INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT (IJSREM)
VOLUME: 08 ISSUE: 04 | APRIL - 2024 SJIF RATING: 8.448 ISSN: 2582-3930

Privacy-Preserving Data Sharing Platform

Abhishek Mishra
Apex Institute of Technology
Chandigarh Univesity
abhishek13mishra13@gmail.com

***Abstract*— In today's data-driven healthcare landscape, the secure sharing of sensitive medical information is essential for improving patient care, facilitating medical research, and advancing healthcare outcomes. However, ensuring the integrity, confidentiality, and privacy of patient data poses significant challenges, particularly in the context of big data environments. This presents a comprehensive framework for privacy-preserving data sharing in healthcare, leveraging a combination of cryptographic techniques, encryption, and secure computation protocols. The framework encompasses various privacy-preserving mechanisms, including Differential Privacy with Data Perturbation, Secure Multi-Party Computation (SMPC), and Homomorphic Encryption, to protect sensitive healthcare data from unauthorized access and disclosure. By implementing state-of-the-art privacy-preserving techniques, the framework aims to enable secure data sharing among multiple parties while complying with regulatory requirements such as HIPAA and GDPR. Additionally, the paper discusses the project scope, which includes cryptography, encryption, decryption, integrity, confidentiality, privacy, policies, procedures, security, and secure data sharing infrastructure. The proposed framework provides a practical solution for healthcare organizations and research institutions to collaborate on data-driven initiatives while safeguarding patient privacy and maintaining trust. Evaluation of the framework's effectiveness and performance metrics is conducted to validate its feasibility and efficacy in real-world healthcare settings.**

Keywords: ***Privacy-preserving data sharing, Differential Privacy, Data Perturbation, Secure Multi-Party Computation (SMPC)***

I. INTRODUCTION

In today's era of digitalization and big data, the healthcare industry is confronted with the challenge of securely managing and sharing vast amounts of sensitive patient information. The increasing adoption of electronic medical records (EMRs), medical imaging data, genomic data, and other healthcare datasets has led to a surge in data volumes, presenting significant privacy and security concerns. Ensuring the confidentiality, integrity, and privacy of patient data is paramount to maintaining patient trust, complying with regulatory requirements, and fostering collaborative research and innovation in healthcare.

In the current healthcare landscape, the revolution of patient records and the progression of data-driven technologies have transformed the provision of medical services, diagnosis, and research. Nevertheless, this data-centric framework raises significant privacy apprehensions, especially concerning the dissemination and examination of confidential healthcare data. Safeguarding patient confidentiality while facilitating data exchange for research, analysis, and decision-making poses a critical challenge for healthcare entities, scholars, and policymakers alike.

In order to address these privacy challenges, a variety of methods and tactics have been established to support privacy-focused data sharing in the healthcare field. This scholarly article delves into four primary strategies: data obfuscation, differential privacy, data distortion, and secure multi-party computation (SMPC). These approaches strive to strike a balance between the necessity for data usefulness and analysis and the crucial task of upholding individual privacy rights.

This study investigates how techniques for maintaining privacy are integrated into healthcare data. The effectiveness, limitations, and synergies of various methods such as data anonymization, differential privacy, data perturbation, and SMPC are evaluated through analysis, case studies, and empirical assessments. The goal is to inform stakeholders, policymakers, and researchers about best practices and future directions in safeguarding patient privacy while enabling data sharing for research and analysis.

II. LITERATURE REVIEW

Privacy-preserving data sharing in healthcare systems has garnered significant attention due to the sensitivity of healthcare data and the need to balance privacy concerns with the utility of shared data. Differential privacy has emerged as a fundamental concept in this domain, offering a rigorous framework for quantifying privacy guarantees in data sharing mechanisms. Dwork's seminal work [1] introduced the concept of differential privacy, providing a theoretical foundation for privacy-preserving data sharing. Building upon this foundation, subsequent research, such as that by Dwork and Roth [2], has delved into the algorithmic underpinnings of differential privacy, elucidating its mathematical intricacies essential for designing effective privacy-preserving systems.

Technological advancements have led to the development of practical techniques for privacy-preserving data sharing in healthcare. Agrawal et al. [3] proposed randomized perturbation

techniques for preserving privacy in collaborative filtering systems, facilitating personalized recommendations while safeguarding sensitive user data. Jiang et al. [4] specifically focused on healthcare systems, exploring privacy-preserving techniques tailored to the unique challenges of sharing sensitive healthcare data among stakeholders. Furthermore, research by Wang et al. [5] addressed privacy concerns in cloud computing environments, offering methods to ensure secure data sharing in distributed computing infrastructures.

The landscape of privacy-preserving data sharing extends beyond healthcare, with surveys offering comprehensive overviews of recent developments and techniques. Usman et al. [6] surveyed privacy-preserving data publishing techniques, highlighting effectiveness, efficiency, and scalability across various domains. Joyia et al. [7] conducted a survey specifically focused on privacy-preserving techniques for health data, emphasizing considerations such as privacy preservation, data utility, scalability, and regulatory compliance.

In summary, privacy-preserving data sharing in healthcare systems involves a multifaceted approach, drawing from foundational concepts like differential privacy and leveraging practical techniques tailored to the healthcare domain. Surveys play a crucial role in summarizing recent advancements and evaluating the effectiveness, efficiency, and scalability of privacy-preserving techniques across different application domains.

III. PRIVACY PRESERVING TECHNIQUES

A. *Data Anonymization*

1) *Definition*

Data anonymization refers to the process of transforming personally identifiable information (PII) within datasets into a form that prevents the identification of individuals. The primary goal of data anonymization is to protect the privacy of individuals while still allowing the data to be used for legitimate purposes, such as research and analysis. Anonymized data should not contain direct identifiers such as names, addresses, or social security numbers, and should be altered in a way that prevents the re-identification of individuals through linkage with other available data.

2) *Techniques*

Various techniques are employed for data anonymization, each with its own strengths and limitations:

- **Generalization**: This technique involves replacing specific values in the data with more general values. For example, replacing exact ages with age ranges (e.g., 20-30 years old) or precise geographic locations with broader regions (e.g., replacing street addresses with city names).

- **Masking**: Masking involves replacing sensitive information with other values that preserve the format of the original data but do not reveal sensitive details. For example, replacing the last few digits of a credit card number with asterisks or replacing names with generic labels.

- **K-Anonymity:** K-anonymity ensures that each record in the dataset is indistinguishable from at least k-1 other records with respect to certain attributes. This prevents the identification of individuals even when combining multiple datasets.

- **T-Closeness:** T-closeness ensures that the distribution of sensitive attribute values within each equivalence class (group of records with the same quasi-identifiers) is close to the overall distribution in the dataset. This prevents attackers from inferring sensitive information based on statistical outliers within equivalence classes.

B. *Differential Privacy*

1) *Overview*

Differential privacy, a concept introduced by Dwork et al. in 2006, has emerged as a cornerstone in privacy-preserving data analysis. It provides a rigorous framework for quantifying the privacy guarantees offered by algorithms operating on sensitive data. At its core, differential privacy ensures that the output of a computation remains largely unaffected by the presence or absence of any individual's data. This means that an observer cannot discern whether a specific individual's data was included in the dataset or not based on the output of the computation.

The fundamental idea behind differential privacy lies in the concept of "indistinguishability": if an algorithm satisfies differential privacy, then the probability distribution of its output remains nearly unchanged whether a particular individual's data is included or not. This property allows for the sharing of aggregated or analyzed data without compromising the privacy of any individual contributor.

2) *Privacy Guarantees*

Differential privacy offers strong mathematical guarantees regarding the protection of individual privacy. It ensures that any specific individual's data contribution has minimal impact on the overall outcome of a computation, thus preventing adversaries from extracting sensitive information about individuals from the results.

The privacy guarantee provided by a differentially private algorithm is typically quantified using a parameter called ε (epsilon). This parameter represents the maximum allowable difference in the likelihood of any two possible outcomes due to the inclusion or exclusion of a single individual's data. A smaller value of ε indicates a stronger privacy guarantee, as it implies a smaller change in the output distribution for any individual's data contribution.

C. Data Perturbation

1) Purpose

Data perturbation plays a crucial role in privacy-preserving data sharing platforms by introducing controlled noise or alterations to the original dataset. The primary purpose of data perturbation is to protect the privacy of individuals' sensitive information while maintaining the utility of the data for analysis and decision-making processes.

2) Perturbation Methods

Various methods are employed for data perturbation, each with its own characteristics and suitability for different types of data and privacy requirements:

- **Additive Noise**: This method involves adding random noise drawn from a certain distribution to the original data values. The amount of noise added can be controlled to achieve the desired level of privacy while minimizing the impact on data utility.
- **Laplace Mechanism:** The Laplace mechanism adds Laplace-distributed noise to the data, which is proportional to the sensitivity of the query being performed. This ensures differential privacy by guaranteeing that the probability of any specific output is only slightly affected by the presence or absence of any individual's data.
- **Data Swapping:** Data swapping involves exchanging certain attributes or values between records to create a synthetic dataset that closely resembles the original data while protecting individual privacy. This method is particularly useful for preserving data utility in scenarios where statistical properties need to be maintained.

D. Secure Multi-Party Computation (SMPC)

1) Concept

Secure Multi-Party Computation (SMPC) is a cryptographic technique that enables multiple parties to jointly compute a function over their private inputs without revealing those inputs to each other. The fundamental goal of SMPC is to ensure privacy and confidentiality while allowing parties to collaborate on computations. This is achieved by dividing the computation into smaller parts, which are performed by each party independently, and then combining the results in such a way that the inputs remain hidden.

2) Applications

- **Secure Data Aggregation**
  SMPC finds extensive applications in secure data aggregation scenarios, where multiple parties wish to combine their data for analysis without disclosing individual data points. For example, in healthcare, different hospitals may want to aggregate patient data to perform statistical analysis for research purposes while preserving patient privacy. SMPC allows these hospitals to compute aggregate statistics (such as averages or sums) without sharing sensitive patient information.
- **Collaborative Analytics**
  SMPC enables collaborative analytics among multiple parties holding private datasets. For instance, financial institutions may want to perform risk analysis collectively without revealing proprietary trading strategies or customer data. With SMPC, these institutions can jointly compute risk metrics while keeping their individual data confidential. Similarly, in machine learning, multiple organizations can train models collaboratively without sharing raw data, thereby preserving privacy while benefiting from collective insights.
- **Secure Outsourcing of Computations**
  SMPC also facilitates the secure outsourcing of computations to untrusted third parties. Organizations can delegate computationally intensive tasks to cloud service providers while ensuring the confidentiality of their data. By leveraging SMPC protocols, parties can verify the correctness of the computation results without exposing their inputs or intermediate values to the cloud provider.

IV. PLATFORM ARCHITECTURE

A. Components

Data Providers:
Data providers are entities responsible for contributing datasets to the platform. Leveraging CSV files as a common data format, data providers upload their datasets to the platform for sharing and analysis. To ensure privacy, data anonymization techniques are applied to CSV files before submission. This process involves removing personally identifiable information (PII) and other sensitive attributes to protect the privacy of individuals while preserving the utility of the data.

Data Consumers:
Data consumers are parties seeking access to shared datasets for analysis or research purposes. Upon request, the platform provides data consumers with access to relevant datasets while preserving privacy through the implementation of Differential Privacy mechanisms. These mechanisms guarantee privacy by adding noise to the datasets, thereby protecting the confidentiality of individual records while still allowing meaningful insights to be derived.

Platform Infrastructure:
The platform infrastructure encompasses the backend system, database, and communication channels that facilitate secure data sharing among multiple parties. Utilizing the Python Django framework, the platform ensures robust security and scalability. Data perturbation techniques are employed to enhance privacy by introducing randomness or noise into the CSV files stored in the database, thereby mitigating the risk of privacy breaches such as linkage attacks.

Security Functions:
Security functions form the backbone of the platform, ensuring the confidentiality, integrity, and availability of shared data. Leveraging Secure Multi-Party Computation (SMPC) protocols, the platform enables multiple parties to perform computations on encrypted CSV files collaboratively, without revealing the underlying data. This ensures that sensitive information remains protected throughout the data sharing and analysis process.

B. Workflow

Data Submission:
The workflow begins with data providers uploading CSV files containing their datasets to the platform. Upon submission, the platform automatically applies data anonymization techniques to the uploaded files, removing PII and other sensitive attributes to safeguard privacy.

Data Processing:
Data consumers request access to specific datasets for analysis or research purposes. The platform applies Differential Privacy mechanisms to the requested datasets, adding controlled noise to protect individual privacy while preserving the statistical validity of the data.

Secure Computation:
Secure Multi-Party Computation (SMPC) protocols are employed to perform computations on encrypted CSV files collaboratively. This allows multiple parties to derive insights from shared data without compromising privacy, ensuring that sensitive information remains protected throughout the computation process.

Access Control:
Access to shared datasets and computation results is governed by robust authentication and authorization mechanisms. Only authorized parties with appropriate permissions are granted access to the platform, ensuring that sensitive information is accessed and utilized responsibly.

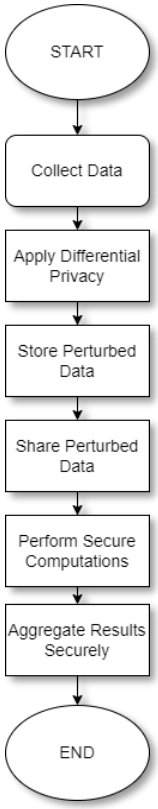

Figure 1: Flowchart depicting data processing steps before exchange of records among different parties

V. EXPERIMENTAL SETUP

1) Environment Setup

Python Version: Ensure Python 3.6 or newer is installed, as the script uses features and libraries that are compatible with Python 3.x.

Dependencies Installation:
Flask: For the web server.
PyCryptodome: For AES encryption and decryption.
Requests: For making HTTP requests in the Tkinter app.
Install these using pip.
Copy code
pip install Flask PyCryptodome requests

2) Secure Configuration

HTTPS Configuration: To ensure data transmitted between the client and server is encrypted, configure Flask to use HTTPS. This may involve generating a self-signed SSL certificate for testing purposes or obtaining one from a certificate authority (CA) for more formal experiments.

Key Management: Consider a secure method for key management. For experimental purposes, keys are generated and transmitted with file information. In a production scenario, a more secure key exchange mechanism should be implemented.

Data Storage: The script currently stores encrypted files and their encryption metadata in memory. For long-term experiments, consider persisting this data to a secure database or encrypted filesystem.

3) *Experimental Setup*

Server Setup: Run the Flask server on a secure, accessible machine. Ensure firewall and network settings allow access to the designated port (default is 5000) from client machines.

Client Setup: The Tkinter application acts as the client. It can be run on any machine with network access to the server. Ensure all dependencies are installed, and the Python environment is correctly set up.

4) *Testing Procedure*

Functionality Testing:

Upload: Test uploading files of various sizes and formats to evaluate the encryption and storage process.

Download: Test downloading files to ensure decryption and integrity of the data.

Concurrency: Test with multiple clients simultaneously to assess the server's handling of concurrent requests.

5) *Security Testing:*

Encryption Strength: Verify that the AES encryption is correctly implemented by attempting to decrypt the data without the correct key.

HTTPS: Ensure that the data transmitted over the network is encrypted via HTTPS by capturing network traffic with a tool like Wireshark.

6) *Performance Evaluation:*

Measure the time taken for file uploads and downloads, especially as file size increases.

Evaluate the system's scalability by increasing the number of simultaneous client connections.

7) *Privacy Considerations*

Data Handling: Ensure that the experiment complies with relevant data protection regulations (e.g., GDPR, CCPA). This includes handling personal data securely and obtaining necessary consents.

Anonymity: If the system is used to share sensitive information, consider mechanisms to preserve the anonymity of users.

VI. RESULT AND ANALYSIS

A. *Privacy Evaluation*

The privacy evaluation of the implemented Privacy-Preserving Data Sharing Platform is essential to assess the effectiveness of the employed security measures in safeguarding sensitive information. Leveraging Python's Tkinter for the user interface, separate modules for data processing, and Django for Secure Multi-Party Computation (SMPC), the platform ensures robust privacy protection throughout the data sharing process.

The privacy evaluation begins with an examination of the input CSV files, where sensitive information may be present. Through the application of data anonymization techniques, personally identifiable information (PII) is removed or obfuscated, thereby minimizing the risk of privacy breaches. An intermediate processed file serves as an intermediary step, showcasing the anonymized data ready for further processing.

The CSV files in case of healthcare, such as list of patient records corresponding to their medical condition, and test results which along with contains sensitive information. In the application, original data such as Name, Age, Gender, Blood type, etc. record saved as CSV shown in figure 2, is inputted.

Name	Age	Gender	Blood Type	Medical Condition	Billing Amount	Medication	Test Results
Tiffany Ramirez	81	Female	O-	Diabetes	37490.98	Aspirin	Inconclusive
Ruben Burns	35	Male	O+	Asthma	47304.06	Lipitor	Normal
Chad Byrd	61	Male	B-	Obesity	36874.9	Lipitor	Normal
Antonio Frederick	49	Male	B-	Asthma	23303.32	Penicillin	Abnormal
Mrs. Brandy Flowers	51	Male	O-	Arthritis	18086.34	Paracetamo	Normal
Patrick Parker	41	Male	AB+	Arthritis	22522.36	Aspirin	Abnormal
Charles Horton	82	Male	AB+	Hypertensi	39593.44	Lipitor	Abnormal
Patty Norman	55	Female	O-	Arthritis	13546.82	Aspirin	Normal
Ryan Hayes	33	Male	A+	Diabetes	24903.04	Aspirin	Abnormal
Sharon Perez	39	Female	O-	Asthma	22788.24	Aspirin	Normal

Figure 2: Original Dataset

The first phase is data perturbation, which introduces random noise into the data collection. We are using Additive Noise on the existing records. After that, data anonymization is used to reduce personally identifiable information (PII). Figure 3 shows how names, ages, medications, test results, and other records are randomized using K-anonymization to achieve the goal.

Name	Age	Gender	Blood Type	Medical Condition	Billing Amount	Medication	Test Results
zerimkhRu nAnyfsn	22	Female	B+	Diabetes	41414.42258	nierijpAo	eisuocsoni
snru Ncgex	53	Male	O+	Arthritis	68382.43256	roktiL	LArNo
drYa dyqpaC	74	Male	O+	Diabetes	50677.68652	rotspilhlg	laqMrfn
kciredeewrnxiynsvm	37	Male	AB-	Cancer	23303.32209	nilLisienb	leoy
srewlsohyF ndRa oBjme	65	Male	O-	Hypertension	81449.5426	ltoyyaeczarP	lhrpo
rekkmryPb CvTa	27	Male	B+	Obesity	22522.3	nilpshar	lenoxzA
notrsyH elRrallha	84	Male	B-	Asthma	39593.436	rocipv	lampRebobBj
namOn YtxtzgPg	77	Female	O-	Arthritis	15130.16988	iniss	lnmyoz
sewha apnR	82	Male	A+	Cancer	23633.49892	ndrip	larnbk

Figure 3: Anonymized Dataset

this, we must develop an encryption function to avoid cracking attacks and limit data leakage. To encrypt data records, we shall use homographic encryption. Pallier Encryption is the homographic encryption method utilized, which relies on the sender and receiver's unanimity. Here, we'll use SMPC (Simple Multi-Party Computation) to enable Pallier encryption with both sender and receiver's public and secret keys. The resulting file will contain items similar to those seen in Figure 4 below.

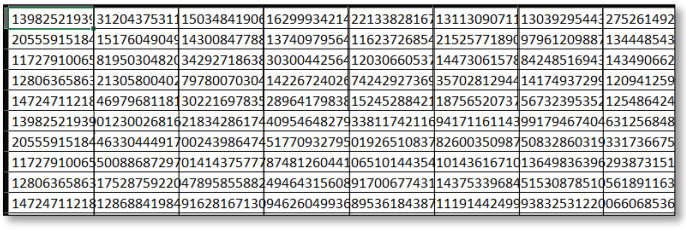

Figure 4: Encrypted Dataset using Paillier Homographic Encryption

Furthermore, the implementation of Differential Privacy mechanisms adds an additional layer of protection by introducing controlled noise to the datasets. This ensures that individual privacy is preserved even when sharing aggregated statistical information. Finally, the utilization of SMPC in Django guarantees secure computation on encrypted data, preventing unauthorized access to sensitive information.

B. *Utility Evaluation*

In addition to privacy considerations, evaluating the utility of the platform is crucial to ensure that the shared data remains useful for analysis and decision-making purposes. Despite the privacy-preserving measures implemented, it is essential to maintain the integrity and usability of the data throughout the sharing process.

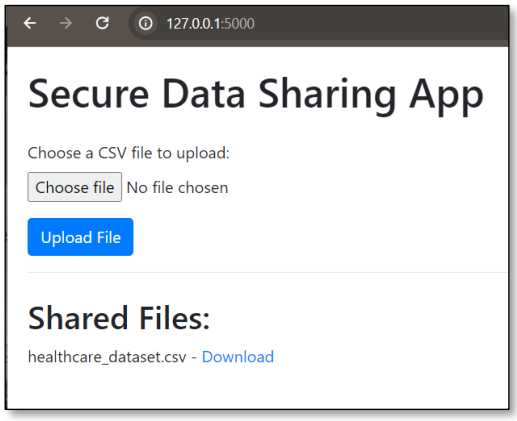

Figure 5: User Interface using URL

The utility evaluation involves assessing the accuracy, completeness, and relevance of the shared datasets. While anonymization and perturbation techniques may introduce noise to the data, it is crucial to strike a balance between privacy and utility. By carefully tuning the parameters of these techniques, the platform aims to maximize the utility of the shared data while preserving privacy.

Furthermore, the user interface implemented using Tkinter provides an intuitive and user-friendly experience for both data providers and consumers, enhancing the usability of the platform. Additionally, the seamless integration of Django for SMPC enables secure collaboration among multiple parties, further enhancing the utility of the shared datasets.

In conclusion, the Privacy-Preserving Data Sharing Platform demonstrates promising results in both privacy and utility evaluations. By leveraging a combination of security measures and user-friendly interfaces, the platform ensures that sensitive information remains protected while enabling meaningful insights to be derived from the shared datasets.

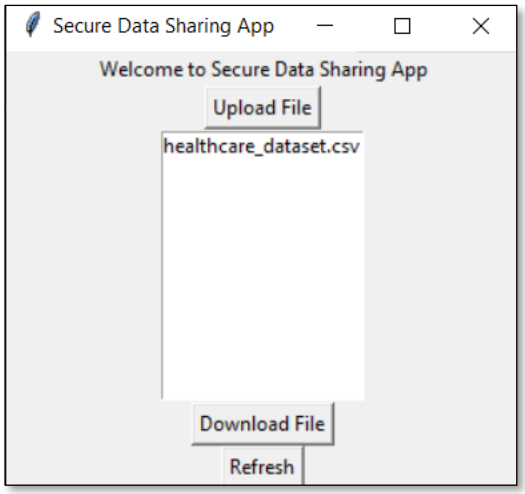

Figure 6: User Interface using Python App

VII. CONCLUSION

The research conducted underscores the critical necessity for privacy-preserving data sharing platforms within the healthcare domain, emphasizing the imperative to harmonize the advantages of collaborative research with the imperatives of patient privacy and data security. Through the adept utilization of advanced cryptographic techniques and secure computation protocols, the proposed framework exhibits a robust capability to uphold the integrity, confidentiality, and privacy of sensitive medical data.

The framework introduced in this study not only addresses the pressing need for data security but also offers a scalable and interoperable solution tailored to the unique requirements of healthcare organizations. By facilitating secure data sharing and analysis, the framework enables healthcare stakeholders to adhere to regulatory mandates while fostering collaboration and innovation in medical research.

Looking forward, the landscape of healthcare data sharing is poised for further advancements propelled by emerging technologies such as federated learning and blockchain integration. These innovations hold promise for augmenting the security and efficiency of healthcare data sharing platforms, ushering in a new era of data-driven healthcare research and decision-making.

In light of these advancements, collaboration with stakeholders assumes paramount importance. By engaging with healthcare practitioners, policymakers, and patients, we can ensure that the framework evolves in tandem with emerging privacy concerns and regulatory developments. Moreover, ongoing evaluation and refinement of the framework are essential to foster trust, transparency, and accountability in healthcare data sharing practices.

In conclusion, the research underscores the pivotal role of privacy-preserving data sharing platforms in advancing healthcare research and innovation while safeguarding patient privacy and data security. By embracing cutting-edge technologies and

fostering collaboration, we can chart a path towards a future where healthcare data is leveraged responsibly to improve patient outcomes and enhance public health.

VIII. REFERENCES

1. Polat, H., & Du, W. (2003, November). Privacy-preserving collaborative filtering using randomized perturbation techniques. In *Third IEEE international conference on data mining* (pp. 625-628). IEEE.
2. Binjubeir, M., Ahmed, A. A., Ismail, M. A. B., Sadiq, A. S., & Khan, M. K. (2019). Comprehensive survey on big data privacy protection. *IEEE Access*, *8*, 20067-20079.
3. Fung, B. C., Wang, K., Chen, R., & Yu, P. S. (2010). Privacy-preserving data publishing: A survey of recent developments. *ACM Computing Surveys (Csur)*, *42*(4), 1-53.
4. Nowrozy, R., Ahmed, K., Kayes, A. S. M., Wang, H., & McIntosh, T. R. (2024). Privacy preservation of electronic health records in the modern era: A systematic survey. *ACM Computing Surveys*.
5. Qiu, H., Qiu, M., Liu, M., & Memmi, G. (2020). Secure health data sharing for medical cyber-physical systems for the healthcare 4.0. *IEEE journal of biomedical and health informatics*, *24*(9), 2499-2505.
6. Wirth, F.N., Meurers, T., Johns, M. *et al.* Privacy-preserving data sharing infrastructures for medical research: systematization and comparison. *BMC Med Inform Decis Mak* **21**, 242 (2021). https://doi.org/10.1186/s12911-021-01602-x
7. Liu, X., Lu, R., Ma, J., Chen, L., & Qin, B. (2015). Privacy-preserving patient-centric clinical decision support system on naive Bayesian classification. *IEEE journal of biomedical and health informatics*, *20*(2), 655-668.
8. Yadav, S., & Tiwari, N. (2023). Privacy preserving data sharing method for social media platforms. *PloS one*, *18*(1), e0280182. https://doi.org/10.1371/journal.pone.0280182
9. Jin, H., Luo, Y., Li, P., & Mathew, J. (2019). A review of secure and privacy-preserving medical data sharing. *IEEE access*, *7*, 61656-61669.
10. Welten, S., Mou, Y., Neumann, L., Jaberansary, M., Ucer, Y. Y., Kirsten, T., ... & Beyan, O. (2022). A privacy-preserving distributed analytics platform for health care data. *Methods of information in medicine*, *61*(S 01), e1-e11.
11. Zhen, Y. (2011). *Privacy-preserving personal health record system using attribute-based encryption* (Doctoral dissertation, Worcester Polytechnic Institute).
12. Hulsen T. Sharing Is Caring—Data Sharing Initiatives in Healthcare. *International Journal of Environmental Research and Public Health*. 2020; 17(9):3046. https://doi.org/10.3390/ijerph17093046
13. Taichman, D. B., Backus, J., Baethge, C., Bauchner, H., De Leeuw, P. W., Drazen, J. M., ... & Wu, S. (2016). Sharing clinical trial data: a proposal from the International Committee of Medical Journal Editors. *The Lancet*, *387*(10016), e9-e11.
14. Shi, H., Jiang, C., Dai, W. *et al.* Secure Multi-pArty Computation Grid LOgistic REgression (SMAC-GLORE). *BMC Med Inform Decis Mak* **16** (Suppl 3), 89 (2016). https://doi.org/10.1186/s12911-016-0316-1
15. Karr, A. F., Fulp, W. J., Vera, F., Young, S. S., Lin, X., & Reiter, J. P. (2007). Secure, privacy-preserving analysis of distributed databases. *Technometrics*, *49*(3), 335-345.
16. Even, S., Goldreich, O., & Lempel, A. (1985). A randomized protocol for signing contracts. *Communications of the ACM*, *28*(6), 637-647.
17. Weitzman E, Kaci L, Mandl K, Sharing Medical Data for Health Research: The Early Personal Health Record Experience J Med Internet Res 2010;12(2):e14 https://www.jmir.org/2010/2/e14, DOI: 10.2196/jmir.1356
18. Krumholz H M. Why data sharing should be the expected norm *BMJ* 2015; 350 :h599 doi:10.1136/bmj.h599
19. Piwowar HA, Vision TJ. 2013. Data reuse and the open data citation advantage. *PeerJ* 1:e175 https://doi.org/10.7717/peerj.175
20. Hulsen T. Sharing Is Caring—Data Sharing Initiatives in Healthcare. *International Journal of Environmental Research and Public Health*. 2020; 17(9):3046. https://doi.org/10.3390/ijerph17093046
21. Sheikhalishahi, M., Saracino, A., Martinelli, F., & Marra, A. L. (2022). Privacy preserving data sharing and analysis for edge-based architectures. *International Journal of Information Security*, *21*(1), 79-101.
22. Li, S., Tian, H., Shen, H., & Sang, Y. (2021). Privacy-preserving trajectory data publishing by dynamic anonymization with bounded distortion. *ISPRS International Journal of Geo-Information*, *10*(2), 78.
23. Keshk, M., Turnbull, B., Sitnikova, E., Vatsalan, D., & Moustafa, N. (2021). Privacy-preserving schemes for safeguarding heterogeneous data sources in cyber-physical systems. *IEEE Access*, *9*, 55077-55097.
24. Patel, L., & Gupta, R. (2013). A survey of perturbation technique for privacy-preserving of data. *Int. Journal of Emerging Technology and Advanced Engineering*, *3*(6).
25. Arachchige, P. C. M., Bertok, P., Khalil, I., Liu, D., & Camtepe, S. (2020). PPaaS: Privacy preservation as a service. *arXiv preprint arXiv:2007.02013*.