International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 12 | Dec - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Anonymized Security Framework for Safeguarding and Distributing Clinical Data

[1]E.Srija, [2]G.S.V Mithelesh, [3]G.Manas Reddy, [4] **Md. Irfan**
[1,2,3]Student,[4]Assistant Professor
[1,2,3,4]Department of Computer Science and Engineering
[1,2,3,4]Guru Nanak Institutions Technical Campus, Hyderabad, Telangana, India

ABSTRACT :

Advancements in Industry 4.0 have significantly improved healthcare through enhanced treatments, communication, remote monitoring, and cost reduction. However, sharing sensitive healthcare data is challenging due to privacy and security concerns. This study presents an attribute-focused privacy-preserving data publishing approach combining fixed-interval methods for numerical attributes and enhanced l-diverse slicing for categorical data. Horizontal and vertical data partitioning ensures privacy without compromising utility. Experiments with real-world datasets show a 13% improvement in classification accuracy and a 12% reduction in information loss compared to existing methods, protecting against identity and attribute disclosures.

I. INTRODUCTION :

Industry 4.0 has brought a surge in digital data generation, particularly in healthcare, where sensitive information is collected for analysis and research. Electronic Health Records (EHRs) are integral to modern healthcare, aiding in clinical decision-making, diagnostics, and communication. However, these records also expose patients to privacy risks, as attackers can combine anonymized data with external knowledge to identify individuals.

Healthcare data sharing is vital for advancements in medical research, but privacy concerns and regulatory constraints make organizations hesitant to share data. Privacy-Preserving Data Publishing (PPDP) methods aim to address these challenges by safeguarding sensitive information while ensuring the data remains useful for research. Common approaches like k-anonymity, l-diversity, and t-closeness reduce data specificity to enhance privacy. However, they are still vulnerable to various disclosure attacks. Cryptographic methods, though highly secure, are computationally demanding and impractical for large datasets, creating a need for more efficient solutions.

Current systems for secure data sharing have additional drawbacks, including insufficient security during data upload and limited accuracy due to dependence on predefined keywords. The need for advanced Privacy-Preserving Data Publishing (PPDP) techniques has grown due to the challenges of securing sensitive data while preserving its analytical value. This project introduces a robust PPDP framework that ensures privacy without compromising usability, minimizing risks like identity and attribute disclosures. It fosters trust and compliance while supporting advancements in healthcare research.Industry 4.0 has transformed healthcare through IoT, wearable devices, and AI, with Electronic Health Records (EHRs) improving diagnostics and treatment. However, EHRs, containing sensitive personal data, are vulnerable to privacy threats like background knowledge attacks. Ethical and regulatory concerns often limit data sharing, despite its importance for medical research. The proposed PPDP framework addresses these issues by securing data and retaining its utility, enabling safe sharing and driving innovation in healthcare.

II. LITERATURE SURVEY :

This article highlights gaps in information privacy research and proposes a cross-disciplinary approach. It critiques Westin's privacy model and suggests a contextual perspective focusing on digital networks,

marginalized populations, and global contexts to enhance privacy theories and technology.

In breast cancer detection, the research combines segmentation and machine learning methods to improve diagnostic accuracy. An adaptive median filter enhances image quality, while a new evaluation parameter assesses K-means and GMM performance. The hybrid technique efficiently distinguishes benign from malignant tumors, enabling faster, more accurate diagnoses. Results confirm its potential for early breast cancer detection.

III. PROPOSED SYSTEM :

The proposed system presents an attribute-focused privacy-preserving data publishing (PPDP) scheme, combining a fixed-interval approach for numerical attributes and an improved l-diverse slicing approach for categorical and sensitive data. The fixed-interval method replaces original values with computed equivalents, while the l-diverse slicing partitions data horizontally and vertically to minimize privacy risks. This system offers several advantages, including reduced storage costs due to efficient data transformation and strong protection for sensitive attributes. Experiments show a 13% improvement in classification accuracy compared to traditional methods, and a 12% reduction in information loss (measured by NCP) compared to similar approaches. The system balances privacy protection and data utility, ensuring secure and effective use of healthcare data in research without compromising its quality.

METHODOLOGIES:

1. **Healthcare Data**

This module allows users to securely log in with a username and password. New users can register their details, while existing users can directly log in. The server manages the user accounts to track upload and download activities, with the username acting as the user ID.

2. **Data Controller**

After logging in, the data user can search for files by name and download encrypted data. Users can also send a trapdoor request to the server, and once the data owner grants permission, the file is downloaded in unencrypted form.

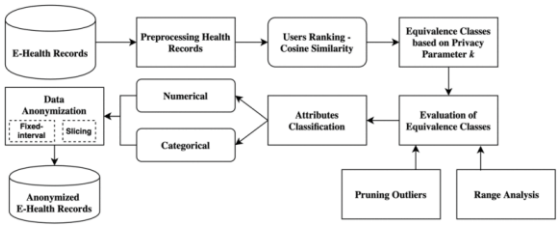
3. **Privacy Protected Data**

In this module, data owners register and upload files to the database. They can also send access requests to data users, maintaining control over the files and ensuring the privacy of the shared data.

4. Data Analyst

The cloud server logs in to view user and data owner information, access stored files, and request encryption keys. It also monitors and handles any potential security risks or attacks related to the stored files.

SYSTEM ARCHITECTURE

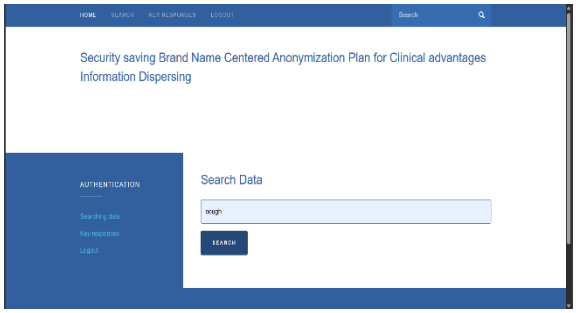


IV. FUTURE ENHANCEMENT:

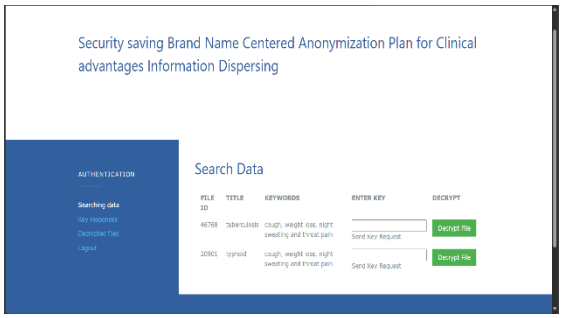
The future scope of this research involves treating quasi-attributes as both sensitive and semi-sensitive. This will allow for stronger protection of these attributes by integrating advanced techniques such as k-anonymity, l-diversity, and t-closeness. Additionally, the approach can be adapted for use in fields beyond healthcare.

V. RESULT AND IMPLEMENTATION :

INPUT IMAGE :



OUTPUT:



VI. CONCLUSION :

In conclusion, the proposed attribute-focused privacy-preserving data publishing scheme effectively balances data utility and privacy protection. By incorporating advanced techniques like fixed-interval and improved l-diverse slicing, it safeguards sensitive attributes while maintaining high data accuracy. This approach not only enhances security but also has potential applications beyond healthcare, ensuring privacy across various domains.

VI. REFERENCES :

[1] T. C. Clark and A. F. Westin, "Privacy and freedom," California Law Rev., vol. 56, no. 3, p. 911, May 1968, doi: 10.2307/3479272.

10.2307/3479272.

[2] P. F. Wu, J. Vitak, and M. T. Zimmer, "A contextual approach to information privacy research," J. Assoc. Inf. Sci. Technol., vol. 71, no. 4, pp. 485–490, Apr. 2020, doi: 10.1002/asi.24232.

[3] J. Andrew and J. Karthikeyan, "Privacy-preserving Internet of Things: Techniques and applications," Int. J. Eng. Adv. Technol., vol. 8, no. 6, pp. 3229–3234, Aug.2019,doi: .35940/ijeat.F8830.088619.

[4] X. Wang, S. Garg, H. Lin, G. Kaddoum, J. Hu, and M. S. Hossain, "PPCS: An intelligent privacy-preserving mobile-edge crowdsensing strategy for industrial IoT," IEEE Internet Things J., vol. 8, no. 13, pp. 10288–10298,Jul.2021,doi: 0.1109/JIOT.2020.3032797.

[5] P. E. Jebarani, N. Umadevi, H. Dang, and M. Pomplun, "A novel hybrid K-means and GMM machine learning model for breast cancer detection," IEEE Access,

vol. 9, pp. 146153–146162, 2021, doi: 10.1109/ACCESS.2021.3123425.

[6] A. Andrushia, K. Sagayam, H. Dang, M. Pomplun, and L. Quach, "Visualsaliency-based abnormality detection for MRI brain images—Alzheimer's disease analysis," Appl. Sci., vol. 11, no. 19, p. 9199, Oct. 2021, doi: 10.3390/app11199199.

[7] G. N. Sundar, D. Narmadha, A. A. A. Jone, K. M. Sagayam, H. Dang, and M. Pomplun, "Automated sleep stage classification in sleep apnoea using convolutional neural networks," Informat. Med. Unlocked, vol. 26, Jan. 2021, Art. no. 100724.

[8] J. A. Onesimu and J. Karthikeyan, "an efficient privacy-preserving deep learning scheme for medical image analysis," J. Inf. Technol. Manag., vol. 12, pp. 50–67, Dec. 2021, doi: 10.22059/jitm.2020.79191.

[9] J. Andrew, S. S. Mathew, and B. Mohit, "A comprehensive analysis of privacy-preserving techniques in deep learning based disease prediction systems," J. Phys. Conf., vol. 1362, no. 1, pp. 1–9, 2019, doi: 10.1088/1742-6596/1362/1/012070.

[10] J. A. Onesimu, J. Karthikeyan, and Y. Sei, "An efficient clustering-based anonymization scheme for privacy-preserving data collection in IoT based healthcare services," Peer Peer Netw. Appl., vol. 14, no. 3, pp. 1629–1649, Feb. 2021, doi: 10.1007/s12083-021-01077-7.