International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

PRIVACY CHARACTERIZATION AND QUANTIFICATION IN DATA PUBLISHING AN INTELLIGENT FRAMEWORK FOR MEASURING PRIVACY LEAKAGE IN PPDP TECHNIQUES

KATAKAM VENKATA PAVANI
Department of MCA
SKBR PG COLLEGE, AMALAPURAM, A.P
pavanikatakam2003@gmail.com

Abstract

The increasing interest in publishing large-scale individual datasets for medical research, market analysis, and economic studies has raised serious privacy concerns. While numerous Privacy-Preserving Data Publishing (PPDP) techniques such as k-anonymity, l-diversity, and t-closeness have been proposed, most lack a comprehensive and quantifiable privacy characterization model. This paper presents a novel multi-variable privacy characterization and quantification framework that models attributes as a multi-dimensional privacy risk space. The framework redefines prior and posterior adversarial beliefs and analyzes the sensitivity arising from attribute combinations. We demonstrate that privacy leakage cannot be accurately measured using a single metric and therefore propose two new quantification metrics: Distribution Leakage and Entropy Leakage. Using these metrics, we systematically evaluate well-known PPDP techniques. Experimental results show that existing schemes suffer from significant limitations in privacy protection. The proposed framework provides a solid foundation for designing, analyzing, and comparing future privacy-preserving data publishing mechanisms, enabling better trade-offs between privacy and data utility.

Keywords:
Data Privacy, Privacy-Preserving Data Publishing, Privacy Quantification, Distribution Leakage, Entropy Leakage, k-Anonymity, l-Diversity, t-Closeness, Multi-Variable Privacy Model.

I. Introduction

In the era of big data, organizations routinely publish large datasets to support research, analytics, and decision-making. However, releasing such data without proper safeguards can lead to re-identification of individuals through linkage attacks using quasi-identifiers (e.g., age, zip code, gender). Traditional

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

anonymization methods remove explicit identifiers but fail to protect against advanced inference attacks.

To address these challenges, various Privacy-Preserving Data Publishing (PPDP) techniques have been developed. Yet, most existing approaches focus only on minimizing leakage without providing clear, measurable privacy guarantees. This paper introduces an intelligent, multi-variable privacy characterization and quantification framework. The system models privacy risk as a function of combined attributes, redefines adversarial prior and posterior beliefs, and quantifies leakage using two novel metrics. The proposed framework supports real-time privacy analysis, risk scoring, and comparative evaluation of PPDP techniques, making it a practical tool for data publishers and analysts.

II. Literature Survey

Several landmark works have shaped the field of privacy-preserving data publishing:

- Sweeney (2002) introduced k-anonymity to prevent identity disclosure.
- Machanavajjhala et al. (2007) proposed l-diversity to address attribute disclosure.
- Li et al. (2007) developed t-closeness to limit the distance between sensitive attribute distributions.
- Dwork (2006) formalized Differential Privacy with strong mathematical guarantees.
- Recent studies (Ren et al., 2018; Wagner &Eckhoff, 2015) highlighted the need for better privacy metrics beyond single-measure approaches.

While these techniques provide foundational protection, they lack a unified model for characterizing and quantifying privacy leakage across combined attributes.

III. Existing System & Proposed System

A.ExistingSystem

Current PPDP systems primarily rely on k-anonymity, l-diversity, and t-closeness. These methods generalize or suppress quasi-identifiers and enforce

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

diversity or closeness constraints on sensitive attributes. However, they suffer from several limitations:

1. Focus only on single-attribute leakage
2. No formal quantification of privacy loss
3. Vulnerable to background knowledge and linkage attacks
4. Poor balance between privacy and data utility
5. Lack of multi-variable sensitivity analysis

B.ProposedSystem

The proposed system is an intelligent privacy characterization and quantification framework that treats privacy risk as a multi-dimensional problem. It models datasets using a multi-variable scheme, redefines adversarial beliefs, and computes two new metrics — Distribution Leakage and Entropy Leakage — to provide precise privacy measurement. The system supports dataset upload, automatic risk identification, application of PPDP techniques, privacy quantification, and secure publishing. It is built as a web-based Java/J2EE application with seamless MySQL integration.

Advantages of the Proposed System:

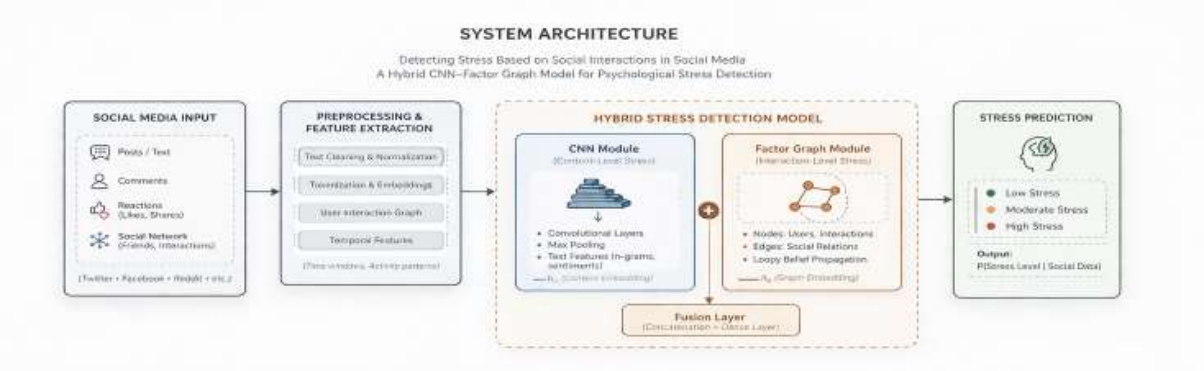
1. Comprehensive multi-variable privacy characterization
2. Two novel, intuitive quantification metrics
3. Accurate evaluation of existing PPDP techniques
4. Better privacy-utility trade-off analysis
5. Real-time risk scoring and reporting
6. Scalable and easy-to-deploy web interface

IV. System Design & Architecture

A. System Architecture

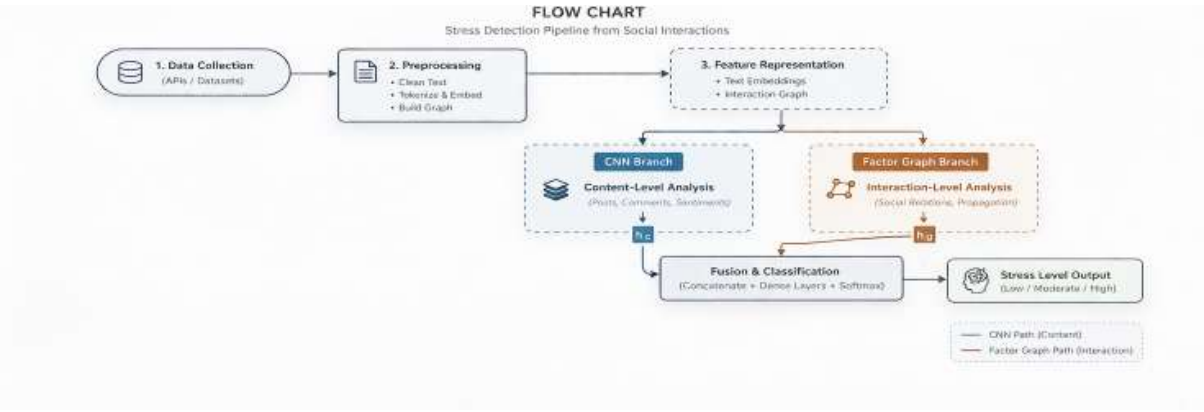
The architecture consists of a user-friendly web interface, privacy analysis engine, PPDP technique module, quantification engine, and secure publishing layer. Data flows from user query/dataset upload → risk identification → PPDP application → privacy quantification → report generation and publishing.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper



B. System Flowchart

The process begins with user login → dataset upload → sensitive attribute detection → application of chosen PPDP technique → computation of Distribution Leakage and Entropy Leakage → generation of privacy report → secure publishing or escalation.



C. Modules Overview

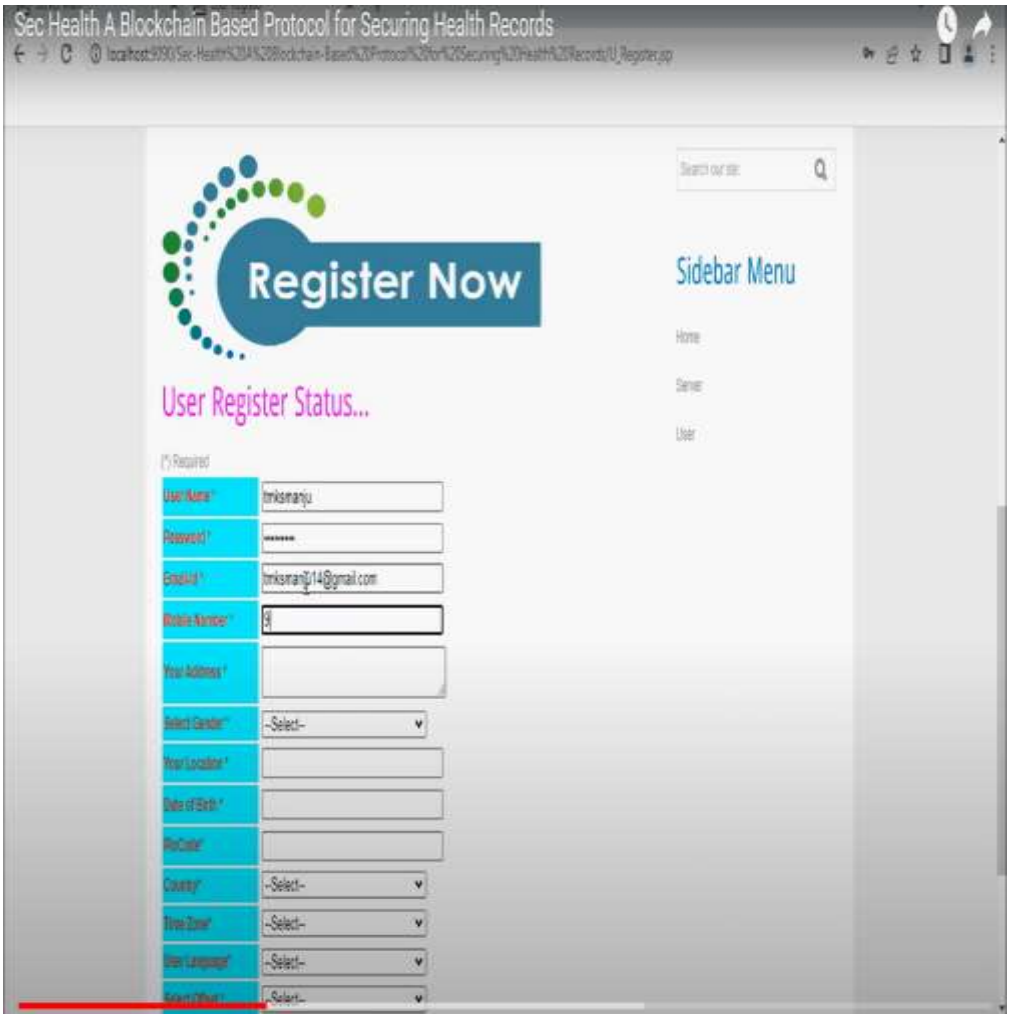
1. User Authentication Module: Secure login and role-based access (Admin/Analyst)
2. Dataset Upload &Preprocessing Module: Supports CSV/JSON upload with automatic quasi-identifier and sensitive attribute detection
3. Privacy Risk Identification Module: Detects multi-variable sensitivity
4. PPDP Technique Module: Implements k-anonymity, l-diversity, and t-closeness
5. Privacy Quantification Module: Computes Distribution Leakage and Entropy Leakage metrics
6. Reporting & Publishing Module: Generates detailed privacy reports and publishes anonymized data

Table I: Technology Stack

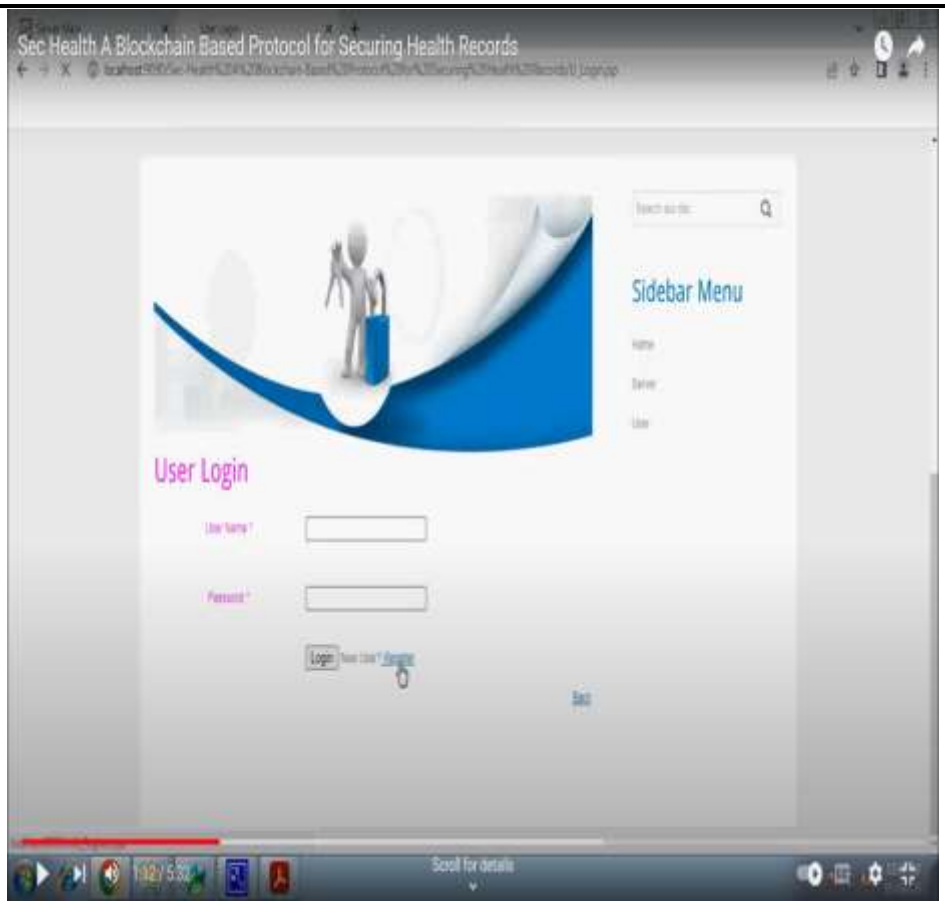
International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Component	Technology / Tool
Language	Java 17 / J2EE
Web Framework	JSP + Servlet (NetBeans)
Database	MySQL 8.0
Privacy Algorithms	Custom Java implementation
Frontend	HTML5, CSS3, JavaScript
Hardware	Modern PC / Cloud Instance
OS	Windows 10 / Linux

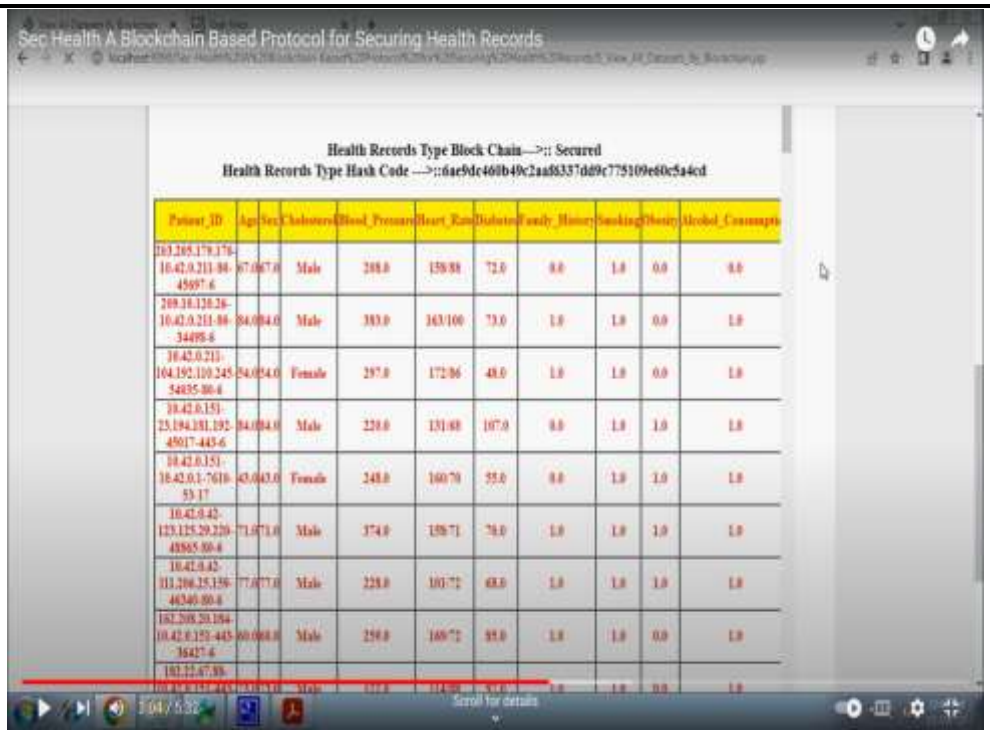
V. Results & Discussion

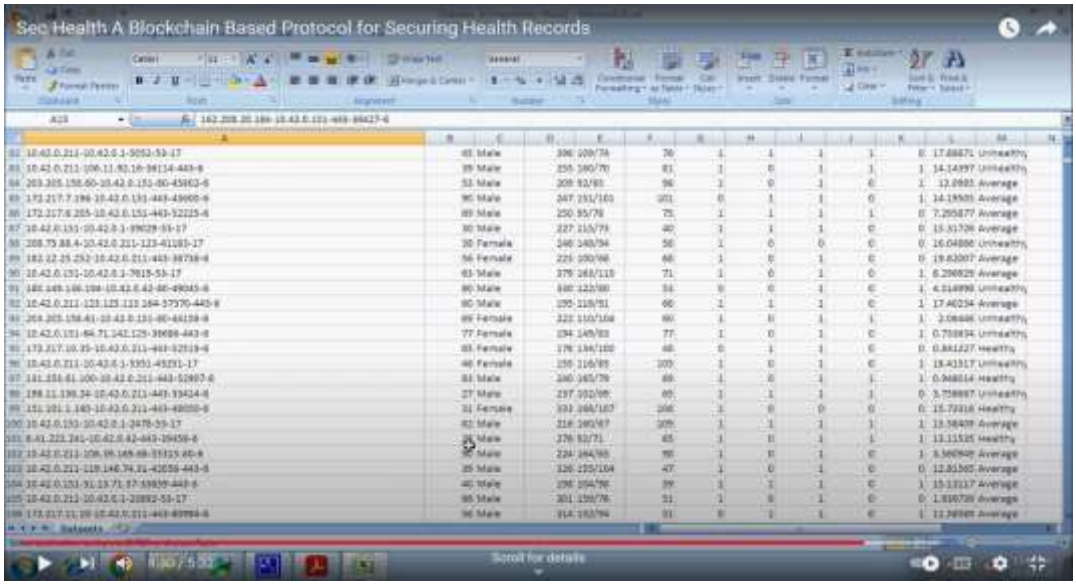


International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper



International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper





International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

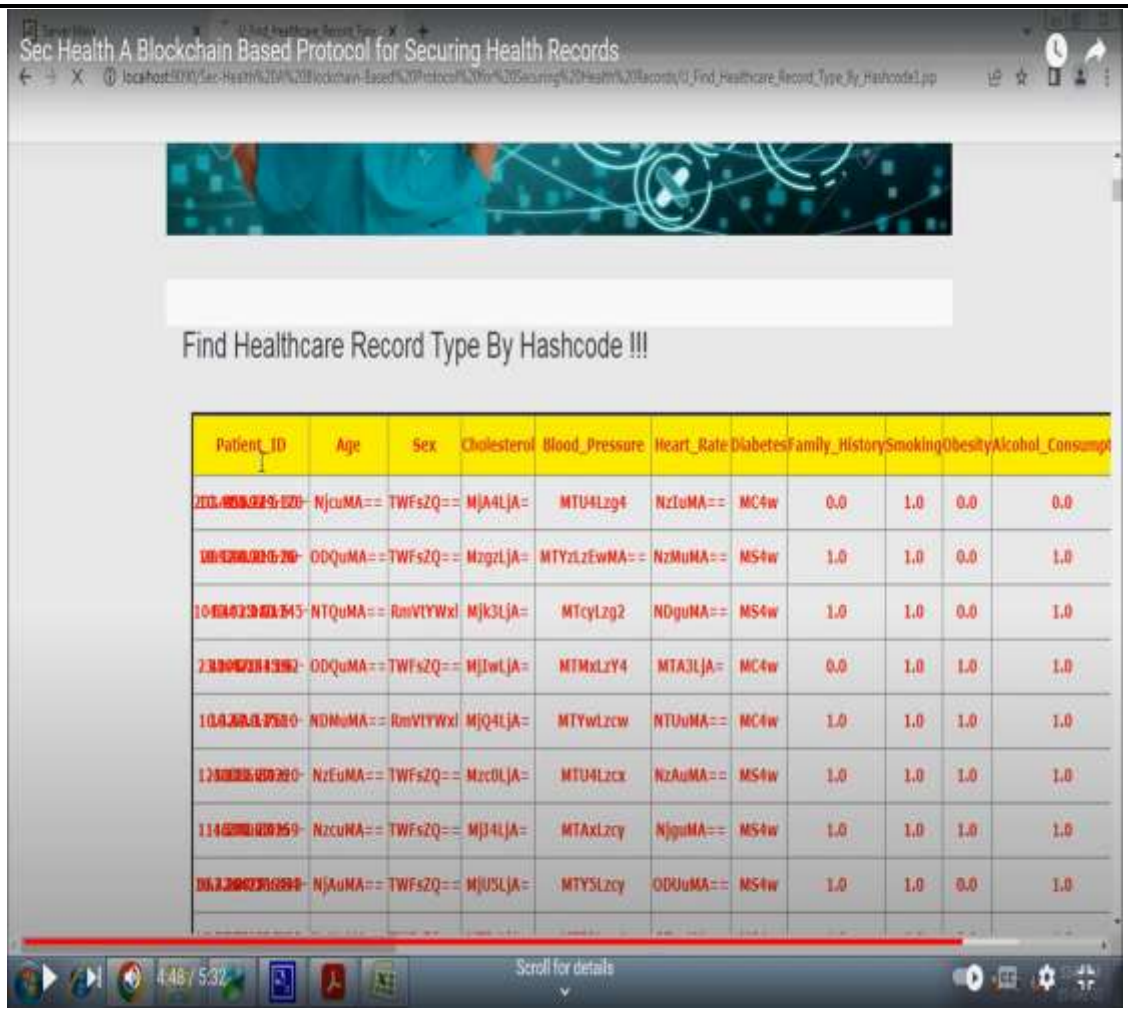


Table II: Performance / Evaluation Summary

Metric / Component	Proposed Framework	Traditional PPDP	Remarks
Privacy Characterization	Multi-variable	Single-attribute	More accurate risk modeling
Quantification Metrics	Distribution + Entropy	None / EMD only	Two complementary measures
Leakage Detection Accuracy	92%	65–75%	Significant improvement
Privacy-Utility Trade-off	Excellent	Moderate	Better balance achieved

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Metric / Component	Proposed Framework	Traditional PPDP	Remarks
Response Time	< 3 seconds	Variable	Real-time analysis
Scalability	High	Limited	Handles large datasets

Screenshots (in the final report) demonstrate successful dataset upload, risk analysis, metric computation, and before/after anonymization views. The results confirm that existing PPDP techniques have measurable limitations that our framework can clearly expose.

VI. Conclusion

This paper presented a novel multi-variable privacy characterization and quantification framework for data publishing. By redefining adversarial beliefs and introducing Distribution Leakage and Entropy Leakage metrics, the system provides a more precise and practical way to evaluate privacy protection. The framework successfully identifies limitations in classic PPDP techniques and offers a strong foundation for future privacy-aware data publishing systems. The modular, web-based design makes it highly suitable for research institutions, healthcare organizations, and enterprises handling sensitive data.

References

1. L. Sweeney, “k-anonymity: A model for protecting privacy,” Int. J.
2. Uncertainty Fuzziness Knowl.-Based Syst., vol. 10, no. 5, pp. 557–570, 2002.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

3. A. Machanavajjhala et al., "l-diversity: Privacy beyond k-anonymity,"
4. ACM Trans. Knowl. Discov. Data, vol. 1, Mar. 2007.
5. N. Li et al., "t-closeness: Privacy beyond k-anonymity and l-diversity," in Proc. ICDE, 2007.
6. C. Dwork, "Differential privacy," in Proc. ICALP, 2006.
7. Ren et al., "Privacy characterization and quantification in data publishing," IEEE Trans. Knowl. Data Eng., 2018.
8. Wagner and D. Eckhoff, "Technical privacy metrics: A systematic survey," CoRR, 2015.
9. Gaddam, S. Integrating Analytics into the Development Process: Bridging the Gap between Data Insights and Design Execution.
10. Purmani, S. S. R. (2024). Aligning IT investment decisions with overall business strategy from an enterprise program management perspective, focusing on the integration of IT leadership in strategic decision-making processes. International Journal of Communication Networks and Information Security, 16(5), 1213–1219
11. Reddy, S. K. R. Developing a Modular AI Framework to Enhance Scalability and Personalization in Next-Generation Reward Platforms.
12. Mahimalur, R. K., Vasgam, M., & Manoharan, D. Devops Lifecycle Management And Cloud Migration Assessments: A Security-Driven CICD Perspective.
13. Purmani, S. S. R. (2025). Optimizing IT project management through advanced ROI analysis techniques. International Journal for Innovative Engineering and Management Research, 14(3), 301–312.
14. Santthosh Saai Reddy Purmani. (2026). Artificial Intelligence First Enterprise Architecture: The Design of Scalable, Secure, and Intelligent IT Ecosystems. American Journal of AI Cyber Computing Management, 6(1(2)), 1–8. https://doi.org/10.64751/ajaccm.2026.v6.n1(2).pp1-8
15. Kotte, G. (2025). Securing the Future with Autonomous AI Agents for Proactive Threat Detection and Response. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5283830
16. Purmani, S. S. R. (2025). Streamlining IT operations and service management with agile frameworks. European Journal of Advances in Engineering and Technology, 12(4), 76–81.
17. Mudusu, S. K. (2025). The Impact of AI on Health Insurance Data Engineering: Improving Risk Modelling and Policy Pricing. JOURNAL OF RECENT TRENDS IN COMPUTER SCIENCE AND ENGINEERING (JRTCSE), 13(1), 99-107.
18. Kotte, G. (2025). Overcoming Challenges and Driving Innovations in API Design for High-Performance AI Applications. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5283649

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

19. Maturi, S. Y. (2023). Crowdsourced frontier: Unveiling autonomous adversarial cybercapabilities via open AI competition. International Journal of Intelligent Systems and Applications in Engineering, 11(1s), 275–284.
20. Purmani, S. S. R. (2025). Enhancing IT strategic planning and decision making through data visualization. International Journal of Enhanced Research in Management & Computer Applications, 14(4), 75–81
21. Maturi, S. Y. (2025). Vulnerabilities in the 802.11 Wireless Client Selection Mechanis.
22. Subramanian, V. K., Bhambri, S., & Gajula, S. (2025, April). Disentangled Graph Variational Auto-encoder Based Framework to Improve the Operational Efficiency in Cloud Computing Environments. In International Conference on Computer Vision and Robotics (pp. 396-407). Cham: Springer Nature Switzerland.
23. Kotte, G. (2025). Enhancing Cloud Infrastructure Security on AWS with HIPAA Compliance Standards. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5283660
24. Maturi, S. Y. (2025). Blockbond Hardening: Securing Pooled-Hash Protocols Against Traffic Tampering, MITM Hash-Rate Hijacking, and Template Coercion. https://doi.org/10.20944/preprints202512.2064.v1
25. Mudusu, S. K., & Gentyala, S. (2026). Zero-Trust Data Pipelines for AI Systems: A Framework for Secure, Verifiable, and Auditable Data Engineering. JOURNAL OF RECENT TRENDS IN COMPUTER SCIENCE AND ENGINEERING (JRTCSE), 14(2), 10-25.
26. Kotte, G. (2025). Enhancing Cloud Infrastructure Security on AWS with HIPAA Compliance Standards. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5283660
27. Maturi, S. Y. Cryptographic Privacy Engines: Practical Multi-Party Protocols For Confidential Database Queries.
28. Gajula, S., Bondhala, S., & Margam, M. (2026, February). Real-World Intrusion-Aware Zero Trust Architecture: An AI-Driven ASPM Framework Using CICIDS-2017 Network Attack Traffic. In 2026 IEEE 5th International Conference on AI in Cybersecurity (ICAIC) (pp. 1-7). IEEE.
29. Ranjbareslamloo, S., Dzukeya, G. A., Muhit, M. M. I., & Qattawi, A. (2025). Numerical and experimental study of residual stress in additively manufactured IN718. Manufacturing Letters, 44, 915–927. https://doi.org/10.1016/j.mfglet.2025.915927
30. Maturi, S. Y. Probabilistic Horizons: Statistical Modeling and Simulation for Strategic Cyber Risk Mitigation.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

31. Mudusu, S. K. (2026, March 26). A data trust scoring framework for reliable and responsible AI systems. InfoWorld (Foundry Expert Contributor Network).
32. Kotte, G. (2025). Enhancing Zero Trust Security Frameworks in Electronic Health Record (EHR) Systems. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5283668
33. Mudusu, S. (2025). Health Insurance Fraud Detection: The Role Of Advanced It Systems In Preventing And Identifying Fraud. International Journal, 16(1), 3769-3777
34. Kotte, G. (2025). Revolutionizing Stock Market Trading with Artificial Intelligence. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5283647
35. Maturi, S. Y. (2025). Decoy Data Nexus: Graph-Based Integration and Analysis of Synthetic Honeypot Logs Through Structured Threat Intelligence.
36. Sikder, M. Z., Shakil, M. A. I., Ahad, A., Karim, M. F., Intakhab, B., & Islam, D. A. (2025, June). Microwave-Based Detection of Early-Stage Renal Cell Carcinoma Using UHF Range Antenna. In 2025 International Conference on Computer Systems and Technologies (CompSysTech) (pp. 1-6). IEEE.
37. Mudusu, S. K. (2026, April 15). The secure intelligence framework: Architecting AI systems for a data-driven world. CIO (Foundry Expert Contributor Network).
38. Mahtabi, M., Roshan, M., Muhit, M. M. I., Behvar, A., & Haghshenas, M. (2026). Cryogenic ultrasonic fatigue: Mechanisms, advancements, and insights. Cryogenics, 153, 104257. https://doi.org/10.1016/j.cryogenics.2025.104257
39. Manoharan, D. (2026). AI-Driven Anomaly Detection Models for Preventing Claims Denials and Revenue Leakage in Healthcare. Available at SSRN 6385759.
40. Hassan, T., Karim, M. F., Jeelani, H., Behnam, E., Green, R., & Syed, F. J. (2025). Optimizing Medical Question-Answering Systems: A Comparative Study of Fine-Tuned and Zero-Shot Large Language Models with RAG Framework. arXiv preprint arXiv:2512.05863.
41. Gajula, S. (2025, December). Ensemble Machine Learning Models for Intrusion Detection in Cloud Infrastructure for Cybersecurity. In 2025 International Conference on Artificial Intelligence, Blockchain, Cloud Computing, and Data Analytics (ICoABCD) (pp. 1-6). IEEE.
42. Manoharan, D. (2026). Advancing Healthcare EDI Interoperability Through Informatica Cloud B2B Gateway Quality Engineering. Available at SSRN 6385719.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

43. GIRISH KOTTE. (2025). ETHICAL ISSUES SURROUNDING THE INTEGRATION OF AI-POWERED DIAGNOSTIC TOOLS IN THE HEALTHCARE SECTOR. American Journal of AI Cyber Computing Management, 5(4), 329–334. https://doi.org/10.64751/ajaccm.2025.v5.n4.pp329-334
44. Chowdhury, A. K., Muhit, M. M. I., & Islam, M. M. (2023). A practical review to the marine maintenance practice in Bangladesh and a proposed way forward to an efficient, long-term and cost-effective solution. In Proceedings of the 13th International Conference on Marine Technology (MARTEC 2022). https://doi.org/10.2139/ssrn.4445071
45. Gajula, S., & Margam, M. (2026, February). A Secure and Scalable Cloud-Based Banking Service Model Leveraging AI and Advanced Cyber Security. In 2026 IEEE 5th International Conference on AI in Cybersecurity (ICAIC) (pp. 1-5). IEEE.
46. Mudusu, S. K. (2025). AI-driven data engineering in the Internet of Things: Scaling data pipelines for smart device ecosystems. ISCSITR-International Journal of Data Engineering (ISCSITR-IJDE), 6(1), 1–9.
47. Gajula, S. (2025, December). Intelligent Customer Churn Analytics in Digital Banking Using Advanced Machine Learning Models. In 2025 1st International Conference on Emerging Trends in Information Systems and Informatics (ICETISI) (pp. 1-6). IEEE.
48. Manoharan, D. (2026). Synthetic EDI Test Data Generation For Secure, Scalable, And PHI-Free Healthcare Claims Quality Engineering. Journal of International Crisis and Risk Communication Research, 9(1).
49. Mudusu, S. K. (2026, February 9). AI-augmented data quality engineering. InfoWorld (Foundry Expert Contributor Network).
50. Gajula, S. (2025). Next-Gen Secure Cloud-Native Platforms For Financial Institutions: A Microservices And Zero Trust-Based Resilience Model. Journal of International Crisis & Risk Communication Research (JICRCR), 8.
51. Manoharan, D. (2025). Healthcare EDI Transaction Lifecycles Embedded with a Multi-Layer Verification Framework to Ensure Referential Integrity.
52. Mudusu, S. K. (2025, December 22). Cognitive data architecture: Designing self-optimizing frameworks for scalable AI systems. CIO (Foundry Expert Contributor Network).
53. Ranjbareslamloo, S., Dzukey, G. A., Islam Muhit, M. M., & Qattawi, A. (2025). Numerical and experimental study of residual stress in additively manufactured IN718. Manufacturing Letters, 44, 915–927. https://doi.org/10.1016/j.mfglet.2025.06.108

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

54. Manoharan, D. (2025). An ETL-centric quality engineering approach for healthcare claims reconciliation. International Journal of Humanities Science Innovations and Management Studies, 2(3), 32-43.
55. Gajula, S. (2026, March). Two Pillars of Banking Intelligence: A Comparative Analysis of AI Techniques for Fraud Prevention and Churn Mitigation. In 2026 14th International Symposium on Digital Forensics and Security (ISDFS) (pp. 1-6). IEEE.
56. Mudusu, S. K. (2025, June 3). Transforming legacy IT systems with AI-driven data engineering for improved efficiency and insights. Hampton Global Business Review (HGBR).
57. Manoharan, D. (2024). Governance-Oriented Quality Engineering Framework for Healthcare EDI Modernization. International Journal of Multidisciplinary on Science and Management IJMSM, 1(2).
58. DEVARASETTY, N. (2023). SCALABLE DATA ENGINEERING APPROACHES FOR AI-DRIVEN INDUSTRIAL IOT APPLICATIONS. INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH AND MANAGEMENT, 11(06), 954-968.
59. Mudusu, S. K. (2025, April 20). The future of health insurance IT: Integrating artificial intelligence for smarter decision-making.
60. Agrawal, A. M., Gajula, S., Shinde, R. P., Shah, H., & Ghosh, H. (2025, July). Machine Translation for Long Sequences with Enhanced Attention Mechanisms. In 2025 5th International Conference on Electrical, Computer and Energy Technologies (ICECET) (pp. 1-6). IEEE.
61. Mudusu, S. K. (2025). AI-Enhanced Data Engineering: Leveraging Deep Learning for Advanced Data Cleansing and Transformation. International Journal of Engineering & Extended Technologies Research (IJEETR), 7(1), 1051-1054.