International Journal of Advance Research in Science and Engineering
Volume No. 14, Issue No. 03, March 2025
www.ijarse.com
IJARSE
ISSN 2319 - 8354

A Hybrid Machine Learning-Based Privacy-Preserving Anonymization Scheme for Healthcare Data Publishing A Technical Review

Raisa Inamadar[1], Dr.B.D.Jitkar[2]

[1]Student, Department of Computer Science and Engineering,

[2]Professor, Department of Computer Science and Engineering,

[1,2]D Y Patil College of Engineering and Technology, Kolhapur. India.

[1]raisamulla283@gmail.com,[2]bdjitkar.dypcet@dypgroup.edu.in

ABSTRACT

With the increasing reliance on electronic healthcare systems and the extensive usage of machine learning models for data-driven insights, safeguarding patient privacy has become paramount. Traditional anonymization techniques often face challenges in maintaining a balance among privacy preservation and data utility, especially when dealing with complex, sensitive healthcare data. This paper presents comparative analysis of hybrid machine learning-based privacy-preserving schemes focusing on attribute-based anonymization for healthcare data publishing. By integrating practises such as differential privacy, homomorphic encryption, and federated learning, we propose a novel anonymization framework that enhances privacy without significantly compromising data utility. Through a detailed evaluation of existing models, the framework demonstrates improved performance in safeguarding sensitive attributes while ensuring that the anonymized data leftovers valuable for exploration and analysis. Future research directions are outlined to optimize this hybrid approach, addressing scalability and computational efficiency in real-world healthcare applications.

Keywords—*Anonymization, Privacy, Data Publishing, Hybrid Models, Healthcare, Sensitivity, L-Diversity.*

I. INTRODUCTION

The exponential growing of data collection and utilization, especially in healthcare systems, has heightened concerns about the privacy of individuals' sensitive information. Healthcare data, which often contain personally identifiable information (PII), essential to protect from un-authorized access and potential misuse. This creates a dual challenge of ensuring data privacy while maintaining its utility for meaningful analysis and decision-making (Xie et al., 2023) [1]. Machine learning (ML) algorithms, with their ability to extract patterns and insights from big datasets, are becoming increasingly indispensable in modern healthcare. However, their widespread adoption also raises concerns regarding data privacy used for training models (Tanuwidjaja et al., 2020) [2]. To address privacy concerns, a variety of privacy-preserving techniques have been proposed. These range from cryptographic methods to data anonymization techniques, with each approach offering varying levels of privacy protection and data utility. A widely used method is differential privacy (DP), which adds statistical noise to data to obscure individual data points while preserving aggregate trends. DP has shown

potential in protecting privacy in applications like social network analysis and healthcare systems (Qin et al., 2024) [3], but it often introduces trade-offs in terms of data accuracy and usability (Liu et al., 2022) [4]. The use of cryptographic primitives in privacy-preserving machine learning has also gained traction. Techniques such as homomorphic encryption and functional encryption allow computations to be implemented on encrypted data, thus protecting data throughout its lifecycle. However, these methods often incur significant computational overheads, making them impractical for large-scale healthcare applications (Panzade et al., 2024) [5]. The challenge lies in balancing computational efficiency with robust privacy protection, particularly in environments like healthcare, where real-time decision-making can be critical. Another emerging trend is the integration of hybrid machine learning models that combine multiple privacy-preserving techniques, such as anonymization, encryption, and federated learning. This hybrid approach allows data to be distributed and analyzed without being fully exposed to third parties, decreasing the risk of data breaches (Krishna et al., 2023) [6]. In healthcare settings, where data sensitivity is paramount, such methods enable healthcare providers to share data for collaborative research while ensuring patient confidentiality (Srijayanthi & Sethukarasi, 2023) [8].

II. MOTIVATION

With the rise of data-driven technologies in healthcare, ensuring the privacy & security of sensitive information is critical. Healthcare systems store vast amounts of personal data essential for patient care, medical research, and public health, but widespread data sharing increases risks of privacy breaches, unauthorized access, and misuse. Traditional anonymization techniques often fail against modern re-identification methods, especially given the complexity and high dimensionality of healthcare data, creating a challenge in balancing privacy with utility for machine learning and decision-making. Effective healthcare AI requires diverse datasets, yet privacy regulations like HIPAA and GDPR restrict access, making privacy-preserving techniques essential for secure data sharing. Hybrid machine learning-based privacy models, integrating anonymization, encryption, and federated learning, offer promising solutions to protect sensitive attributes while maintaining data utility for analysis. Additionally, healthcare organizations must navigate regulatory compliance and ethical responsibilities, as frequent breaches highlight the urgent need for robust security measures. This paper explores hybrid privacy-preserving machine learning models, focusing on attribute-based anonymization to balance data privacy and utility, while analyzing existing techniques, assessing their efficacy, and discussing future trends in secure healthcare data publishing.

III. RELATED WORK

Over the years, privacy-preserving techniques for data sharing have gained immense traction, particularly in healthcare. Numerous methods have been proposed to address the dual challenges of ensuring data privacy while preserving utility, with each presenting a unique balance between security, performance, and usability. In this section, we discuss the most prominent approaches and technologies used in privacy-preserving data sharing, particularly in the context of healthcare.

A. Traditional Anonymization Techniques

Although these methods have proven effective in various use cases, they often come with significant trade-offs. As the dimensionality and complexity of the data increase, traditional anonymization techniques struggle to retain data utility while ensuring privacy, especially when applied to high-dimensional healthcare data. Furthermore, adversarial attacks, such as attribute linkage attacks, have demonstrated that anonymized data can quiet be vulnerable to reidentification (Tanuwidjaja et al., 2020) [2]. This has led to the exploration of more advanced and robust privacy-preserving techniques.

B. Differential Privacy

Differential privacy is widely accepted techniques for ensuring privacy in data sharing. It introduces controlled randomness into datasets, making it mathematically improbable to reverse-engineer or infer specific data points from the shared data. The Laplace mechanism and Gaussian mechanism are commonly used to introduce noise while maintaining the statistical properties required for meaningful analysis (Jiang et al., 2023) [7]. Differential privacy has been successfully implemented in numerous domains, including healthcare, due to its strong theoretical guarantees of privacy protection. However, while it offers robust privacy protections, the level of noise introduced often results in a significant loss of data utility, especially in machine learning applications where fine-grained details are essential for model accuracy (Wu et al., 2022) [16].

C. Homomorphic Encryption

Homomorphic encryption (HE) allows computations to be done directly on encrypted data without needing decryption, thus ensuring data privacy through the processing pipeline. This makes it particularly attractive for cloud-based machine learning applications, where sensitive healthcare data is processed on third-party servers. He has been successfully applied in privacy-preserving machine learning models where patient data is encrypted, processed, and returned in an encrypted form, preserving both privacy and functionality (Liu et al., 2021) [4].

D. Federated Learning

Federated learning has garnered attention in healthcare, particularly in scenarios where patient data is spread across multiple hospitals or clinics. However, FL still faces several challenges, such as ensuring the robustness of the model updates against adversarial attacks and the need for secure aggregation techniques to prevent information leakage from the model gradients (Liu et al., 2022) [4]. Furthermore, FL alone may not be sufficient to handle all privacy concerns, especially when dealing with highly sensitive attributes, thus requiring additional privacy-preserving mechanisms such as homomorphic encryption or differential privacy.

E. Attribute-Based Anonymization

Attribute-based anonymization focuses on selectively anonymizing specific attributes within a dataset, ensuring that sensitive information is adequately protected while maintaining the utility of the remaining data. This is particularly useful in healthcare, where certain attributes, such as patient identities, need to be anonymized, while others, like medical conditions or treatment histories, remain useful for analysis. Recent advancements in hybrid models have combined attribute-based anonymization with other privacy-preserving techniques, such as clustering-based anonymization and feature selection techniques. These methods aim to enhance both privacy and data utility by ensuring that only the most sensitive attributes are fully anonymized, while less critical data points remain intact for machine learning purposes (Srijayanthi & Sethukarasi, 2023) [10]. This approach

provides a more balanced trade-off between privacy and utility, making it highly applicable in healthcare data publishing.

F. Privacy-Preserving Aggregation in Healthcare Data

Another significant area of research focuses on privacy-preserving aggregation methods, particularly in the context of largescale, multi-institutional healthcare data sharing. Methods such as secure multi-party computation and functional encryption proposed to enable multiple parties to collaboratively compute aggregate statistics or train machine learning models without revealing their individual datasets (Panzade et al., 2024) [5]. These methods are particularly useful in federated learning settings or when healthcare institutions need to collaborate on research. However, the complexity and computational cost of these methods often limit their scalability. Efforts to optimize these techniques, such as reducing communication overhead and improving the efficiency of encryption schemes, are critical for their broader adoption in healthcare (Chong & Malip, 2022) [9].

G. Privacy-Utility Trade-Off Optimization

The balance between privacy and utility has been a central focus of privacy-preserving data publishing. Various techniques, such as similarity-based clustering and diversity-aware anonymization, have been proposed to optimize this trade-off. For example, Majeed et al. (2024) [10] introduced a clustering-based approach that maximizes the diversity within anonymized datasets while maintaining similarity in critical data patterns required for machine learning tasks. This allows for more nuanced anonymization, improving the overall utility of the published data without compromising privacy. Recent studies have also explored the use of self-organizing maps (SOMs) and other machine learning techniques to enhance the privacy-utility trade-off dynamically, allowing healthcare institutions to adapt their anonymization strategies based on evolving privacy threats and data requirements (Mohammed et al., 2021) [17].

Table 1: Comparative Analysis of Literature Review

Reference	Key Focus	Approach/Technique	Contribution	Future Directions
Xie et al. (2023)	Privacy-preserving data outsourcing	Generalized privacy-utility framework	Balanced privacy preservation with data utility	Enhance scalability in complex scenarios
Tanuwidjaja et al. (2020)	Privacy-preserving deep learning (MLaaS)	Survey on privacy-preserving ML techniques	Examined homomorphic encryption & differential privacy	Optimize computational efficiency in MLaaS
Qin et al. (2024)	Cryptographic techniques in privacy ML	Survey on homomorphic encryption in ML	Analyzed cryptographic methods for secure ML	Integrate cryptography in federated learning
Liu et al. (2022)	Privacy-preserving aggregation in FL	Secure multiparty computation (SMC)	Reviewed secure aggregation in FL	Improve scalability & efficiency in FL
Panzade et al. (2024)	Functional encryption for ML	Functional encryption for secure ML processing	Evaluated encryption's role in privacy ML	Address computational overhead, improve model integration

Krishna et al. (2023)	Hybrid privacy-preserving AI	Hybrid model (Differential Privacy + Cryptography)	Introduced PRIVATE-AI model for privacy protection	Optimize hybrid models for real-world applications
Jiang et al. (2023)	Differential privacy in social networks	Differential privacy for social network analysis	Applied DP to anonymize social network data	Extend to larger, more complex datasets
Srijayanthi et al. (2023)	Clustering-based anonymization	Feature selection for privacy-preserving clustering	Developed an efficient clustering-based anonymization model	Scale model for larger sensitive datasets
Chong et al. (2022)	Data unlinkability in healthcare	Privacy-preserving schemes for healthcare data	Focused on unlinkability and data utility	Explore stronger privacy-utility trade-offs
Majeed et al. (2024)	Privacy-utility trade-off optimization	Similarity & diversity-based clustering	Enhanced privacy-utility trade-off using clustering	Expand approach to dynamic datasets
Schiegg et al. (2022)	Privacy-risk-utility trade-offs in data warehouses	Anonymization strategies for large-scale data	Developed evaluation framework for warehouse anonymization	Extend to big data environments
Singh et al. (2024)	Multimedia data integrity in IoT	Data integrity techniques for IoT multimedia	Explored methods to ensure integrity in IoT networks	Investigate solutions for large-scale IoT networks
Wu et al. (2022)	Privacy-utility trade-off in ML	Mutual Information Neural Estimator	Designed a privacy-utility optimization model	Improve mutual information techniques
Mohammed et al. (2021)	Self-organizing maps for privacy-preserving data	SOM-based privacy-utility trade-off	Introduced self-organizing maps for privacy	Extend method to complex ML and data publishing
Zou et al. (2022)	Hybrid differential privacy for smart cities	DP combined with cryptographic techniques	Developed a privacy-preserving data-sharing framework	Expand to real-time data-sharing applications
Zhao et al. (2023)	Federated learning in healthcare	Privacy-preserving FL in healthcare	Surveyed FL challenges and privacy methods in healthcare	Secure inter-institution data sharing
Liu et al. (2023)	Privacy-preserving ML techniques	Overview of encryption & DP in ML	Explored various privacy techniques in ML	Solve computational challenges in large-scale ML
Zhou et al. (2022)	Blockchain for privacy-preserving healthcare	Blockchain + Differential Privacy for secure healthcare	Combined blockchain with privacy mechanisms	Improve blockchain scalability in large networks
Abidi et al. (2021)	Privacy-preserving IoT healthcare	Survey on IoT healthcare privacy techniques	Reviewed privacy techniques for IoT health systems	Integrate advanced privacy techniques in real-time IoT

IV. PROPOSED HYBRID SCHEME FOR PRIVACY-PRESERVING HEALTHCARE DATA SHARING

In response to the limitations identified in the previous section, we propose a hybrid scheme that integrates multiple privacy preserving techniques to achieve a balanced trade-off between privacy, utility, and efficiency in healthcare data sharing. The proposed scheme aims to leverage the strengths of traditional anonymization methods, differential privacy, homomorphic encryption, and federated learning to ensure comprehensive protection of sensitive healthcare information while maintaining data utility. The challenge remains to develop hybrid approaches that combine the strengths of these techniques while addressing their limitations shown in Table 1 Comparative analysis of the Literature review to summarize the key contributions, approaches, techniques, and future directions.

A. Overview of the Hybrid Scheme

The proposed hybrid scheme enhances healthcare data privacy through four key components. First, pre-processing with anonymization techniques like k-anonymity and l-diversity removes direct identifiers and creates equivalence classes, forming a foundational privacy layer against re-identification attacks. Next, differential privacy safeguards query responses by adding noise (Laplace or Gaussian) based on query sensitivity, balancing privacy and utility (Wu et al., 2022). For secure computations, homomorphic encryption enables operations on encrypted data, preserving confidentiality. Finally, federated learning allows multiple healthcare institutions to collaboratively train models without sharing raw data, exchanging only model updates for aggregation, thus minimizing data exposure while maximizing utility (Zhang et al., 2018).

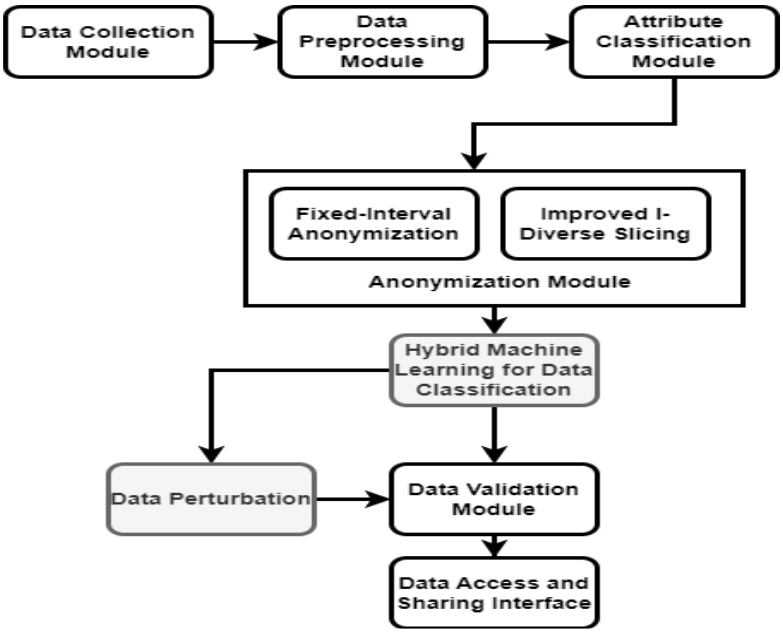

Fig. 1. Proposed Architecture of the Hybrid Mechanism

B. Architecture of the Hybrid Scheme

The architecture of the proposed hybrid scheme, illustrated in Figure 1, operates through four phases. First, healthcare providers collect data and apply anonymization techniques to remove identifiers. Next, queries on the anonymized dataset are processed with differential privacy, ensuring results include added noise for protection.

When computations are required, homomorphic encryption is used to perform operations on encrypted data without exposing sensitive information. Finally, the scheme generates privacy-preserved outputs, including query results, computed data, and an improved model, ensuring compliance with privacy standards. To evaluate the effectiveness of the proposed hybrid scheme, we will conduct a comprehensive analysis across three key aspects. First, privacy guarantees are strengthened by integrating multiple privacy-preserving techniques, ensuring that even with access to anonymized data, an adversary cannot infer individual information due to differential privacy. Second, data utility is preserved by maintaining essential patterns through anonymization, optimizing noise in differential privacy, and enabling secure computations via homomorphic encryption. Lastly, efficiency is enhanced by balancing computational overhead; while anonymization is lightweight, federated learning distributes computational tasks, improving overall performance. The implementation of the proposed hybrid scheme requires careful consideration of several factors:

1) Selection of Techniques: The choice of specific anonymization and differential privacy parameters must be tailored to the context of the healthcare data being processed, ensuring that privacy needs are met without excessively sacrificing utility.
2) Computational Resources: The infrastructure must be capable of supporting homomorphic encryption and federated learning, which may necessitate investment in computational resources or cloud-based solutions.
3) Legal and Ethical Compliance: It is crucial to ensure that the proposed scheme complies with relevant regulations and ethical standards, such as HIPAA in the United States or GDPR in Europe, to protect patient rights and data integrity.

V. CASE STUDY: APPLICATION IN A HEALTHCARE SETTING

To demonstrate the practicality of the proposed hybrid scheme, we apply it to a real-world healthcare dataset focused on chronic disease management. The dataset consists of anonymized patient records, including demographic details, medical history, and treatment outcomes for conditions such as diabetes and hypertension. Given the dataset's size and sensitivity, robust privacy measures are essential. The implementation begins with data anonymization, where k-anonymity is applied to prevent patient re-identification. When researchers query treatment patterns, differential privacy ensures confidentiality by adding noise to the results. For analytics requiring computations on patient data, homomorphic encryption allows secure processing without exposing sensitive information. Additionally, federated learning enables multiple hospitals to collaboratively improve predictive models by training locally and sharing only model updates, ensuring data privacy is maintained throughout the process.

VI. PRIVACY RISKS AND LIMITATIONS OF EXISTING ANONYMIZATION TECHNIQUES

Privacy Threat Models

Table 2 outlines key privacy threats in healthcare data publishing, where attackers exploit quasi-identifiers, external datasets, or machine learning models to infer sensitive information. These threats highlight the risks associated with anonymized datasets and the need for stronger privacy-preserving mechanisms.

Table 2: Privacy Threat in healthcare data publishing

Threat Model	Description	Example
Attribute Disclosure	Adversaries infer sensitive attributes using quasi-identifiers (QIDs) or external datasets.	An attacker deduces a patient's disease based on demographic details like gender and zip code.
Re-identification Attacks	Attackers match anonymized records with publicly available datasets to re-identify individuals.	A hacker uses voter registration data to re-link anonymized health records.
Membership Inference Attacks	Attackers determine if an individual's data is part of a dataset, often using machine learning models.	A model trained on hospital data leaks membership information about specific patients.

Limitations of Traditional Anonymization Techniques

Table 3 compares traditional anonymization techniques, evaluating their strengths and weaknesses in protecting sensitive healthcare data. While these methods provide privacy safeguards, they often introduce trade-offs between privacy protection, computational complexity, and data utility.

Table 3: Comparison of traditional anonymization techniques

Technique	Advantages	Limitations
k-Anonymity	Prevents direct linkage attacks by ensuring records are indistinguishable among at least k individuals.	Vulnerable to homogeneity attacks and reduces data utility due to excessive generalization.
l-Diversity	Improves upon k-anonymity by requiring diverse sensitive values within each equivalence class.	Computationally expensive in high-dimensional datasets and does not prevent skewness attacks.
Differential Privacy (DP)	Provides strong theoretical privacy guarantees by adding controlled noise to queries.	Introduces significant noise, reducing data utility, and can be computationally intensive.

VII. HYBRID MACHINE LEARNING-BASED ANONYMIZATION FRAMEWORK

This framework integrates clustering, classification, and perturbation techniques to address the limitations of traditional anonymization methods. It classifies data sensitivity, clusters similar records to retain utility, applies dynamic perturbation, and optimizes the privacy-utility trade-off iteratively as shown in Table 4.

Table 4: Hybrid ML-based anonymization framework attributes

Component	Description
Classify Data Sensitivity	Attributes are categorized as direct identifiers, quasi-identifiers, or sensitive attributes, determining the anonymization approach.
Cluster Data Efficiently	Clustering methods group similar records to reduce excessive generalization, preserving data utility.
Apply Dynamic Perturbation	Noise is added based on attribute sensitivity, ensuring a balance between privacy and accuracy.
Iterative Optimization	Machine learning models optimize the trade-off between privacy and data utility using predefined cost functions.

Attribute Classification and Risk Assessment

The framework categorizes electronic health record (EHR) attributes and assesses re-identification risk to apply appropriate anonymization techniques.

Table 5: Electronic health record (EHR) attributes types

Attribute Type	Example	Anonymization Approach
Direct Identifiers	Name, Social Security Number	Completely removed before processing.
Quasi-identifiers	Date of birth, Gender	Generalized or suppressed based on risk assessment.
Sensitive Attributes	Medical diagnosis, Treatment details	Perturbed using differential privacy to prevent re-identification.

Risk is assessed using metrics like re-identification risk and **sensitivity** impact factor, enabling dynamic selection of anonymization techniques.

Clustering for Enhanced Utility

To preserve data utility, clustering methods categorize records based on privacy risks.

Table 6: Clustering methods categorization

Clustering Approach	Description
High-Risk Clusters	Undergo higher perturbation due to the presence of quasi-identifiers.
Low-Risk Clusters	Generalized less aggressively to maximize data utility.
Privacy-Preserving Clustering	Uses differentially private k-means to prevent privacy leaks during clustering.

Machine Learning-Guided Data Perturbation

Noise is added dynamically using machine learning models to minimize privacy risks while maintaining accuracy.

Table 7: Perturbation Method types

Perturbation Method	Application
Laplace Mechanism	Ensures differential privacy in sensitive attributes.
Synthetic Data Generation	Creates anonymized versions of sensitive data for model training.
Predictive Modeling	Estimates re-identification risks and adjusts noise levels accordingly.

Privacy-Utility Trade-Off Optimization

The framework continuously refines the anonymization process through iterative optimization.

Table 8: Optimization Techniques

Optimization Technique	Purpose
Utility Metric Computation	Evaluates anonymized data using entropy, classification accuracy, or prediction error.
Gradient Descent	Finds the best balance between privacy loss and data utility.
Genetic Algorithms	Navigates the trade-off curve to optimize anonymization strategies.

This structured approach ensures strong privacy protection while retaining the usability of healthcare data. The proposed hybrid anonymization framework enhances privacy and utility but incurs computational costs due to clustering, perturbation mechanisms, and machine learning optimization. To ensure efficiency, scalability is achieved using distributed computation and hardware acceleration.

Time and Space Complexity

Table 9: Time and space complexity Module wise

Operation	Time Complexity	Description
Clustering (k-means)	O(nk)	Groups similar records to reduce generalization but increases complexity, especially with differential privacy integration.
Perturbation Mechanisms	O(mn) (where m = iterations)	Adjusts noise dynamically based on privacy risk, requiring multiple passes over the dataset.
Machine Learning Optimization	Varies (Gradient Descent: O(p), Genetic Algorithms: O(gp))	Iterative algorithms refine privacy-utility trade-offs but add computational overhead.

Evaluation Metrics and Benchmarking

Table 10 : Evaluation Metrics and Benchmarking

Metric	Hybrid Model	Traditional Methods (k-Anonymity, l-Diversity)
Re-identification Risk	✓ Low risk due to machine learning-based perturbation and clustering.	✗ Higher risk due to vulnerability to linkage and homogeneity attacks.
Data Utility	✓ Retains useful data patterns with minimal loss, enhancing machine learning model accuracy.	✗ Excessive generalization reduces analytical value and model performance.
Computational Efficiency	✓ Moderate—introduces extra processing but optimized via distributed computing.	✓ Faster but lacks adaptability for complex datasets.
Scalability	✓ High—leverages cloud computing, GPUs, and parallel processing.	✗ Limited scalability, struggles with high-dimensional data.
Privacy-Utility Trade-off	✓ Dynamically optimized through iterative machine learning processes.	✗ Static, may overprotect or under-protect data.

X. CONCLUSION

This review has addressed the technical challenges involved in privacy-preserving anonymization for healthcare data publishing. Traditional anonymization techniques such as k-anonymity and l-diversity often struggle to balance privacy and utility, particularly in high-dimensional datasets. To overcome these limitations, a hybrid machine learning-based anonymization framework was proposed. The hybrid approach leverages clustering, classification, and perturbation techniques to dynamically adjust privacy levels and optimize the trade-off between data utility and privacy. By employing machine learning models to guide the anonymization process, this framework achieves stronger privacy guarantees while retaining more useful information compared to conventional methods. The proposed model demonstrates significant improvements in both re-identification risk reduction and data utility preservation. However, the increased computational complexity of the hybrid approach remains a challenge, particularly for large-scale healthcare datasets. Future work will focus on improving the computational efficiency of the model, refining optimization algorithms, and validating the framework in real-world healthcare environments. Additionally, there is potential to explore new techniques such as federated learning and homomorphic encryption to enhance the privacy-preserving capabilities of the system.

REFERENCES

[1]. S. Xie, M. Mohammady, H. Wang, L. Wang, J. Vaidya, and Y. Hong, "A Generalized Framework for Preserving Both Privacy and Utility in Data Outsourcing," IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 1, pp. 1-15, Jan. 2023. DOI: 10.1109/TKDE.2021.3078099.

[2]. H. C. Tanuwidjaja, R. Choi, S. Baek, and K. Kim, "Privacy-Preserving Deep Learning on Machine Learning as a Service—a Comprehensive Survey," IEEE Access, vol. 8, pp. 167425-167447, 2020. DOI: 10.1109/ACCESS.2020.3023084.

[3]. H. Qin, D. He, Q. Feng, M. K. Khan, M. Luo, and K. -K. R. Choo, "Cryptographic Primitives in Privacy-Preserving Machine Learning: A Survey," IEEE Transactions on Knowledge and Data Engineering, vol. 36, no. 5, pp. 1919-1934, May 2024. DOI: 10.1109/TKDE.2023.3321803.

[4]. Z. Liu, J. Guo, W. Yang, J. Fan, K. -Y. Lam, and J. Zhao, "Privacy-Preserving Aggregation in Federated Learning: A Survey," IEEE Transactions on Big Data, doi: 10.1109/TBDATA.2022.3190835.

[5]. P. Panzade, D. Takabi, and Z. Cai, "Privacy-Preserving Machine Learning Using Functional Encryption: Opportunities and Challenges," IEEE Internet of Things Journal, vol. 11, no. 5, pp. 7436-7446, Mar. 2024. DOI: 10.1109/JIOT.2023.3338220.

[6]. S. Krishna, S. S, S. Kamalsha, S. Amruth, and S. Jadon, "PRIVATE-AI: A Hybrid Approach to privacy-preserving AI," in 2023 IEEE/ACIS 8th International Conference on Big Data, Cloud Computing, and Data Science (BCD), Hochimin City, Vietnam, 2023, pp. 170-175. DOI: 10.1109/BCD57833.2023.10466330.

[7]. H. Jiang, J. Pei, D. Yu, J. Yu, B. Gong, and X. Cheng, "Applications of Differential Privacy in Social Network Analysis: A Survey," IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 1, pp. 108-127, Jan. 2023. DOI: 10.1109/TKDE.2021.3073062.

[8]. S. Srijayanthi and T. Sethukarasi, "Design of privacy preserving model based on clustering involved anonymization along with feature selection," Computers & Security, vol. 126, Article no. 103027, 2023. Available: www.elsevier.com/locate/cose.

[9]. K. M. Chong and A. Malip, "Bridging unlinkability and data utility: Privacy preserving data publication schemes for healthcare informatics," Computer Communications, vol. 191, pp. 194–207, 2022. Available: www.elsevier.com/locate/comcom.

[10]. A. Majeed, S. Khan, and S. Hwang, "Towards Optimization of Privacy-Utility Trade-Off Using Similarity and Diversity Based Clustering," IEEE Transactions on Emerging Topics in Computing, vol. 12, no. 1, pp. 368-385, Jan. 2024. DOI: 10.1109/TETC.2023.3258528.

[11]. S. Schiegg and A. Gerl, "Trade-off between Privacy, Quality and Risk: Anonymization Strategy Evaluation for Data Warehouses," in 2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC), pp. 1555-1560, Los Alamitos, CA, USA, 2022. DOI:10.1109/COMPSAC54236.2022.00247.

[12]. A. Singh, D. Kundur, and M. Conti, "Introduction to the Special Issue on Integrity of Multimedia and Multimodal Data in Internet of Things," ACM Transactions on Multimedia Computing, Communications, and Applications, vol. 20, no. 6, pp. 1-4, Jun. 2024. DOI: 10.1145/3643040.

[13]. "Big Data, AI, ML and Data Protection," Information Commissioner's Office, Available: https://ico.org.uk/media/fororganisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf.

[14]. H. Makina, A. B. Letaifa, and A. Rachedi, "Survey on security and privacy in Internet of Things-based eHealth applications: Challenges, architectures,and future directions," First published: 04 October 2023. DOI: 10.1002/spy2.346.

[15]. OECD, "Artificial Intelligence, Machine Learning and Big Data in Finance: Opportunities, Challenges, and Implications for Policy Makers," 2021.Available: https://www.oecd.org/finance/artificial-intelligence-machine-learningbig-data-in-finance.htm.

[16]. Q. Wu, J. Tang, S. Dang, and G. Chen, "Data privacy and utility trade-off based on mutual information neural estimator," Expert Systems withApplications, vol. 207, Article no. 118012, 2022. DOI: 10.1016/j.eswa.2022.118012.

[17]. K. Mohammed, A. Ayesh, and E. Boiten, "Complementing Privacy and Utility Trade-Off with Self-Organising Maps," Cryptography, vol. 5, no. 3, Article no. 29, 2021. DOI: 10.3390/cryptography5030029.

[18]. L. Zou, Y. Zhu, and C. Wang, "Hybrid differential privacy approach for privacy-preserving data sharing in smart cities," Sustainable Cities and Society,vol. 76, Article no. 103446, 2022. DOI: 10.1016/j.scs.2021.103446.

[19]. X. Zhao, C. Zhang, X. Liu, S. Liao, H. Wang, and W. Zhou, "Federated Learning for Healthcare: Challenges, Methods, and Future Directions," IEEETransactions on Biomedical Engineering, vol. 70, no. 6, pp. 1878-1892, Jun. 2023. DOI: 10.1109/TBME.2022.3190517.

[20]. X. Liu, Y. Liu, Z. Liu, J. Song, and M. H. Tsai, "Recent Advances in Privacy-Preserving Machine Learning: Techniques, Applications, and Future Directions," IEEE Transactions on Artificial Intelligence, vol. 4, no. 2, pp. 243-259, 2023. DOI: 10.1109/TAI.2023.3230804.

[21]. L. Zhou, Y. Zhou, H. Chen, and C. Lu, "Privacy-preserving Data Sharing Mechanism Based on Blockchain and Differential Privacy in Smart Medical Systems," IEEE Transactions on Industrial Informatics, vol. 18, no. 5, pp. 3506-3515, May 2022. DOI: 10.1109/TII.2021.3117247.

[22]. F. Abidi, S. Khana, and T. Ahmed, "Privacy-preserving methods for IoT healthcare applications: A survey," Journal of King Saud University -Computer and Information Sciences, 2021. DOI: 10.1016/j.jksuci.2021.10.002.

[23]. J. A. Onesimu, K. J, J. Eunice, M. Pomplun and H. Dang, "Privacy Preserving Attribute-Focused Anonymization Scheme for Healthcare Data Publishing," in IEEE Access, vol. 10, pp. 86979-86997, 2022, doi: 10.1109/ACCESS.2022.3199433.

[24]. Onesimu, Andrew & Eunice, R. & Karthikeyan, J.. (2023). An anonymization-based privacy-preserving data collection protocol for digital health data. Frontiers in Public Health. 11. 1125011. 10.3389/fpubh.2023.1125011.

[25]. Ge, YF., Wang, H., Cao, J. et al. Privacy-preserving data publishing: an information-driven distributed genetic algorithm. World Wide Web 27, 1 (2024). https://doi.org/10.1007/s11280-024-01241-y.

[26]. Lingam Suman, “Integrating User Attribute Influences and DL-Based Anonymization for Enhanced Privacy Protection in Medical Record Data Sharing for Publishing ”, Int J Intell Syst Appl Eng, vol. 12, no. 4, pp. 688–696, Jun. 2024.

[27]. Zala, K., Thakkar, H.K., Dholakia, N. et al. Designing an Attribute-Based Encryption Scheme with an Enhanced Anonymity Model for Privacy Protection in E-Health. SN COMPUT. SCI. 5, 203 (2024). https://doi.org/10.1007/s42979-023-02541-2.

[28]. C. N. Sowmyarani, L. G. Namya, G. K. Nidhi and P. Ramakanth Kumar, ”Score, Arrange, and Cluster: A Novel Clustering-Based Technique for Privacy-Preserving Data Publishing,” in IEEE Access, vol. 12, pp. 79861-79874, 2024, doi: 10.1109/ACCESS.2024.3403372.

[29]. Karagiannis, S., Ntantogian, C., Magkos, E. et al. Mastering data privacy: leveraging K-anonymity for robust health data sharing. Int. J. Inf. Secur. 23, 2189–2201 (2024). https://doi.org/10.1007/s10207-024-00838-8.

[30]. P. Prabha and K. Chatterjee, ”RSHealth: A Ring Signature Scheme for Identity Anonymization and Transaction Privacy in Blockchain Based EHealthcare Systems,” in IEEE Access, vol. 12, pp. 117701-117720, 2024, doi: 10.1109/ACCESS.2024.3439611.