different levels of security and performance[18].

4.1.1 MD5

a.Description: One of the most common cryptographic hash techniques, MD5 creates a 128-bit hash value, typically expressed as a 32-character hexadecimal integer. Since it is a one-way function, computationally it is inconvenient to work backward and deduce the original data from the hash value.

b.Application: MD5 has been deployed worldwide to check the integrity of data as a checksum. However, it is rarely ever recommended for secure applications that require cryptographic capability since flaws have been discovered that allow hash collisions to occur, where two different inputs swap for the same hash[19].

4.1.2 Sha512

a. Description: This approach is utilized in case the anonymization of the data set is carried out using unique or particular data. For instance, delete all identities, account numbers as well and other private data from a database.

b. Application: SHA-512 extends the SHA-2 family by generating a hash value that is 512 bits. It manifests as a hexadecimal integer with characters. It is supposed to be more reliable than MD5, providing higher resistance against collision attacks and surviving common cryptography attacks[20].

4.1.3 CRC-32

a.Description: Another example of a 32-bit non-cryptographic hash method is CRC32. It is among the most popular methods for checking accidental alternations of raw data and one of the most common check methods for transmission and storage errors.

b. Application: CRC32 is in use to provide integrity during file transfers and other related processes of data storage. While it's good with mistake detection, its low collision tolerance and loss of security against advantageous manipulation make it unfit for usage in cryptography.

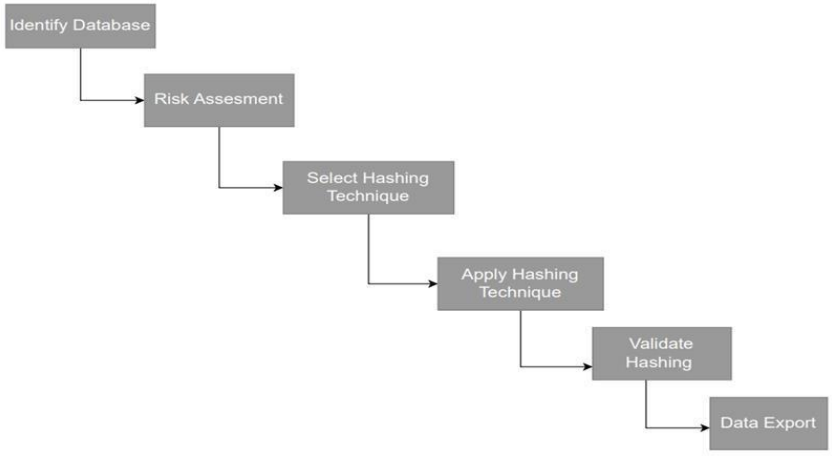
4.1.4 XHASH64

a.Description One such method is the use of Bottom coding, which could be used when there are a large number of factors and also when the factors have a wide range of outcomes. For example, income levels may be described as "low," "medium," or "high," but no other information may be given.

b. Application: XXHASH64 is a fast, non-cryptographic hash function that computes a 64-bit hash. It has been designed to hash data very fast and later updated to hash large datasets instantly.

Figure 2 provides a high-level overview of our hashing process. Once data is initially extracted from the database, a careful risk analysis is performed to determine the saliency of the material and associated privacy considerations. The study helps decide which hashing algorithm will work best. For instance, MD5, despite known weaknesses, may be selected for more salient data because of its performance. However, SHA-256 or SHA-512 might be used instead in the case of more sensitive material, as they offer a greater level of security. In practice, salting is often combined with hashing to further extend the level of security and reduce the likelihood of hash collisions-that is, when two independent sources result in the same hash value[21].Finally, when the risk analysis is performed and optimal hashing method has been decided upon, the data will be hashed. To do this, all data items must be rendered into a code of a hash yet the data items shall remain trusted and useful to study. Security with varying sensitivity degrees can be depicted if, for instance, in a dataset, patient genders are hashed using MD5 while the races of the patients use SHA-512. The hashed data is transmitted for consumption and thereafter carefully and keenly checked for its accuracy and precision to ensure that little or no security concern associated with transmission or storage. In as much as hashing has numerous advantages in security and computer productivity, it also has some limitations. The key feature is the ability to quickly and accurately verify information, which becomes necessary in terms of data integrity and safety techniques for validation. Of course, it does become somewhat relevant what kind of hashing algorithm is applied: whereas stronger algorithms like SHA-512 provide increased security at the same time as increased demands on processing, faster ones, like MD5, are simpler but easier to be compromised. This problem of hash collisions means salting and other methods have to be employed when trying to ensure each input is uniquely hashed. By adopting and using these hashing algorithms with care, an organization will be able properly to improve its data privacy practices without sacrificing private data security along with its usability.

Figure 2. A visual representation of the Hashing Process.



Hashing for this advanced transformation technique has several advantages and disadvantages. The strength of hashing is found in computational efficiency for its utilization in fast audits of integrity of data and password authentication. However, an appropriate hashing algorithm should be adapted since the use of small algorithms might lead to the exposure of their problems. Hash functions also face the problem of hash collision, where different inputs produce the same hash value. Salting is a practice where arbitrary data is added to each input before hashing, in an attempt to alleviate this problem[22].

The following Table 3 is an initial dataset that shows sensitive patient data in a healthcare context. Each record has a unique identifier for the PatientId, and each of the records also has gender, ethnicity, marital status, and preferred language. An example is FB2ABB23-C9D0-4D09-8464-49B70B982F0F0, who in this sample is a married male, African-American, and speaks Icelandic. Although this unique and identifiable information is indispensable for internal use, disclosure could pose a serious threat to privacy. By applying hashing methods that improve data security, the corporation generates the dataset displayed in Table 4.

The sensitive data would be PatientGender, PatientRace, PatientMaritalStatus, and PatientLanguage. These are hashed through various cryptographic methods. For example, XXHASH64 for the Patient Language, MD5 for the Patient Gender, SHA512 for the Patient Race, and CRC32 for the Patient Marital Status. The end result of this set of changes is to retain utility from original sensitive data for certain types of analysis, but in the changing of those values into unreadable hashed values.

Table 3. Original Dataset

PatientId	PatientGender	PatientRace	PatientMaritalStatus	PatientLanguage
FB2ABB23-C9D0-4D09-8464-49B70B982F0F0	Male	Unknown	Married	Icelandic
64182B95-EB72-4E2B-BE77-8050B71498CE	Male	African-American	Separated	English
DB22A4D9-7E4D-485C-916A-9CD1386507B	Female	Asian	Married	English

Table 4. Hashed Dataset

PatientId	PatientGender (MD5)	PatientRace (SHA512)	PatientMaritalStatus (CRC32)	PatientLanguage (XXHASH64)
FB2ABB23-C9D0-4D09-8464-49B70B982F0F0	$1$00IZNdMg$3Qyw6QCoTt25vG.MSOZI.	b90e54bb3e16afad4067bf777716387559083b84a3a794788f824od4d808fa3760f9d6f7d39673b17779868eb4f6f628f8e76101256a71476fe5a5587bb3c22	03cd7b91	ce023865e6969833
64182B95-EB72-4E2B-BE77-8050B71498CE	1UpOHW8F$Rbiks1HN9IYoHYKd3c6gehN/	C10988cfc79f1174893cf32c9c5266fb9a4984b05bb24170d50293262cfdb62be73b2994a87do705548a72367bbc04b2937e382a5ec743a90375da30a0ef56d3	d4296c7c	f7eefde29f3e16f2
DB22A4D9-7E4D-485C-916A-9CD1386507B	1Z0Y.V1HS$y1W6TuhjoMpu7YZQD6qp0	C33ea9d0ba47653638e278a8d3cab55856c064a148112d377b55fbe89b5cb17f6bfcfd709e9e989f32b245eca33640f2ecf857253d09b4ece807884f8aef35c051	03cd7b91	f7eefde29f3e16f2

We will point out the importance of the choice of an algorithm concerning how voluminous and sensitive the data is. SHA-256 is very often used in blockchain systems because of high security, but there is a deficit of processing speed. In this work, we are going to suggest using SHA-512, where this calculation expense can be eliminated since SHA-512 has higher security and possibly faster computing for some applications. The problem of its readiness to employ Table 4 remains its absence of jointly signed or declared hash values. These hash values can be validated by a process similar to that of mining in blockchains, wherein every hash is checked against a set of criteria and digitally signed to ensure their integrity and authenticity. Since this technique ensures that the system recognizes and accepts the hash values, the entire concept of data security is positively improved[23]. Since every transaction block from blockchain mining is being hashed and does need to meet certain specifications, then large processing power is needed in this regard. The participants could just keep hashing the block data while keeping minor modifications to find a suitable hash value that matches the testing criteria. This valid hash is then fed back into the network, which other nodes use to verify that indeed it is correct by hashing the respective block levels and comparing the result to the hash provided. If a hash meets all the criteria and is verified by the majority of nodes, then it is added to the blockchain.This allows every block to be valid and reliable. Likewise, the verification process

followed in this work can be followed using multi-nodes or organizations that digitally sign and verify the obtained hash values electronically. This ensures authenticity and system acceptability while considering all the security measures and data integrity. In this study, using a digital signature technique at every generated hash value may improve the validity in light of the hash functions. That is to say, a signature by a reliable source on each hash value ascertains that its value creation and verification were performed on conformities with the system's needed security. Secondly, the responsibility of validation can also be distributed among multiple nodes or even be taken by a decentralized validation process, similar to the mining process in blockchains. This latter approach is more secure since there is less risk of a single point of failure. Ensuring this provides confidence in the system, given that different parties independently verify and agree on hash numbers. Finally, ShinyAnonymizer implements various advanced hashing algorithms that notably enhance data security and integrity. Whereas SHA-512 is strong in cryptographic security, MD5 derives its strength from performance speed. CRC32 does error detection quite well, while XXHASH64 ensures good performance. Besides all that, salting is used with respect to the potential weaknesses issue to keep the hashed data secure, dependable, and appropriate for varied uses. The proper selection of a hashing method concerning data sensitivity and application requirements assures security and functionality. Careful selection of a hashing algorithm according to application requirements and data sensitivity ensures security and functionality.

5. Enhancements to ShinyAnonymizer Data Encryption Techniques

Data encryption is one of the most important and sophisticated transformation technologies used in various contemporary data security techniques. Data encryption transforms plaintext data into an unintelligible form, normally referred to as ciphertext, with the use of algorithms. Encryption keys ensure that the data will be accessible only to those who have the correct decryption keys. Encryption of sensitive data during transportation, retrieval, and analysis is essential to avoid any conceivable chance of unauthorized access[24].If data encryption is implemented, the selection of the appropriate encryption algorithm is a significant challenge. Among common symmetric encryption methods are DES, XDES, Blowfish, and Advanced

Encryption Standard (AES512). The selection methodology is based on the type of protected data, computation speed, and special security requirements. Given the excellent balance between security and efficiency, AES512 is regarded as a general-purpose standard that is regularly used to secure the data in transit or storage[25].

5.1 Technical Details of Encryption Techniques

5.1.1. DES

a.Description: DES is a symmetric block cipher that encrypts data in 64-bit blocks using a 56-bit key. Because of its susceptibility to brute-force assault, one of the most popular encryption algorithms has grown rather weaker nowadays.

b. Application: The inappropriately short length of keys and the vulnerability to computer powers have made the once powerfully effective DES encryption algorithm for the protection of sensitive data worthless.

5.1.2. XDES

a. Description A process applied to remove unique and identifiable data points and the direct identifier from a dataset is described in. This could mean the removal of any names, social security numbers, and any other direct identifier when scrubbing a database.
b. Application: This is, in effect, an expansion of DES. It implements many stages of DES encryption with different keys, increasing the key sizes and, thereby, the security.

5.1.3 BLOWFISH

a. Blowfish is a block cipher with symmetric keys. It encrypts 64-bit blocks of data using keys of variable length that range from 32 to 448 bits. It was generated to replace DES; it is faster and more reliable because it emits flexibility and efficiency.

b. Application: Blowfish has been applied to various forms of applications for the purpose of encrypting data. These include data, CDs, and virtual private networks amongst other vital

uses. As a result of the numerous applications in addition to the stringent safety precautions taken, it generally suits most of these applications.

5.1.4. AES512

a.Description:The AES-512 methodology takes 512 bits of input and key size, which provides resistance to cryptanalysis. The drawback of this approach is that it expands the allowed area. b.Application: AES-512 applies to any domain that requires high security and speed. It has a limited amount of chip space that may be devoted to a block cipher. Such will relate to satellite network communications, as well as entertainment applications.

Figure 3 summarizes our data encryption process. Accordingly, the initial step of the data encryption process is to select the proper database to which it should connect. The first step make sure the right data repository is used for all subsequent activities. Once connected, all necessary cryptographic functionalities shall be enabled, including the PostgreSQL pgcrypto module. Once everything is set up, the important move has to be finding out about the sensitive data kept in the database. Extensive studies has been done based on the sensitivity of the data that must be guarded and what the aftermath of the disclosure would be. This would then be followed by risk assessment to find potential dangers or vulnerabilities in the data. It can also be helpful in choosing the best technique of encryption.

After the completion of risk assessment processes, techniques of encryption that would be suitable are selected. Under the demands of security and efficiency, AES512 is an ideal option. Other methods can be used like RSA, Blowfish, or Triple DES, depending on the requirements. Having selected such algorithms, the sensitive data now becomes illegible; it gets transformed into a ciphertext and is incomprehensible to third parties.

After encryption, there is a verification phase to make sure that the data is safely encrypted without losing its validity or usefulness. Finally, the encrypted files are prepared for safe transmission or storage to ensure they remain safe throughout their total existence.

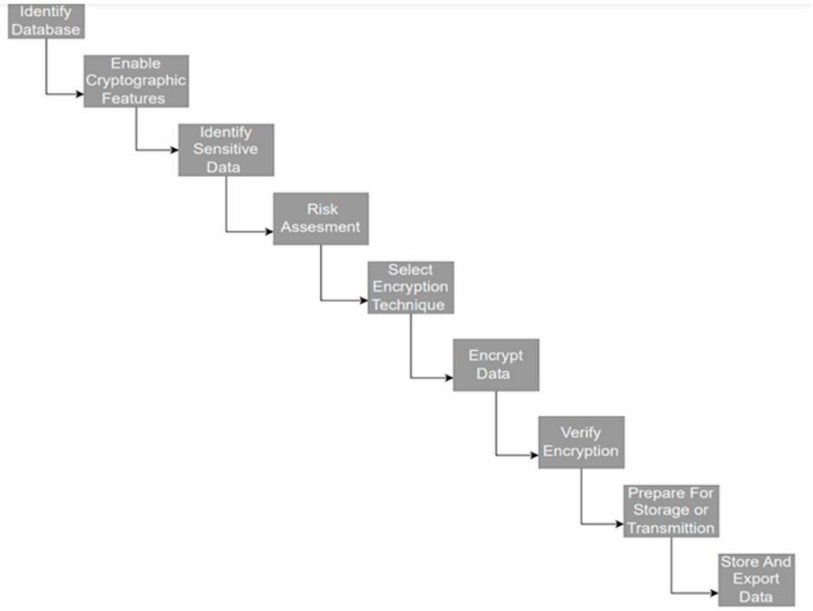


Figure 3. A visual representation of the encryption process.

Handling sensitive patient data in the healthcare industry is entrusted to maintain privacy and other regulatory requirements, such as GDPR and HIPAA. Think of a hospital that has a huge database of patients with their names, ages, gender, marital status, and languages. The data will be useful in conducting studies in epidemiology, clinical trials, and medical analysis. Due to the vulnerabilities of data breaches, all these information must first be anonymized before it can be used for other purposes. The following example shows how the hospital takes this data away using the enhanced ShinyAnonymizer tool and keeps patient privacy while maintaining the analytical value of the data. In the original dataset, there was information from three patients.

- Patient 1: An Unidentified married male of unknown race, speaking Icelandic, and is traceable only through his PatientId

- Patient 2: An African-American male residing alone, correctly traceable similarly by his PatientId.
- Patient 3: Asian female, good English speaker, married to a husband.

Table 5. Original Dataset.

PatientId	PatientGender	PatientRace	PatientMaritalStatus	PatientLanguage
FB2ABB23-C9D0-4D09-8464-	Male	Unknown	Married	Icelandic
64182B95-EB72-4E2B-BE77-	Male	African-American	Separated	English
DB22A4D9-7E4D-485C-916A-	Female	Asian	Married	English

Table 6 shows a scenario where multiple encryption schemes are used on different fields. Since the data points are sensitive in nature, the hospital employs the enhanced ShinyAnonymizer to anonymize each field by using strong encryption techniques. For example, PatientId yields a secure anonymous word once its anonymization is executed using the Data Encryption Standard. The XDES has been used for encrypting gender data; Blowfish for encoding race data, and AES512 has been used to encrypt marital status and language preferences

Table 6. Encrypted Dataset.

PatientId	PatientGender (DES)	PatientRace (XDES)	PatientMaritalStatus (BLOWFISH)	PatientLanguage (AES512)
FB2ABB23-C9D0-4D09-8464-49B70B982F0F0	J9..raL7sarlpgZHu3M	N7ez8ptklHSN2	$2a$06$dX.ox5Kw5UTea.PI1fpejeo/jxaVaKbQIq.9OrVnPHwb61NsI5Nis	\303\015\004\007\003\002\215\247\001\230\252:\270\\d\3228\001\003M/\
64182B95-EB72-4E2B-BE77-8050B71498CE	_J9..cJLyoOKPqPnzXro	uA/swfmosamBU	$2a$06$TTudssu1SSODVb9N6gWWO.wNydggPaPirel6UUFBF.feeA0ZNPM8u	\303\015\004\007\003\002\256\223\215\230\304\265\263\211z\01~L&\315
DB22A4D9-7E4D-485C-916A-9CD1386507B	_J9..hqKDaOxM7uNNmWU	\cumibnjFeBjF2	$2a$06$86GkCFeqVPctzovZQYYkS.vYP0qcu5QX3rdIP9f/Jj33ZaHO6uGzy	\303\015\004\007\003\002\302\\\304\341T\264\010\244n\3228\001\35

In the enhanced dataset, personal data will be removed in such a way that it will not be possible to re-identify any patient using the anonymized information, while the structure and

statistics needed for research may be kept. Using such encryption techniques will therefore safely provide data to be transmitted and assessed by the medical facility. Besides providing insight into patient profiles and health patterns, researchers can also use them in the course of exploratory data analysis and, by doing so, contribute to medical improvements, having privacy for the patients. Enhanced ShinyAnonymizer strikes the right balance between privacy and utility, hence allowing medical centers to commercially exploit their data assets while adhering to strict privacy laws.

Powerful encryption algorithms are provided with ShinyAnonymizer to achieve the latest and most secure data security and confidentiality. Some of the applicable algorithms and techniques include DES, XDES, Blowfish, and AES512, which go deeply into the very implementation of the encryption process for protecting private information against unwanted access while securing its analytical value. This approach involves flexibility for safe data management and study, assuring that all privacy problems have been overcome.

5.2. Complete Data Security: Combining Hashing, Encryption, and Anonymization

In the present digital era, new information and knowledge basically come from data itself. Therefore, complete data protection is something that one should not miss. With changing regulations, growing privacy concerns, and increasing incidents of data loss, a broader approach is in demand. The aim of this session is hence to take a closer look at how hashing, encryption, and data anonymization can help protect sensitive information while sustaining data integrity and to discuss both moral and legal demands of a society that may value privacy in general.

The first line of defense with anonymization of data deals with the maintaining of distinct identities inside the databases. The private information or PII must be removed or changed using techniques like generalization, suppression, and perturbation. When hashing and anonymization

are applied together, a quite sophisticated method is to be used. Sensitive characteristics from hashed identifying codes are used to create pseudonyms. The result is that this linkage allows significant data to be linked across a large number of data sets while offering protection for personal privacy without disclosure of first names related to the individuals in question. The integration of encryption with anonymization reduces the possibility of re-identification during the anonymization process itself because the critical attributes are already secured before the data is sent. Hashing is a vital procedure for verifying the correctness of data, because it turns data into hashes, which are unrecoverable types. Hashing is useful for merging numerous techniques, even outside of its typical use. The encryption process is made more difficult by hashed values, which are used as passwords or credentials for encryption techniques. Furthermore, encrypted IDs provide safe data linking without disclosing actual names when used in conjunction with data anonymization. Combining data anonymization, encryption, and hashing results in a comprehensive strategy that addresses the full data lifespan. Data security might be ensured by this integrated approach from the point of collection through all phases of processing, collaboration, and storage. It reduces the possibility of data unlawful access and re-identification.

This approach also preserves the analytical value of the data. Once eligible research and evaluations attain an appropriate balance between privacy and utility, they will benefit further from the corresponding and anonymized sets of data. In other words, data anonymization, encryption, and hashing together ensure a flexible yet robust backbone for generic data security.This is a symbiotic approach that overcomes the dynamic data environment challenges while sensitive data is protected. These three approaches will lead the future of data security: developing trust and best practices in the fast-growing digital footprint while technology is evolving and so are concerns about privacy.

5.3. Integration of Data Analysis Visualization and Statistics Paradigms

ShinyAnonymizer integrates statistical concepts into data analysis and visualization, helping scientists to gain meaningful insights from anonymized information with dynamism. Through the application, customers can perform an exploratory data analysis (EDA) of their health data and anonymize it using various statistical techniques in record time. By providing description statistics, a researcher can identify patterns and trends, hence assisting them in identifying the delivery, primary trend, and variability of anonymized data.

It further provides an easy interface that can combine paradigms for data display, allowing scientists to analyze anonymized data. There are various visualization techniques that exist in finding patterns and relations within data, including spreadsheets, box plots, and heatmaps.The representations can be used interactively because interactive dashboards allow real-time enhancements and fast input concerning anonymized data. Another important factor is efficiently communicating research output. It therefore enables the user to generate highly useful and appealing visuals that are easy to share with clients or colleagues.

Bringing the user closer to merging intricate methods and straightforward presentation, this tool will marry together powerful statistical analysis with interactive graphics, enriching the user experience by opening up intelligent analysis of the newly anonymized health data in a much more intelligent way. In a nutshell, ShinyAnonymizer uses scientific rules, data analysis, and visualization to offer health data privacy, while the researcher can perform deep statistical and graphical analyses. The general procedure enhances detailed and well-informed analysis with better insight through a deeper understanding of the modified datasets.

6. The Proposed Enhanced ShinyAnonymizer Tool

6.1. Workflow and System Architecture

ShinyAnonymizer is designed for business users who want to anonymize data with the support of a computer science expert. All the steps that make up the process are presented in Figure 4. The above figure depicts the anonymization process. We want to facilitate the adoption of data anonymization to a broader class of users that does not need to be necessarily acquainted with anonymization methods. We have made broad use of several techniques which are very easy for the non-expert user to set up and use.



Figure 4. System workflow.

This is mainly done by connecting, retrieving, and displaying data already existing in other databases, Excel and CSV files, and other formats. With the availability of several types of R scripts, ShinyAnonymizer is able to access a variety of relational databases and file formats currently in use. This is done in each case by having ShinyAnonymizer specify the appropriate connection settings to connect to the right data source and then access and analyze the available data. It is then necessary that the second step be presented to enable active discovery of publicly accessible data. The clients may play with the various privacy models and select what will be the data to be processed later. In step three, the client can select hashing, anonymization, and encryption techniques of various types. During the last phase of the process, participants are able to conduct various forms of analyses, including the validation through a number of different

privacy models of visible data. To perform precisely the same process described above, a three-tiered architecture made up of privacy models, GUIs, and Data Source APIs has also been developed and appears in Figure 5. It is a web-based tool where users can learn about data and understand through graphs and other forms of visualizations how different approaches to privacy affect that data. For building this layer, HTML, JavaScript, and CSS were used with the Shiny R package. After establishing the connection with the external data provider, the data were transformed and stored into an internal PostgreSQL database. The result would then be tabular representation of the data, where then the user can take an active role in each data field to decide which should be hashed, encrypted, or anonymized.

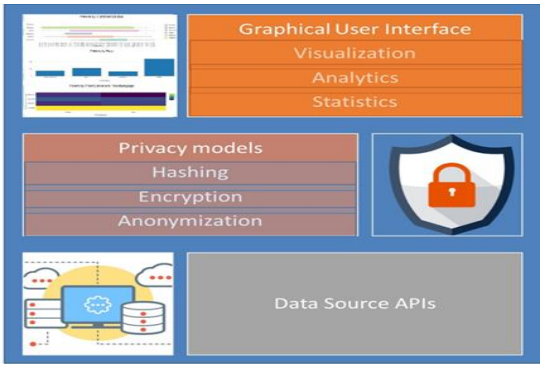


Figure 5. System architecture.

6.2. The GUI

Because the intention of this dynamic domain is to maintain privacy in health data management, the development of a proposed upgraded version of ShinyAnonymizer is spot on. Finding novel methods for enhancing individual data and preserving usability is becoming increasingly important in the dynamic domain because maintaining health data is intricate work. The Enhanced ShinyAnonymizer fixes this with many improvements and changes to fix any shortcomings of the first version. Nobody could undervalue the importance of privacy in health data handling. It did not stem from laws and morals but also from maintaining and establishing confidence in health institutions among the public with great strength. This means that with

fewer electronic medical records, there are going to be increased risks of data breaches that will put people's privacy at risk and damage the public's trust. The ShinyAnonymizer Enhanced Version is a pragmatic solution to broader safety concerns for the health sector when it put into use anonymizing health data. This development reflects a frame of thinking in hitting the perfect balance between user privacy and open data access. To understand the application's growth, it is important to understand ShinyAnonymizer's history and its purpose in the medical field. Even though the initial idea was only to anonymize health data, with the changing circumstances, it became clear that something better had to be developed. The ShinyAnonymizer Enhanced Version does more than basic anonymization by adding more advanced features, while also fixing issues present in the first version. This is discussed on this page. It aims to introduce ShinyAnonymizer as a novel, integrated health data management platform for privacy protection. The research study aims to present the development of an improved version of ShinyAnonymizer by pinpointing and solving problems of current privacy-preserving technologies in use within the healthcare industry. Meeting these objectives enables the research to contribute to the active debate on privacy in the management of health data. It provides a workable approach to balancing patient privacy and data availability for researchers and medical practitioners. The program above also provides a GUI-based tool with an easy-to-use interface that allows importing statistics, info-graphics, and numerous methods of anonymizing data, as depicted in Figure 6.

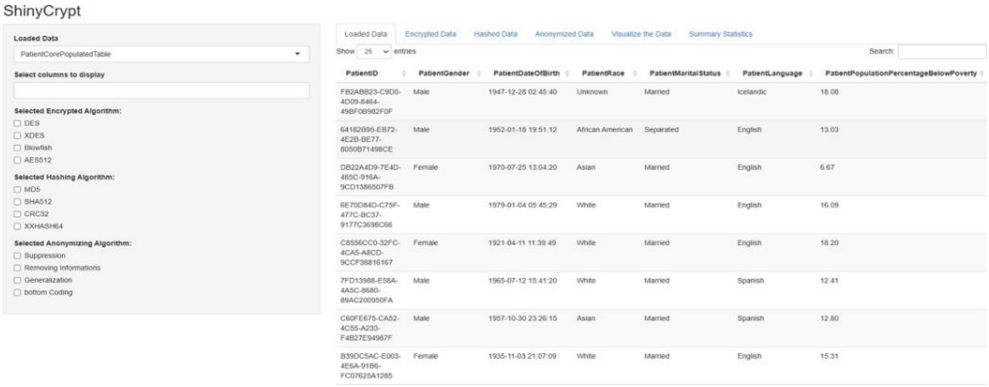


Figure 6. ShinyAnonymizer interface.

In summary, the enhanced ShinyAnonymizer is a very good example of Privacy-Enhancing Technology, which enables handling health data in an ethically and securely correct manner efficiently and in a user-friendly way.

7. Real World-Use Cases

Its performance in real-world scenarios demonstrates that ShinyAnonymizer Enhanced Version is capable of offering protection for patient privacy.

7.1. Clinical Research

Nowadays, the tool of ShinyAnonymizer can be helpful in the research field as far as the anonymization of patient data is concerned. Researchers can now apply modern anonymization methods, integrate electronic health information, and conduct comprehensive studies without having to make any trade-off on patients' privacy. Because the program is flexible in handling a variety of data types, it can be applied in a wide range of experimental settings

7.2. Healthcare Analytics

ShinyAnonymizer proves to be quite helpful for healthcare businesses based on data analytics, considering these are the ways through which cost-effectiveness and improvements in health services could be envisaged. The enhanced version provides flexible anonymization methods that make it easier to handle large datasets in a secure way while maintaining privacy regulations.The integrity and confidentiality of the data could be kept, enabling analytics teams in healthcare to learn useful insights while sensitive patient data remains protected.

7.3. Research on Public Health

In healthcare research, data is being collected from a large number of sites. The health professionals prefer using ShinyAnonymizer because of the wide range of anonymization

possibilities and flexibility regarding different formats of the data. The program protects users' identities but at the same time allows fast study and criticism of the retrieved data collection.

7.4 Collaborative Research Efforts

If ShinyAnonymizer is used, one can have secure contact much faster. Sharing of data in joint projects necessitates improved verification. Academics will then share anonymized datasets without the fear of being re-identified. Collaboration and teaching of scientists about anonymized data are enabled through statistics and data analysis.

7.5. Educational Environments

Concerning the possible educational use, the ShinyAnonymizer has the potential to become an indispensable tool for students with a knack for manipulating data and health analytics. It allows students to work on real health data and gain such important management and anonymization skills. The added value of this application is indeed more practical experience; hence, increasing the chances of their success in managing health data. In the end, the ShinyAnonymizer Enhanced Version gives good reason to invest more in handling healthcare data.

This flexibility and efficiency of the tool regarding privacy maintenance and insightful data analysis are reflected in the broad range of settings within which it could be employed. It would also prove helpful in instances where scientists, corporations, or healthcare workers face challenges with privacy handling sensitive patient data.

A part of the integral use involves handling multiple types of health data. Because of the different procedures, kinds of data, and norms adopted, electronic medical data is complex. The ability of ShinyAnonymizer Enhanced Version to process and anonymize both structured and unstructured data strengthens its usability within the modern health care sector.

The other important aspect of using the tool is that it can handle a number of different health data types. By nature, health data systems are likely to be complicated by different types of data and norms. The enhanced version of ShinyAnonymizer further increases its practicality in

everyday settings through a fully interactive, intuitive user interface. The study of data by researchers and clinicians, along with their statistical evaluations and visualizations in a comprehensive and clear way, may support decision-making that is informed and educated. This allows for effective communication regarding research findings by developing attractive and insightful plots.

8. Discussion and Implications in the Contents of Privacy Regulations

Indeed, the practical uses of ShinyAnonymizer are hugely promising in their contribution to making this dream of individually controlled privacy policies a reality where the use and storage of health data are no longer a liability. As a result, the technology lets full compliance with stringent confidentiality requirements, such as the HIPAA in the US and the General Data Protection Regulation in the EU, be achieved with the anonymization of patient data. Strict security and caution engage them in the protection of private patient data. ShinyAnonymizer, updated to its latest version, provides an agile way to anonymize health data and to eliminate, as such, most of the legal problems. These applications show the awareness of privacy and secrecy and exist in corporate, academic, medical, and scientific fields. The regulations also provide that medical institutions should create security procedures regarding changes in management affecting personal information. The Enhanced Version of ShinyAnonymizer will provide the functionality of setting business policy in such a way that prohibits illegal activities and guarantees, under all circumstances, the safe processing of data. Moreover, based on guidelines derived from such policies, risk decreases and trust increases. The improved version is integrated with contemporary state-of-the-art privacy-preserving techniques: hashing, encryption, and anonymization. It covers requirements related to privacy, HIPAA, and GDPR, offering complete protection against re-identification. It stresses how fundamental the tool is in terms of maintaining privacy regulations and ensuring the correctness of programs by using an integral security approach that avails multiple techniques. Due to the ease of usage of the tool,

investigators could now collaborate, with anonymized data, investigational medical facilities, and educational institutions, while still preserving privacy and ethical principles. The ShinyAnonymizer Enhanced Version also has implications for security procedures involving data throughout all parts of its life cycle, from collection to removal. Many sets of privacy laws explain how information shall be collected, processed, used, and disposed of throughout its life. Security measures regarding health data would help organizations reduce the risk of legal complications or damage to reputation due to data leakage. The enhanced version advocates a responsible approach toward data and ethics with its simple graphical user interface. It articulates enterprise responsibility pertaining to the right to privacy and allows flexibility in da-ta processing. The tool is designed to easily support investigators, healthcare facilities, and educational institutions in communicating effectively with anonymized data while meeting ethical and privacy standards.

9. Results and Experiments

The Enhanced Version of ShinyAnonymizer, by its modern and visually attractive design, serves as a very good foundation for testing data and producing an accurate assessment of the results. Researchers can show full insight from anonymized information on the development of relationships, patterns, and trends by using different visualization options. This idea should adjust data structures in the direction of disclosing new opportunities and sharing study outcomes within dynamic dashboards with readable graphics. This constant review at times enriches the applicability and comprehensibility of findings from data models. Additionally, with estimates and statistical data visualizations for mean, variance, and median along with graphical visualizations, properties of the dataset are also mined deeply.

Laws also require security procedures to be executed prior to changes being made in personal information by healthcare organizations themselves. Therefore, with the ShinyAnonymizer Enhanced Version, it is possible to enable business to create policies that prevent illegal activity

and ensure data processing security under any circumstance. Consequently, following such rules reduces risks and builds trust. The advanced version of the tool provides additional features supporting enhanced privacy-preserving mechanisms, such as hashing, encryption, and data anonymization. It also provides a road map for compliance with legislation on data protection. Compliant with GDPR and HIPAA regulations, full data protection against reidentification of stolen data is ensured.

9.1. Case Study: Clinical Study on Patient Attributes

The objective of this clinical study is to investigate the connection between different patient features, including gender, race, marital status, and language. In order to examine such relationships, we have developed three different types of visualizations:

1. Box plot Below is a graph showing the age of the patient obtained from the PatientId for each of the different marital statuses. On the y-axis is the number of the ages while on the x-axis, the different marital statuses are placed: single, married and divorced.

To show general patterns, summary data are given for each marital status category: these include mean and median ages. The following measures of statistics are used to analyze the range of age for each category of marital status: (1) and (2)

- Mean (m):

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

where N denotes age of the i-th patient and x_i - the number of patients within a marital status group.

- Median (M):

$$M = \begin{cases} x\left(\frac{N+1}{2}\right), & \text{if } N \text{ is odd} \\ \frac{x\left(\frac{N}{2}\right) + x\left(\frac{N}{2}+1\right)}{2}, & \text{if } N \text{ is even} \end{cases}$$

where xi represents i-th patient order statistic, that is, the smallest value in the sorted data [28].

2. Bar plot: This chart represents the distribution of patients according to their gender. The number of patients is carried on the y-axis, while the gender categories, such as male and female, are on the x-axis. In each line, the number of patients in each of the gender categories is shown. Other data may be used, such as the percentage of males and females, to indicate clearly the gender distribution. Where the median number reflects the centre tendency, and the total % reflects an average gender distribution[29,30]. The total number of patients by gender can be worked out with formula 3 below:

- Percentage (P):

$$P = \left(\frac{n_g}{N}\right) \times 100$$

3. Area plot: It provides an overview of the variation in accessibility for patients depending on their language. On the x-axis, it displays different times maybe in years, while on the y-axis, it indicates the number of patients that can understand the particular language. Each section in this graph represents one language, but the joint pieces represent variations in the distribution of each language. The average and median of the total number of patients understanding each language within every monitoring period are as explained below[31]. Formula 4 and Formula 5 are used to ascertain the variation in language dispersal:

- Average number of patients who comprehend a language during a certain period of time:

$$\mu_L = \frac{1}{T}\sum_{t=1}^{T} n_{L,t}$$

where, T is the total number of periods, nL,t denotes the number of patients who under-stand the language L at time t.

• Median percentage of patients able to understand a language:

$$M_L = \begin{cases} n_{L,}\left(\frac{T+1}{2}\right), & if\ T\ is\ odd \\ \frac{n_{L,}\left(\frac{T}{2}\right)+n_{L,}\left(\frac{T}{2}+1\right)}{2}, & if\ T\ is\ oven \end{cases}$$

where nL,(t) is the number of patients that understand language L at the t-th highest statistic in a given time period.

This will be achieved by highlighting the trends in language choice over time, combining the descriptive data with visual assistance to understand changes in language distribution at different points in time. From the figures below, we see how the speech patterns of the patients have changed and what is the normal language they use. A dashboard of these examples and their respective explanations are shown in Figures 7 and 8.

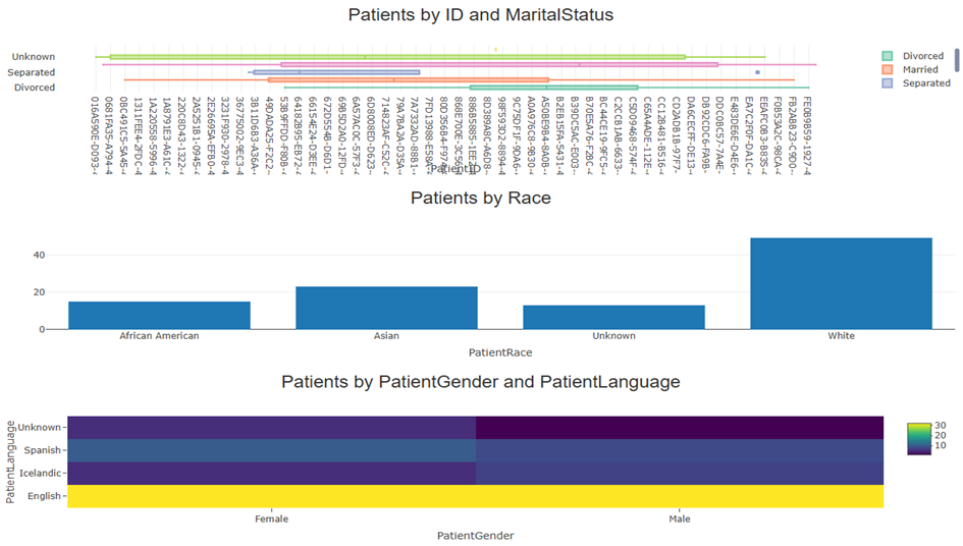


Figure 7. Graphical representation of patients.

```
 PatientID         PatientGender      PatientDateOfBirth                PatientRace        PatientMaritalStatus PatientLanguage
Length:100         Length:100         Min.   :1921-03-26 14:38:51.7999  Length:100         Length:100           Length:100
Class :character   Class :character   1st Qu.:1944-11-02 13:55:33.7189  Class :character   Class :character     Class :character
Mode  :character   Mode  :character   Median :1958-03-04 19:43:17.3429  Mode  :character   Mode  :character     Mode  :character
                                      Mean   :1957-03-31 23:35:33.6269
                                      3rd Qu.:1969-12-16 07:14:43.6652
                                      Max.   :1988-11-25 02:59:36.3730
PatientPopulationPercentageBelowPoverty
Min.   : 1.70
1st Qu.:12.70
Median :15.44
Mean   :22.10
3rd Qu.:18.69
Max.   :98.40
```

Figure 8. Descriptive statistics of patients.

.

Moreover, we want to visualize the detailed data by using three kinds of visualizations box plot, bar plot, and area plot together with descriptive statistics like mean and median. Descriptive statistics and the visualizations that have been suggested give us the opportunity to identify interesting patterns and trends in the data, such as-but not limited to age distribution of patients by marital status groups, gender distribution in patients, and the trend in language assistance over time. Applying these statistical methods may drive certain pattern detection or issues that drive the utilization of products and services within healthcare. This further promotes the health of the

patients since the doctors, academics, and elected leaders are educated to make informed decisions.

9.2. Comparison with Other Works

There are several options to provide data anonymization, and some of them are open-source. In the context of this paper, the most famous ones are SdcMicro, Anonymizer, and ARX. In SdcMicro, the focus is on the anonymization of data to protect it from unauthorized accesses. The anonymized data or documents then can be utilized for general purposes and for research purposes. It provides a set of functions for the risk assessment. It performs data loss and its impact on continuous variables by using the statistics connected. The SdcMicro package allows for several risk and productivity evaluations and also anonymization methods. Along with different steps of anonymization, it may also provide estimates, methods for evaluating and displaying the risk and utility. However, the interfaces of SdcMicro are not user-friendly, which makes it difficult to effectively use by non-experts. Its scalability and flexibility are limited and large-scale applications may be at a disadvantage due to these issues[32].

Anonymizer is yet another anonymization tool that speeds up the process of anonymizing data, especially PII, by combining salting and hashing or "convenience functions." It is an extremely lightweight tool with no advanced cryptographic and/or risk assessment methods. When large datasets were to be processed, it performed inadequately, where detailed analysis and reporting tasks were at hand. Its design allows for fast and easy anonymization operations[27].

Speaking about data anonymization, ARX is designed to be both versatile and deployable. It supports multiple anonymization methods, including data analysis methods and re-identification risk-reduction methods. Among common anonymization methods used by ARX, there are also k-anonymity, t-closeness, l-diversity, and differential privacy. ARX is adequate for a broad scope of tasks because of its decent rate and expressiveness[26]. However, its interface is too complicated for non-technical people, and it takes loads to install and operate efficiently. While ❌ denotes the trait with its absence, Table 7 compares those qualities with ✔ in respect of the

specified aspect. For instance, SdcMicro and ARX do not have "Data Encryption" or "Hashing," while the enhanced version of ShinyAnonymizer and Anonymizer does. Other features such as scalability, flexibility, and efficiency all share the same rating, depending on how friendly the "User Interface" is.

Table 7. Comparison with other works.

Aspect	ShinyAnonymization Enhanced Version	SdcMicro	Anonymizer	ARX
Data Anonymization	High	Moderate	Moderate	High
Data Encryption	✔	✘	✔	✘
Hashing	✔	✘	✔	✘
User Interface	User Friendly	Complex	User Friendly	Complex
Flexibility	High	Low	Moderate	✔
Scalability	High	Low	Moderate	✔
Performance	High	Low	Moderate	✔
Compliance with Laws	✔	✔	✔	✔

Besides the intuitive interface, what really makes this new ShinyAnonymizer great is the respect of privacy regulations. Advanced functionalities include data anonymization, encryption, and hashing. ShinyAnonymizer easily scales and is flexible with SdcMicro for everything from big healthcare analytics to smaller-scale examination. Due to its simple interface, those nontechnical persons can easily learn and take its functionality, unlike ARX's complex setup. Again, due to strong encryption techniques applied, ShinyAnonymizer keeps data safe while anonymizing them. In this respect, ShinyAnonymizer covers the exact requirement which was totally missing in Anonymizer. This application manages huge datasets very accurately outperforming their speed. With comprehensive data analysis and visualization options, ShinyAnonymizer provides the user with information on data analysis that may lead to decisions and discoveries one can put a lot more faith in. In general, the ShinyAnonymizer Enhanced Version stands out from its competition by combining significant anonymization algorithms, encryption, ease of use interfaces, speed, and flexibility. It can therefore become an excellent choice for companies wanting an in-depth and efficient manner of dealing with sensitive data without facing any privacy regulations compromise.

10. Beyond ShinyAnonymizer: Future Directions

Besides what ShinyAnonymizer is currently doing, there are several interesting directions that could be taken in order to enhance the capabilities of current and future privacy-preserving technologies dealing with sensitive data:

10.1. Artificial Intelligence and Learning

First, AI and machine learning drive the development of increasingly complex anonymization methodologies. The future anonymization solutions will precisely and effectively mask sensitive data in big datasets with the aid of AI algorithms. This will preserve the analytical value of the data while reducing privacy violations. From the following Formula for data utility, one would be able to see how good AI is in anonymizing data:

$$U = 1 - \frac{L_{anonymized} - L_{original}}{L_{original}} \quad (6)$$

where Lanonymized is the missing function for the anonymized data and Loriginal is the missing function for the original data. This formula sets up how well anonymized data preserves the meaning of the original data.

10.2: Blockchain Technology

It also saves a lot of money as the number of mediators and organizations involved in the process is reduced. Additionally, this technology can greatly increase the standards for ensuring privacy and medical record security. Such techniques will allow organizations to build transparent and tamper-proof verification schemes for the receipt and modification of health-related data. The integrated approach of ShinyAnonymizer and blockchain technology could be

beneficial in case of higher levels of personal information protection against illegal access and modification[33]. A blockchain's integrity can be depicted in the following examples (7):

$$I = \sum_{i=1}^{N} \left(\frac{1}{H_i} \right)$$

Where N is the total number of blocks and Hi is the hash of the block that is i-th in place and guarantees integrity for each of them.

10.3. Differential Privacy

The addition of noise in responses to queries makes the individual records indistinguishable, yet valuable information can still be derived. Sophisticated types of user privacy guarantee may be enabled by differential privacy models that can constitute part of future anonymization systems[34]. The following is the representation of the mathematical expression for differential privacy 8:

$$M(D) = f(D) + Lap\left(\frac{\Delta f}{\epsilon} \right)$$

When query function name is f(D) , the sensitivity level of the function is Δf, private differential value for dataset D, is denoted by M(D), and privacy restriction specified by ϵ.

10.4: Federated Education

Federated learning is a new model of data analysis with privacy. In federated learning systems, machine learning models are developed among various dispersed computers or devices in manners that need not transmit data. Thus, it permits companies to collaborate on data analysis while sensitive data remains localized and confidential. Therefore, the features of federated learning could be one of the features included in future anonymization software versions to conduct collaborative research without risking the privacy of individuals' data [35]. The following formulates the federated learning formula, where:

$$\min_{w} = \left(\frac{1}{N} \sum_{i=1}^{N} L_i(w) \right)$$

Let N clients, a set of variables w,, and a local loss rate for client i is $Li(w)$.

10.5. Homomorphic Encryption

Homomorphic encryption keeps the data secret but permits functional analysis because it can allow computations on encrypted data. For example, to permit the investigation by organizations, ShinyAnonymizer can utilize homomorphous encryption techniques alongside anonymization in order not to risk leakage and illegal access. In the following, some abbreviations are defined for homomorphic encryption: (10)

$$E\big(f(x)\big) = f\big(E(x)\big)$$

where x denotes plaintext data, f is the computation function, and E is the encryption function. It guarantees the feasibility to perform computations on encrypted data.There is, therefore, a huge need for innovation and advancement of data management concepts that protect privacy. Future versions of anonymization tools like ShinyAnonymizer could perhaps assure protection of privacy far more reliably while allowing for essential data analysis and collaboration in the healthcare industry and beyond, by use of techniques such as AI, blockchain, differential privacy, federated learning, and homomorphic encryption[36].

11. Conclusions

It is more important that privacy-enhancing technologies, such as ShinyAnonymizer, be developed and implemented first, so sensitive data may be handled within the rule of law, in particular for the healthcare domain. We have analyzed the key characteristics of enhanced research techniques, the technical limitations, and potential uses of ShinyAnonymizer. Also, we debated how changes in privacy regulations may be influenced by the development of privacy-enhancing technologies in the coming years. ShinyAnonymizer strikes an optimal balance between usefulness and security for a pressing need to protect patient privacy due to its robust

hashing, encryption, and anonymization methods. Also, it enables smart data analysis and collaboration. Flexibility is shown in its advantageous application to clinical trials, group research, epidemiology, healthcare analysis, and educational institutions. It is more confident when the procedures are in adherence to the security regulations such as GDPR and HIPAA.Future improvements in blockchain, artificial intelligence, homomorphic encryption, federated research, and differential privacy will further extend what is possible ethically when it comes to the management of data and cooperation on data. Since ShinyAnonymizer treats sensitive data with consideration and respect, its development and deployment will continue to be useful in health and related areas. ShinyAnonymizer represents one of the most relevant steps toward developing privacy-preserving data managing systems, a key component in realizing improvements in the healthcare sector and beyond.

12. Bibliography of the Chapter

1. Shojaei, P.; Vlahu-Gjorgievska, E.; Chow, Y.-W. Security and privacy of technologies in health information systems: A systematic literature review. *Computers* **2024**, *13*, 41. https://doi.org/10.3390/computers13020041.
2. Almulihi, A.H.; Alassery, F.; Khan, A.I.; Shukla, S.; Gupta, B.K.; Kumar, R. Analyzing the Implications of Healthcare Data Breaches through Computational Technique. *Intell. Autom. Soft Comput.* **2022**, *32*, 1763–1779. https://doi.org/10.32604/iasc.2022.023460.
3. Kondylakis, H.; Despoina, M.; Glykokokalos, G.; Kalykakis, E.; Karapiperakis, M.; Lasithiotakis, M.-A.; Makridis, J.; Moraitis, P.; Panteri, A.; Plevraki, M.; et al. EvoRDF: A framework for exploring ontology evolution. In *The Semantic Web: ESWC 2017 Satellite Events, Proceedings of the ESWC 2017 Satellite Events, Portorož, Slovenia, 28 May–1 June 2017*; Revised Selected Papers 14; Springer International Publishing: Cham, Switzerland, 2017. https://doi.org/10.1007/978-3-319-70407-4_20.
4. Adeniyi, A.O.; Arowoogun, J.O.; Okolo, C.A.; Chidi, R.; Babawarun, O. Ethical considerations in healthcare IT: A review of data privacy and patient consent issues. *World J. Adv. Res. Rev.* **2024**, *21*, 1660–1668. https://doi.org/10.30574/wjarr.2024.21.2.0593.
5. Bakare, S.S.; Adeniyi, A.O.; Akpuokwe, C.U.; Eneh, N.E. Data privacy laws and compliance: A comparative review of the EU GDPR and USA regulations. *Comput. Sci. IT Res. J.* **2024**, *5*, 528–543. https://doi.org/10.51594/csitrj.v5i3.859.
6. Guide, A. *Cybersecurity Resource. Implementing the Health Insurance Portability and Accountability Act (HIPAA) Security Rule*; National Institute of Standards and Technology: Gaithersburg, MA, USA, 2024. https://doi.org/10.6028/NIST.SP.800-66r2.
7. Williamson, S.M.; Prybutok, V. Balancing privacy and progress: A review of privacy challenges, systemic oversight, and patient perceptions in ai-driven healthcare. *Appl. Sci.* **2024**, *14*, 675. https://doi.org/10.3390/app14020675.
8. Clayton, E.W.; Embí, P.J.; A. Malin, B. Dobbs and the future of health data privacy for patients and healthcare organizations. *J. Am. Med. Inform. Assoc.* **2023**, *30*, 155–160. https://doi.org/10.1093/jamia/ocac155.
9. Vardalachakis, M.; Kondylakis, H.; Koumakis, L.; Kouroubali, A.; Katehakis, D. ShinyAnonymizer: A Tool for Anonymizing Health Data. In Proceedings of the 5th International Conference on Information and Communication Technologies for Ageing Well and e-Health, Crete, Greece, 2–4 May 2019; pp. 325–332.
10. Paul, M.; Maglaras, L.; Ferrag, M.A.; Almomani, I. Digitization of healthcare sector: A study on privacy and security concerns. *ICT Express* **2023**, *9*, 571–588. https://doi.org/10.1016/j.icte.2023.02.007.
11. Dhirani, L.L.; Mukhtiar, N.; Chowdhry, B.S.; Newe, T. Ethical dilemmas and privacy issues in emerging technologies: A review. *Sensors* **2023**, *23*, 1151. https://doi.org/10.3390/s23031151.

12. Hossain, M.T. Privacy and Security for Trustworthy AI/ML in Multi-Agent Critical Infrastructures: An Analysis of Adversarial Dynamics and Protective Strategies. Doctoral Dissertation, University of Nevada, Reno, Reno, NV, USA, 2024. Available online: https://scholarwolf.unr.edu/home (accessed on 4 August 2024).
13. Paramesthi, P.; Jati, S.P.; Suryoputro, A. The use of Electronic Medical Record (EMR) in hospitals during the COVID-19 pandemic in Indonesia: A systematic literature review. *BKM Public Health Community Med.* **2024**, *40*, e11727. https://doi.org/10.22146/bkm.v40i01.11727.
14. Vardalachakis, M.; Kondylakis, H.; Tampouratzis, M.; Papadakis, N.; Mastorakis, N. Anonymization, Hashing and Data Encryption Techniques: A Comparative Case Study. In Proceedings of the 2023 International Conference on Applied Mathematics & Computer Science (ICAMCS), Lefkada Island, Greece, 8–10 August 2023; pp. 129–135.
15. Marques, J.; Bernardino, J. Analysis of Data Anonymization Techniques. In Proceedings of the 12th International Conference on Knowledge Engineering and Ontology Development, Budapest, Hungary, 2–4 November 2020; pp. 235–241.
16. Turgay, S.; İlter, I. Perturbation methods for protecting data privacy: A review of techniques and applications. *Autom. Mach. Learn.* **2023**, *4*, 31–41. https://doi.org/10.23977/autml.2023.040205.
17. Vovk, O.; Piho, G.; Ross, P. Methods and tools for healthcare data anonymization: A literature review. *Int. J. Gen. Syst.* **2023**, *52*, 326–342. https://doi.org/10.1080/03081079.2023.2173749.
18. Hasan, H.A.; Al-Layla, H.F.; Ibraheem, F.N. A Review of Hash Function Types and their Applications. *Wasit J. Comput. Math. Sci.* **2022**, *1*, 75–88. https://doi.org/10.31185/wjcm.52.
19. Kumar, K.K.; Ramaraj, E.; Srikanth, B.; Rao, A.S.; Prasad, P. Role of MD5 Message-Digest Algorithm for Providing Security to Low-Power Devices. In Proceedings of the 2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 25–27 May 2022; pp. 352–358.
20. Al-Shareefi, F.; Al-Barmani, Z. Comparing two cryptographic hash algorithms: SHA-512 and whirlpool—A case study on file integrity monitoring. *BIO Web Conf.* **2024**, *97*, 00093. https://doi.org/10.1051/bioconf/20249700093.
21. Yusuf, A.D.; Abdullahi, S.; Boukar, M.M.; Yusuf, S.I. Collision resolution techniques in hash table: A review. *Int. J. Adv. Comput. Sci. Appl.* **2021**, *12*, 757–762. https://doi.org/10.14569/ijacsa.2021.0120984.
22. Joshua, T. A Secure Model for Student Results Verification Using Salted Hash Functions. Master's Thesis, Kampala International University, Kampala, Uganda, 2023. Available online: https://irbackend.kiu.ac.ug/server/api/core/bitstreams/c74b76c4-3f75-4393-990a-2ce5638b2ad0/content (accessed on 4 August 2024).
23. Kuznetsov, O.; Peliukh, O.; Poluyanenko, N.; Bohucharskyi, S.; Kolovanova, I. Comparative Analysis of Cryptographic Hash Functions in Blockchain Systems. In Proceedings of the CPITS-2023-II: Cybersecurity Providing in Information and Telecommunication Systems, Kyiv, Ukraine, 26 October 2023; pp. 81–94. Available online: https://ceurspt.wikidata.dbis.rwth-aachen.de/Vol-3550/paper7.html (accessed on 4 August 2024).
24. Atadoga, A.; Farayola, O.A.; Ayinla, B.S.; Amoo, O.O.; Abrahams, T.O.; Osasona, F. A Comparative Review of Data Encryption Methods in the USA and Europe. *Comput. Sci. IT Res. J.* **2024**, *5*, 447–460. https://doi.org/10.1002/asi.24369.
25. Rameel, M.; Asif, Z. Fortifying Information Security: A Comparative Analysis of AES, DES, 3DES, RSA, and Blowfish Algorithm. *Communications* **2024**, *2*, 5.
26. Prasser, F.; Eicher, J.; Spengler, H.; Bild, R.; Kuhn, K.A. Flexible data anonymization using ARX—Current status and challenges ahead. *Softw. Pract. Exp.* **2020**, *50*, 1277–1304. https://doi.org/10.1002/spe.2812.
27. Hendricks, P. Anonymizer: Anonymize Data Containing Personally Identifiable Information. R Package Version 0.2.0. 2015. Available online: https://github.com/paulhendricks/anonymizer (accessed on 4 August 2024).
28. Newbold, P.; Miller, W.L.; Thorne, R. *Statistics for Business and Economics*, 9th ed.; Pearson Education Limited: Harlow, UK, 2013.
29. Tukey, J.W. Exploratory Data Analysis. Addison-Wesley Publishing Company Reading, Mass.—Menlo Park, Cal., London, Amsterdam, Don Mills, Ontario, Sydney 1977, XVI, 688 S. *Biom. J.* **1981**, *23*, 413–414. https://doi.org/10.1002/bimj.4710230408.
30. Healy, K. *Data Visualization: A Practical Introduction*; Princeton University Press: Princeton, NJ, USA, 2018.
31. Shumway, R.H.; Stoffer, D.S. *Time Series Analysis and Its Applications: With R Examples*, 4th ed.; Springer Nature: Dordrecht, The Netherlands, 2017.
32. Templ, M.; Kowarik, A.; Meindl, B. Statistical disclosure control for micro-data using the R package sdcMicro. *J. Stat. Softw.* **2015**, *67*, 1–36. https://doi.org/10.18637/jss.v067.i04.
33. Vaigandla, K.K.; Karne, R.; Siluveru, M.; Kesoju, M. Review on blockchain technology: Architecture, characteristics, benefits, algorithms, challenges and applications. *Mesopotamian J. CyberSecurity* **2023**, *2023*, 73–85. https://doi.org/10.58496/mjcs/2023/012.
34. Vasa, J.; Thakkar, A. Deep learning: Differential privacy preservation in the era of big data. *J. Comput. Inf. Syst.* **2023**, *63*, 608–631. https://doi.org/10.1080/08874417.2022.2089775.

35. Wen, J.; Zhang, Z.; Lan, Y.; Cui, Z.; Cai, J.; Zhang, W. A survey on federated learning: Challenges and applications. *Int. J. Mach. Learn. Cybern.* **2023**, *14*, 513–535. https://doi.org/10.1007/s13042-022-01647-y.
36. Munjal, K.; Bhatia, R. A systematic review of homomorphic encryption and its contributions in healthcare industry. *Complex Intell. Syst.* **2023**, *9*, 3759–3786. https://doi.org/10.1007/s40747-022-00756-z.

Appendix A - *Publications*