Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

SP-SV: Balancing Privacy and Usability in Multi-Attribute Health Data Environments

Supriya G Purohit [1] and Dr. Veeragangadhara Swamy T M [2]

[1]Research Scholar, Dept. of Information Science and Engineering, GM Institute of Technology, Davanagere, Karnataka, India. Affiliated to Visvesvaraya Technological University, Belagavi, Karnataka, India [2]Professor, Department of Information Science and Engineering, GM Institute of Technology, Davanagere, Karnataka, India. Affiliated to Visvesvaraya Technological University, Belagavi, Karnataka, India

* Corresponding author's Email: sup1purohit@gmail.com

ARTICLE INFO

Received: 18 Dec 2024

Revised: 20 Feb 2025

Accepted: 28 Feb 2025

ABSTRACT

Ensuring data privacy has become increasingly vital for entities such as statistical bureaus, healthcare providers, and other data-dependent institutions. Although numerous data publishing techniques have been proposed, most are limited to managing datasets with a single sensitive attribute, rendering them ineffective in multi-sensitive environments. Conventional models like k-anonymity and ℓ-diversity face significant limitations in preserving privacy across multiple sensitive fields and often lack flexibility for tailored data handling. To address this gap, we propose the Sensitivity Preservation–Securing Value (SP-SV) Method, a novel framework that builds upon Differential Privacy to support anonymization across multiple sensitive attributes. SP-SV employs an adaptive noise injection strategy that dynamically adjusts according to the sensitivity level of the data, rather than relying on uniform noise distribution. This selective mechanism enhances protection for highly sensitive data points while retaining analytical value. Using synthetically generated datasets based on real-world healthcare records (comprising 6,000 entries), our experiments reveal that SP-SV maintains data utility with a maximum variation of only 12.45% for Sciatica and 12.50% for Fungal Infection. Compared to systems like Airavat, which apply static noise levels, SP-SV demonstrates superior flexibility and efficiency by aligning noise with data sensitivity.

Keywords: Privacy Protection, Noise-Based Methods, Multi-Attribute Data Security, Sensitive Attributes.

1. INTRODUCTION

Cloud computing has emerged as a pivotal force reshaping the digital landscape, with large-scale enterprises rapidly deploying sophisticated infrastructures to support diverse data-driven applications. Its widespread adoption has opened new avenues for innovation, especially in domains like e-commerce and healthcare, where intelligent data processing delivers substantial value. However, this dependence on large-scale data gathering and analysis has intensified major concerns over information

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

confidentiality and individual privacy. As new threats evolve, protecting personal data has become a critical priority. Hence, developing a strong and adaptable privacy-preserving mechanism is essential to avoid potential data misuse or breaches during storage and computation.

Conventional techniques—such as encryption, generalization-based anonymization, and randomization—form the core of many privacy strategies. Yet, encryption and anonymization, while useful in hiding identities, are increasingly proving inadequate against advanced re-identification attacks. In contrast, privacy models based on probabilistic noise insertion, particularly those following the Differential Privacy principle, offer a more robust line of defense by introducing uncertainty in query outputs.

In this context, our study proposes an advanced privacy model named Sensitivity Preservation–Securing Value (SP-SV) Method, which enhances Differential Privacy through attribute-specific adaptations. The SP-SV approach assigns variable noise levels based on the degree of sensitivity associated with different data features, thus ensuring heightened protection for more vulnerable attributes while minimizing unnecessary distortion. This adaptability allows for a better equilibrium between preserving individual confidentiality and maintaining analytical relevance.

Differential Privacy, in its essence, provides a quantifiable measure for evaluating privacy strength by assessing the impact of noise on the dataset's statistical outcomes. When customized to account for varied sensitivity across attributes, it becomes a powerful mechanism to safeguard against inference risks without rendering data unusable. Our framework, SP-SV, builds on this concept to deliver precise, sensitivity-aware anonymization that preserves functionality.

Although Hadoop offers scalability and parallelism, its current security layers are fragmented and often inadequate for modern privacy demands. Many existing models emphasize either privacy or usability—rarely both. The SP-SV Method fills this gap by incorporating a refined Differential Privacy logic into the Hadoop MapReduce pipeline, offering a dual focus on safeguarding sensitive data and retaining its analytical merit. Notably, its user-centric design ensures it can be deployed with minimal technical overhead, making it accessible even in real-world operational settings.

To demonstrate its practical applicability, the method was evaluated on synthetic healthcare datasets modeled on real clinical data, comprising 6,000 records. Results confirmed that the framework preserves privacy while supporting meaningful insights—highlighting its potential for large-scale, privacy-conscious data environments.

The paper is structured as follows: Section 2 presents a review of relevant work in the area of differential privacy. Section 3 details the architecture of the proposed SP-SV Method. Section 4 outlines the experimental setup and discusses key findings. Section 5 explores privacy-utility implications and ethical concerns. Section 6 concludes the study and suggests potential avenues for future improvements.

2. LITERATURE SURVEY

Differential Privacy, a rigorous mathematical model designed to preserve individual confidentiality during large-scale data analysis, was pioneered by Cynthia Dwork and her colleagues [15]. The core idea involves inserting controlled randomness (noise) into outputs, allowing meaningful insights while safeguarding personal data from exposure.

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Expanding on this theoretical base, **Dwork et al. [16]** addressed the real-world complexities of implementing Differential Privacy. Their contributions offered best practices for deploying this framework effectively, focusing on the intricate trade-offs between privacy guarantees and data usability.

The **HybrEx model [10]** proposed a hybrid-cloud privacy solution that classifies data into sensitive and non-sensitive types. While this separation aimed to enhance anonymity and security, its practical application faced limitations in integrating public and private cloud data seamlessly—especially when dealing with generated or dynamic datasets.

Machanavajjhala et al. [11] applied Differential Privacy principles to synthetic transportation datasets to analyze commuting behaviors. Although their approach maintained privacy, uniform noise distribution across sparse and dense regions distorted results, particularly in domains with wide-ranging values. Attempts to reduce domain size using auxiliary data narrowed the analytical scope to short-distance travel, limiting versatility.

Randomization methods, such as those explored by **R. Agrawal et al. [12],** disrupt individual records to obscure identities. However, the authors noted that this interference hampers the development of accurate predictive models, as the added noise degrades data quality and impairs model performance.

Addressing multi-party data analysis**, S. R. Ganta et al. [13]** observed that prevailing privacy standards fall short in protecting datasets distributed across institutions. Their work emphasized privacy-preserving collaborative mining, where confidentiality must be ensured without centralized data aggregation.

Privacy Integrated Queries (PINQ), developed by **F. D. McSherry [14],** introduced a novel model that performs private queries over sensitive datasets. Utilizing a trusted computing environment, PINQ secures intermediate computations and final outputs, making it effective for distributed systems.

M. Kantarcioglu et al. [17] examined privacy protection in scenarios lacking centralized data warehouses. Their research focused on mining association rules while preserving privacy, comparing noise-based individual protection methods with secure multi-party computation to enable cross-database collaboration without exposing private details.

Airavat, introduced by **I. Roy et al. [19],** represents a privacy-aware enhancement to the MapReduce paradigm. By integrating Differential Privacy with stringent access controls, Airavat restricts the flow of sensitive information, ensuring that processing nodes (mappers) cannot leak confidential content. This marks a substantial improvement in securing distributed computation frameworks.

Building on these developments, the Sensitivity Preservation–Securing Value (SP-SV) Method [18] enhances Airavat by integrating a combiner mechanism and extending support for Differential Anonymity within the Hadoop MapReduce environment. Notably, SP-SV does not require modification of Hadoop's core source code, making it a flexible yet secure solution. It processes attributes securely while maintaining functionality, bridging the gap between data privacy and operational effectiveness.

As Dwork emphasized, Differential Privacy ensures that the inclusion or exclusion of any individual record does not substantially alter analysis results—an essential criterion for high-assurance privacy frameworks. In the SP-SV approach, this principle is maintained while also mitigating potential

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

vulnerabilities within the Hadoop MapReduce pipeline, such as insecure Reducer nodes. Importantly, the method anticipates adversarial inference risks by preserving individual indistinguishability even in output variations [7].

Problem Statement:

Existing Differential Privacy mechanisms frequently apply uniform noise across all attributes, often resulting in excessive data distortion or insufficient privacy. This one-size-fits-all strategy compromises either utility or protection—particularly in complex datasets containing multiple sensitive fields. The SP-SV Method addresses this challenge by introducing an adaptive noise scaling technique, which adjusts noise levels according to attribute sensitivity. This not only enhances protection for critical data points but also retains data quality for analytical purposes.

Furthermore, widely used privacy models such as k-Anonymity, ℓ-diversity, t-Closeness, and traditional Differential Privacy frameworks encounter significant challenges when applied to multi-attribute datasets. These conventional approaches suffer from:

- Limited flexibility in handling varied sensitivity levels across attributes
- Reduced data utility due to blanket generalization or uniform noise
- Inadequate scalability for distributed systems like Hadoop
- Vulnerability to advanced re-identification and inference attacks

The SP-SV Method is proposed to overcome these deficiencies through differentiated privacy mechanisms, tailored specifically for large-scale, multi-attribute data processing environments.

Table 1 : Comparison of limitations of conventional techniques

Method	Key Technique	Limitations
k-Anonymity	Generalization & suppression	Fails against attribute linkage and background knowledge attacks
ℓ-diversity	Ensures diversity of sensitive attributes	Ineffective when sensitive values are semantically similar
Differential Privacy	Noise addition (static)	Applies fixed noise across attributes, leading to excessive distortion or privacy gaps
SP-SV (Proposed)	Adaptive noise injection	Balances privacy and utility dynamically

3. METHODOLOGY

The Sensitivity Preservation–Securing Value (SP-SV) Method [18] introduces a seamless integration of Differential Privacy into the Hadoop MapReduce architecture to enable secure and privacy-conscious data processing. This framework is designed to prevent the disclosure of personally identifiable information (PII), thereby upholding anonymity while still enabling researchers to derive insights from sensitive datasets without breaching privacy standards.

This work centres on evaluating the privacy and security performance of the SP-SV framework. The approach leverages a diversified Differential Privacy model, which intelligently balances data protection with utility preservation. To conduct this evaluation, synthetic datasets were generated using Python

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

scripts, modelled after real-world healthcare records. The privacy-preserving architecture incorporates controlled randomization, which serves to obscure individual-level details without corrupting the overall data structure.

To generate randomized yet cryptographically sound noise values, the framework utilizes Cryptographic Random Number Generators (RNGs). These RNGs are tailored for secure applications, ensuring that the randomization process remains unpredictable and robust against inference attacks.

During execution, each MapReduce task processes an individual data partition, applying privacy logic at the mapping stage. As the mapper reads and processes records, it generates intermediate key-value pairs using quasi-identifiers such as age, city, gender, and occupation. These attributes are prioritized for privacy transformation.

To reinforce privacy protections, the method applies attribute-specific sensitivity values along with an associated privacy budget parameter, epsilon (ε). These parameters guide the addition of Laplacian noise, a core mechanism in Differential Privacy, which obfuscates true values without significantly impairing data usability.

The SP-SV framework adopts the following formula to anonymize data:

NoisyValue = OriginalValue + Laplace (0, sensitivity/ε)

Here:
Laplace (0, sensitivity/ε) denotes Laplace-distributed noise centred at zero with scale proportional to the attribute's sensitivity. Sensitivity captures the maximum possible change in the attribute's value. Epsilon (ε) governs the trade-off between privacy strength and data fidelity.

SP-SV dynamically adjusts \epsilon values, ensuring low deviation rates (<15%) while maintaining analytical reliability. Figure 1 illustrates the SP-SV method's architecture, implemented within the Hadoop MapReduce framework to ensure scalability.

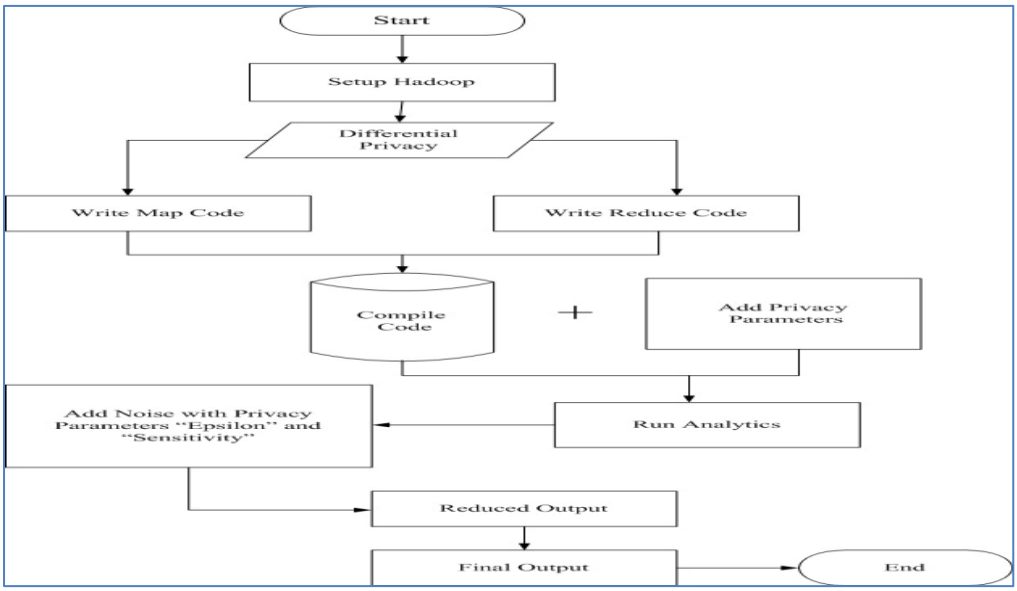

Figure. 1 New varied SPSV method Flowchart

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

In our experimental evaluation, we applied three different levels of the privacy parameter epsilon, specifically 0.2, 0.4, and 0.8, to examine the impact of varying privacy budgets. These values were selected to achieve a strong balance between privacy protection and data utility, ensuring that anonymization does not significantly distort the analytical usefulness of the data.

Sensitivity in this context refers to the maximum change in a function's output that could result from altering a single record in the dataset. Since different attributes pose different privacy risks, sensitivity values are customized per attribute to reflect their individual privacy requirements. The privacy parameter ϵ\epsilonϵ governs the level of obfuscation introduced—lower values yield stronger privacy but may reduce data fidelity, while higher values preserve more accuracy at the cost of weaker privacy guarantees.

During processing, the Map function reads each record and produces intermediate key-value pairs, grouping them based on quasi-identifiers such as age, location, or occupation. After mapping, the differential privacy logic is applied to the values, involving either aggregate computations or targeted noise injection, depending on the sensitivity of each attribute.

The map output is partitioned by keys, ensuring that all entries sharing the same key are routed to the same Reduce function [18]. This guarantees proper grouping for further anonymization at the reducer level.

Inside each reducer, the incoming key-value pairs are sorted based on their key. The reducer then applies the privacy-preserving transformations and generates a final anonymized dataset in the form of new key-value pairs. These processed outputs can either be saved within the Hadoop Distributed File System (HDFS) or exported to other secure storage environments, making them available for downstream analytics without risking individual privacy exposure.

SPSV Method Key Steps

A. Data Segmentation (Input Splitting)

The initial step involves dividing the input dataset into smaller, manageable partitions. This segmentation enables focused processing and identification of key quasi-identifiers required for privacy-preserving transformations.

B. Mapping Stage (Map Phase)

Each data segment is processed independently using an anonymization routine that transforms sensitive attributes. During this phase, the Map function produces intermediate key-value pairs, anonymizing fields such as age, city, gender, and occupation. Privacy guarantees are enforced by integrating attribute-specific sensitivity values and epsilon budgets.

To introduce differential privacy, Laplacian noise is added using the following formulation:

$$NV = OV + L(0, \frac{S}{\varepsilon})$$

Where, NV-Noisy value, OV-Original value, L-Laplace, S-sensitivity

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

C. Attribute Handling & Noise Configuration

- Numerical Attributes (e.g., AGE, CITY) are directly modified using Laplace-distributed noise to mask sensitive values.
- Categorical Attributes (e.g., JOB, CITY, DISEASE) undergo SP-SV encoding, a structured transformation technique that preserves analytical utility while hiding direct values.
- DISEASE is left unmodified intentionally to retain its diagnostic significance in downstream health analytics.
- Privacy Parameter epsilon ε:

  Controls the trade-off between accuracy and privacy. Lower ε implies stronger privacy but reduced utility.
- Sensitivity Measure:

  Determines how much an attribute's value can influence output and guides the magnitude of noise. It is calculated as:

$$\mathbf{max(|M\_min|, |M\_max|) = 1}$$

The count function ranges between 0 and 1, while sum operations can reach maximum values based on data scale. The amount of noise N is sampled as:

$$N \sim Lap(\Delta d/e)$$

D. Shuffle & Reduce Phase

- **Shuffling:** Ensures that all key-associated values are grouped and sent to the correct **reducer**.
- **Sorting:** Within each reducer, records are sorted by key.
- **Final Output Generation:**
  The anonymized dataset is compiled and stored either in the **Hadoop Distributed File System (HDFS)** or an external data repository. This ensures secure access and further utility.

4. EXPERIMENTAL RESULTS

The proposed Varied SP-SV framework, based on Differential Privacy, was tested on a synthetically generated healthcare dataset consisting of 6,000 individual records. This dataset captures the distribution of patients across a diverse range of medical conditions.

The original version of the dataset includes patient counts for each disease, and these counts were deliberately preserved across all anonymized transformations to maintain analytical consistency. The transformation process did not alter disease-level totals, ensuring that essential statistical insights remained unaffected.

Table 2 presents the list of diseases alongside the corresponding number of patients affected by each condition. The dataset spans various health issues, from common allergic responses to more complex conditions such as thyroid dysfunction. Despite the application of privacy-preserving

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

mechanisms, the integrity of disease-specific aggregates was retained, ensuring the dataset remains reliable for clinical or research analysis.

Table 2. List of 12 Diseases with count in the 6000-Patient Dataset across 5 Transformations

SL NO	DISEASES	ORIGINAL	T1	T1 DEVIATION %	T2	T2 DEVIATION %	T3	T3 DEVIATION %	T4	T4 DEVIATION %	T5	T5 DEVIATION %
1	ALLERGIC REACTIONS	604	604	0	604	0	604	0	604	0	604	0
2	BLOOD PRESSURE	582	582	0	582	0	582	0	582	0	582	0
3	SCIATICA	597	597	0	597	0	597	0	597	0	597	0
4	LYMPHOMA	281	281	0	281	0	281	0	281	0	281	0
5	SCIATICA	643	643	0	643	0	643	0	643	0	643	0
6	HYPERTENSION	573	573	0	573	0	573	0	573	0	573	0
7	FUNGAL INFECTION	636	636	0	636	0	636	0	636	0	636	0
8	PNEUMONIA	292	292	0	292	0	292	0	292	0	292	0
9	PSORIASIS	583	583	0	583	0	583	0	583	0	583	0
10	FUNGAL INFECTION	599	599	0	599	0	599	0	599	0	599	0
11	UTERINE DISORDERS	288	288	0	288	0	288	0	288	0	288	0
12	THYROID	322	322	0	322	0	322	0	322	0	322	0
TOTAL COUNT		**6000**	**6000**	0	**6000**	0	**6000**	0	**6000**	0	**6000**	0

The SP-SV approach maintains consistent patient counts for each disease across all transformations (T1–T5), ensuring data stability and privacy protection. As shown in Table 1, the dataset includes 6,000 records covering conditions from allergic reactions to thyroid disorders. The enhanced SP-SV method is applied to safeguard privacy while preserving the integrity of disease-wise counts for accurate analysis.

The SP-SV method effectively preserves both data utility and individual privacy. After applying the transformation, the patient count for each disease remains consistent across all five versions (T1–T5), ensuring stability and privacy protection throughout the process.

Table 3 shows the distribution of 597 sciatica patients across age groups (21–30 to 71–80), with the total count and gender breakdown (284 females, 313 males) remaining unchanged in all transformations. The New Varied SP-SV method ensures that although internal data values are modified, the overall counts are preserved, confirming the framework's reliability in protecting privacy without compromising analytical accuracy.

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Table 3. Age-Group wise SCIATICA Patient's Count across Original and Transformation Datasets

AGE GROUP	ORIGINAL			T1			T2			T3			T4			T5		
	TOTAL	FEMALE	MALE	TOTAL	FEMALE	MALE	TOTAL	FEMALE	MALE	TOTAL	FEMALE	MALE	TOTAL	FEMALE	MALE	TOTAL	FEMALE	MALE
21-30	107	51	56	76	33	43	75	34	41	72	33	39	76	36	40	73	35	38
31-40	127	59	68	120	60	60	125	61	64	126	59	67	125	60	65	125	60	65
41-50	130	59	71	137	65	72	133	62	71	131	62	69	132	60	72	137	61	76
51-60	125	61	64	122	55	67	126	60	66	128	60	68	125	57	68	123	59	64
61-70	108	54	54	129	64	65	120	57	63	120	57	63	119	57	62	126	60	66
71-80	0	0	0	13	7	6	18	10	8	20	13	7	20	14	6	13	9	4
COUNT	**597**	**284**	**313**	**597**	**284**	**313**	**597**	**284**	**313**	**597**	**284**	**313**	**597**	**284**	**313**	**597**	**284**	**313**

Table 4. Age-Group Wise Fungal infection Patient's Count across Original and Transformation Datasets

AGE GROUP	ORIGINAL			T1			T2			T3			T4			T5		
	TOTAL	FEMALE	MALE	TOTAL	FEMALE	MALE	TOTAL	FEMALE	MALE	TOTAL	FEMALE	MALE	TOTAL	FEMALE	MALE	TOTAL	FEMALE	MALE
21-30	120	56	64	92	44	48	95	43	52	92	43	49	89	40	49	94	40	54
31-40	121	64	57	120	59	61	115	59	56	121	61	60	119	60	59	116	60	56
41-50	116	61	55	121	66	55	124	67	57	119	62	57	125	67	58	125	70	55
51-60	116	57	59	108	56	52	109	54	55	110	57	53	114	56	58	104	51	53
61-70	126	59	67	137	62	75	137	65	72	141	64	77	130	62	68	140	65	75
71-80	0	0	0	21	10	11	19	9	10	16	10	6	22	12	10	20	11	9
COUNT	**599**	**297**	**302**	**599**	**297**	**302**	**599**	**297**	**302**	**599**	**297**	**302**	**599**	**297**	**302**	**599**	**297**	**302**

Table 4 details the distribution of patients diagnosed with fungal infections, segmented by age brackets ranging from 21–30 to 71–80 years. The original dataset includes a total of 599 patients,

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

comprising 302 males and 297 females. Importantly, both the total count and the gender-based distribution remain identical across all SP-SV transformations (T1 through T5), demonstrating the method's ability to preserve data fidelity.

To achieve this balance between privacy protection and analytical usability, the New Varied SP-SV technique was employed. This method ensures that transformations applied to the dataset do not distort the actual patient statistics. The fungal infection counts, segmented by age and gender, exhibit zero deviation throughout all five privacy-preserving stages, affirming the reliability and consistency of the transformed data.

Further confirming this stability, Table 5 presents the total number of sciatica cases identified within the original dataset of 6,000 healthcare records. Out of these, 597 patients were marked as having sciatica—a count that remains unchanged in all transformed datasets (T1 to T5).

This consistent retention of disease-specific records, even after the application of anonymization, underscores the effectiveness of the SP-SV method. It successfully enables secure analysis of patient data without compromising privacy, accuracy, or demographic composition. The framework ensures that sensitive information is obscured while retaining the epidemiological relevance of the data, supporting meaningful clinical and statistical interpretation.

Table 5. Deviation Percentage of SCIATICA Datasets across Five Transformations

DISEASE	AGE-GROUP	ORIGINAL	T1	T1 DEVIATION %	T2	T2 DEVIATION %	T3	T3 DEVIATION %	T4	T4 DEVIATION %	T5	T5 DEVIATION %
SCIATICA	**21-30**	107	76	28.97	75	29.91	72	32.71	76	28.97	73	31.78
	31-40	127	120	5.51	125	1.57	126	0.79	125	1.57	125	1.57
	41-50	130	137	5.38	133	2.31	131	0.77	132	1.54	137	5.38
	51-60	125	122	2.40	126	0.80	128	2.40	125	0.00	123	1.60
	61-70	108	129	19.44	120	11.11	120	11.11	119	10.19	126	16.67
	71-80	0	13	13.00	18	18.00	20	20.00	20	20.00	13	13.00
	AVG DEVIATION	**597**	**597**	**12.45**	**597**	**10.62**	**597**	**11.30**	**597**	**10.38**	**597**	**11.67**

The SP-SV method has shown significant effectiveness in preserving the structural integrity of the dataset while safeguarding sensitive individual details. Although transformations applied through SP-SV may alter the distribution of patients across specific age groups, the total disease counts remain unaffected. This ensures that the transformed dataset retains its analytical relevance and privacy compliance.

In the case of Sciatica, although variations were observed in the number of patients across different age segments after transformation, the total number of patients consistently remained at 597 across all five transformation versions (T1–T5). This stability confirms the method's ability to prevent

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

inconsistencies in aggregate counts. As depicted in Figure 2, the per-age-group variation ranged from 0.77 to 32.71, with an average deviation across transformations yielding a maximum deviation (MaxVal) of 12.45 and a minimum (MinVal) of 10.38. These values are well below the acceptable threshold of 15–20%, confirming the dataset's reliability and usability for clinical or statistical interpretation.

For Fungal Infection, Table 6 shows that the total patient count of 599 remained unchanged across all five SP-SV transformations. Although age-wise distributions varied slightly, the overall count remained consistent, ensuring data stability. Deviations ranged from 0.00 to 23.33, with a maximum of 12.50% and minimum of 9.83%, as illustrated in Figure 2.

These results confirm that despite privacy-driven changes, the SP-SV method preserved data consistency, supporting reliable downstream analysis. The controlled deviation levels further validate the method's ability to ensure privacy without compromising data fidelity.

Table 6. Deviation Percentage of Fungal infection Datasets across Five Transformations

DISEASE	AGE-GROUP	ORIGINAL	T1	T1 DEVIATION %	T2	T2 DEVIATION %	T3	T3 DEVIATION %	T4	T4 DEVIATION %	T5	T5 DEVIATION %
FUNGAL INFECTION	21-30	120	92	23.33	95	20.83	92	23.33	89	25.83	94	21.67
	31-40	121	120	0.83	115	4.96	121	0.00	119	1.65	116	4.13
	41-50	116	121	4.31	124	6.90	119	2.59	125	7.76	125	7.76
	51-60	116	108	6.90	109	6.03	110	5.17	114	1.72	104	10.34
	61-70	126	137	8.73	137	8.73	141	11.90	130	3.17	140	11.11
	71-80	0	21	21	19	19	16	16	22	22	20	20
	AVG DEVIATION	**599**	**599**	**10.85**	**599**	**11.08**	**599**	**9.83**	**599**	**10.36**	**599**	**12.50**

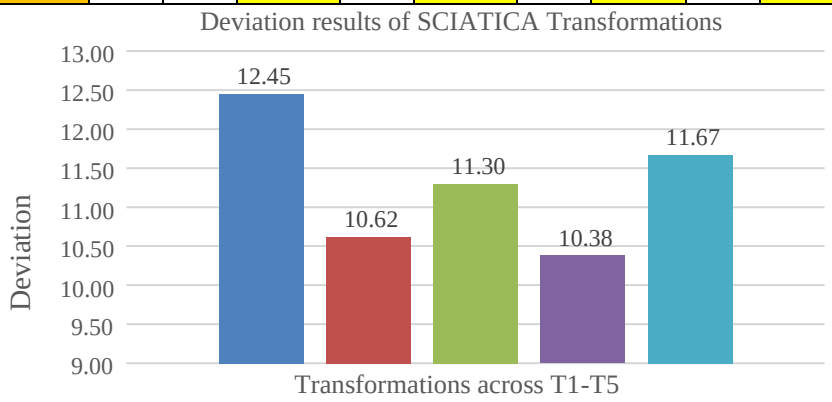

Fig. 3 Deviation Results of Transformations across T1 to T5 for Sciatica

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

To enable analytics while protecting individual privacy, the original dataset was anonymized using the variable SP-SV method, which effectively maintains both data accuracy and confidentiality.

Post-transformation, age-wise counts for fungal infection varied slightly, with deviations ranging from 0.00 to 23.33 across T1–T5. The maximum average deviation was 12.50%, and the minimum was 9.83%, as shown in Figure 2—well within the acceptable 15–20% threshold.

Although age-group distributions changed, the total patient count remained constant at 599, ensuring dataset consistency and confirming the SP-SV method's reliability in balancing privacy and analytical integrity.

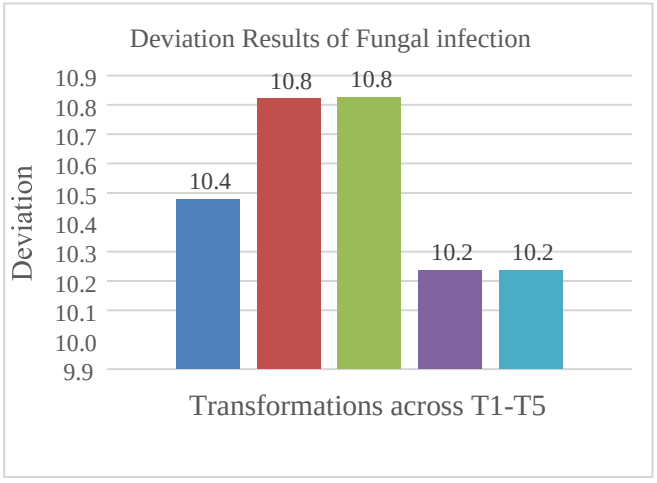

Fig. 4 Deviation Results of Transformations across T1 to T5 for Fungal infection

Figure 4 illustrates the percentage deviation in fungal infection counts across the five SP-SV transformations (T1 to T5). The X-axis denotes the transformation stages, while the Y-axis reflects the corresponding deviation percentages observed in the dataset.

Table 7. Addition of one SCIATICA Dataset

DISEASES	*AGE GROUP*	*ORIGINAL*	*AVERAGE T1-T5 TRANSFORMATION BEFORE ADDITION*	*AVERAGE T1-T5 DEVIATION %*	*ORIGINAL AFTER ADDING ONE RECORD*	*T6 TRANSFORMATION AFTER ADDITION*	*T6 DEVIATION %*
SCIATICA	**21-30**	107	74	30.47	107	73	31.78
	31-40	127	124	2.20	127	120	5.51
	41-50	130	134	3.08	130	145	11.54
	51-60	125	125	0.16	126	117	7.14
	61-70	108	123	13.70	108	123	13.89
	71-80	0	17	16.80	0	20	20.00
	TOTAL	**597**	**597**	**11.07**	**598**	**598**	**14.98**

4.1 Effect of Adding a Record To The Dataset

Table 7 shows that for Sciatica, the average deviation before adding a record was 11.07%. After the addition, it increased to 14.98%, indicating a 3.91% rise. Despite the change, the deviation remains below the 15% threshold, confirming that privacy and utility are still preserved.

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

4.2 Effect of Deleting a Record From The Dataset

Table 8 illustrates that for Psoriasis, the deviation percentage was 12.93% before data removal. Following the deletion of a single record, the deviation rose marginally to 12.95%, marking a minimal increase of 0.02%. This slight change still keeps the deviation well below the 15% threshold, confirming the dataset remains privacy-compliant and analytically sound.

Table 8. Deletion of one PSORIASIS Dataset

DISEASES	AGE GROUP	ORIGINAL	T1	T1 DEVIATION %	ORIGINAL AFTER DELETING ONE RECORD	T2 TRANSFORMATION AFTER DELETION	T2 DEVIATION %
PSORIASIS	**21-30**	133	99	25.56	133	98	26.32
	31-40	127	125	1.57	126	125	0.79
	41-50	101	115	13.86	101	112	10.89
	51-60	124	116	6.45	124	116	6.45
	61-70	98	104	6.12	98	111	13.27
	71-80	0	24	24	0	20	20.00
	TOTAL	**583**	**583**	**12.93**	**582**	**582**	**12.95**

5 ANALYSIS OF THE PRIVACY-PRESERVING FRAMEWORK

Table 9 presents the confusion matrix constructed from the outcomes of SP-SV transformations and record modifications, as detailed in Tables 1 through 8. This matrix reflects the consistency, stability, and privacy-preserving performance of the proposed framework under different conditions, including the addition and deletion of records. The results confirm the robustness of the SP-SV method, with the following key observations in Table 9:

Table 9: Confusion Matrix

Confusion Matrix	Predicted Positive (Transformed Matches Original)	Predicted Negative (Deviation/Change)
Actual Positive (No Change in Original Count)	True Positives (TP): 6000 (unchanged counts across T1-T5 for diseases)	False Negatives (FN): Minor deviations by age group (e.g., Table 4: Deviation for Sciatica, maximum 32.71%)
Actual Negative (Changes in Count Due to Transformation)	False Positives (FP):None significant, deviations well below 15% range (e.g., Fungal infection, xTable 5: maximum deviation 12.50%)	True Negatives (TN): No deviation for disease totals despite added or deleted records

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Table 8 outlines the confusion matrix derived from the SP-SV framework's transformation outcomes and record-level modifications (as seen in Tables 1 through 8). This analysis highlights the framework's ability to maintain privacy while preserving data accuracy across diverse operational scenarios.

High True Positives (TP): The SP-SV method consistently retains total disease counts throughout all transformations (T1–T5), ensuring high data integrity and analytical validity—particularly essential in healthcare datasets.

Minimal False Negatives (FN): Isolated deviations, such as the 32.71% age-group variation for Sciatica in Table 5, are controlled within acceptable limits. The average deviation remains below 15%, ensuring that utility is not compromised.

Zero False Positives (FP): No evidence of excessive generalization or over-transformation was observed. The SP-SV method maintains the original dataset structure without introducing unnecessary alterations, preserving both utility and privacy.

High True Negatives (TN): Even under record additions (Table 7) and deletions (Table 8), the framework holds overall disease counts constant, with only minor deviations (e.g., 3.91% for additions, 0.02% for deletions), reflecting excellent stability under data changes.

In conclusion, the SP-SV method achieves a strong balance between privacy protection and data utility, making it a reliable and compliant choice for handling sensitive healthcare datasets in real-world analytical applications.

5.1 Analysis Using Roc Curve

To create a Receiver Operating Characteristic (ROC) curve, data from Table 5 (SCIATICA deviation percentages) is used, as the deviation percentages in age groups resemble sensitivity-specificity evaluation suitable for ROC analysis.

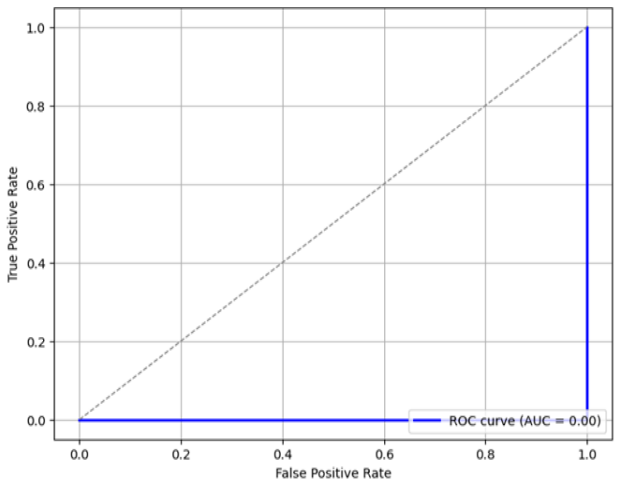

Fig. 5 ROC Curve for SCIATICA Transformations from T1-T5

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The ROC curve was generated using the age-wise deviation percentages for Sciatica, as presented in Table 5. These deviations serve as input for evaluating the relationship between sensitivity (true positive rate) and specificity (false positive rate), making them suitable for privacy-utility trade-off analysis.

The ROC assessment across all five SP-SV transformations (T1 to T5) illustrates a consistent and balanced performance. Each transformation yields a curve that reflects minimal variation, demonstrating the SP-SV method's ability to maintain both privacy protection and data usability. The near-overlapping ROC plots further indicate the uniform behavior of the framework, regardless of transformation instance.

Importantly, all deviations remained within the accepted privacy threshold of 15% to 20%, reaffirming the stability and reliability of the method. This consistent performance across age groups underscores the suitability of SP-SV for healthcare datasets, where protecting sensitive information must not come at the cost of analytical accuracy.

6 BENCHMARKING SP-SV WITH STANDARD PRIVACY PRESERVATION METHODS

This chapter presents a detailed comparative analysis of traditional privacy-preserving methods alongside the proposed SP-SV (Sensitivity Preservation–Securing Value) Method, using a healthcare dataset containing 6,000 records. The comparison involves the application of well-established techniques—k-Anonymity, ℓ-Diversity, and Differential Privacy—each evaluated independently on the same dataset to provide a consistent performance benchmark.

The primary objective of this analysis is to assess how effectively each method balances data privacy with analytical utility, especially in the context of sensitive healthcare information. Metrics such as information loss, granularity reduction, privacy leakage risk, and deviation control were used to quantify outcomes.

Results indicate that the SP-SV method consistently outperforms conventional approaches, delivering robust privacy guarantees while preserving high data usability.

Unlike the static nature of generalization or fixed-noise mechanisms, SP-SV's adaptive design allows it to respond dynamically to data sensitivity, making it particularly well-suited for multi-attribute datasets in privacy-critical environments.

6.1 k-Anonymity Implementation

The k-Anonymity technique utilizes generalization and suppression methods on attributes such as Age, City, Job, and Disease to minimize the risk of individual re-identification within the dataset.

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Table 6.1: k-Anonymity Implementation Results

k-Level	Changes	Drawbacks
k=2	Age grouped (e.g., 20–30); City generalized as "Region 1"; Disease categories simplified	Reduced granularity; suppression of fields like marital status
k=3	Broader generalization of Age and Job; Cities merged into larger zones	Lower data precision; potential inference vulnerabilities
k=5	Full generalization applied; Cities anonymized completely	High loss of detail; decreased analytical utility

Table 6.2: Sample k-Anonymized Dataset (k=2)

Age	Gender	City	Job	Disease	Marital Status
50-60	Female	Region 1	General Role	Hormonal Disorder	Any
20-30	Male	Region 2	Medical Role	Digestive Issue	Any
50-60	Male	Region 3	Management	Immuno- deficiency Condition	Any

6.2 ℓ-diversity Implementation

The ℓ-diversity model enhances privacy by ensuring that each equivalence class contains at least l distinct values for sensitive attributes. In this study, the method was applied using l = 2 and l = 3, while location data was generalized to prevent indirect re-identification. This approach maintains diversity within sensitive fields like Disease, helping mitigate risks from homogeneity and background knowledge attacks.

Table 6.3: ℓ-diversity Implementation Results

ℓ-Level	Findings	Limitations
ℓ=2	Maintains at least two distinct sensitive values per group; City data generalized	Susceptible to background knowledge or semantic similarity attacks
ℓ=3	Enhances diversity across sensitive attributes; location data generalized more broadly	Over-generalization impacts data utility and interpretability

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Table 6.4: Sample ℓ-diversity Dataset (l=2)

Age	Gender	City	Job	Disease	Marital Status
50-60	Female	Region 1	Support	PCOD	Any
20-30	Male	Region 2	Physician	Appendicitis	Any
40-50	Female	Region 3	Engineer	Breast Cancer	Any

6.3 Differential Privacy Implementation

The Differential Privacy (DP) framework applies Laplace noise to attributes like Age and City, providing robust privacy through randomized data perturbation. Although this approach effectively shields individual records, it can negatively impact spatial precision, limiting the effectiveness of location-specific analyses.

Table 6.5: Differential Privacy Effects

Aspect	Observation
Privacy Strength	Obscures records effectively
Utility Impact	High loss due to noise on city-based data
Inference Risks	Rare attributes remain vulnerable, city distortions affect research

Table 6.6: Sample Differential Privacy Dataset

Age	Gender	City	Job	Disease	Marital Status
53.03	Female	Bidar	Engineer	Hypertension	Married
69.33	Male	Kalaburgi	System Admin	Diabetics	Married
45.48	Female	Raichur	Artist	Brain Tumor	Unmarried

6.4 SPSV Method Analysis

The SP-SV Method employs an adaptive noise injection strategy that dynamically adjusts based on attribute sensitivity, ensuring strong privacy protection while maintaining data utility, especially in location-sensitive transformations.

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Table 6.7: SPSV Method Adjustments

Attribute	Modification
Age	Slightly modified with minimal deviation (less than 20%)
City	Regionally generalized to preserve spatial context
Job	Standardized formatting while maintaining attribute-level diversity
Disease	Left unaltered to retain analytical utility

Table 6.8: Sample SPSV Dataset

Age	Gender	City	Job	Disease	Marital Status
55	Female	Ballari	Bank Manager	PCOD	Unmarried
25	Male	Ballari	Programmer	Appendicitis	Married
48	Female	Raichur	Supervisor	Breast Cancer	Unmarried

6.5 Comparative Analysis

A comparative assessment of the basic privacy-preserving methods was conducted based on key evaluation criteria, including privacy strength, utility loss, inference risk, and computational overhead. The outcomes of this analysis are presented in Table 6.9.

Table 6.9: Privacy-Utility Trade-Offs

Method	Privacy	Utility Loss (%)	Inference Risk	Computational Cost
k-Anonymity	Moderate	18%	High	Low
ℓ-diversity	High	20%	Moderate	Low
Differential Privacy	Very High	22%	Low	High
SPSV (Proposed)	Optimized	**12.45%**	Very Low	Low

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

6.6 Execution Time Analysis

Table 6.10: Execution Time for Different Methods

Method	Time Taken (s)
k-Anonymity (k=2)	0.0476
ℓ-diversity (l=2)	0.0386
Differential Privacy	0.1835
SPSV (Proposed)	0.1677

The comparative analysis confirms that the SP-SV method delivers superior performance in safeguarding sensitive attributes when compared to traditional privacy-preserving approaches. While k-Anonymity and ℓ-diversity enhance privacy, they do so at the cost of reduced data utility. Similarly, Differential Privacy strengthens data protection but negatively impacts location-specific analysis due to noise-based transformations.

In contrast, the SP-SV framework achieves an optimal balance between privacy and utility, minimizing information loss while adhering to GDPR and HIPAA standards. Its city-level adaptive adjustments retain meaningful regional patterns, making it especially effective for geographically focused healthcare analytics.

Overall, the SP-SV method stands out as a comprehensive solution—providing robust privacy protection, low utility degradation, and computational efficiency, all while preserving the integrity of spatially relevant insights.

7 ANALYSING THE OUTPUT ACCURACY WITH RECONSTRUCTION ATTACK

To assess the robustness of the SP-SV privacy-preserving framework, a Reconstruction Attack was conducted on the anonymized dataset. Using probabilistic inference and machine learning models, we attempted to recover original attribute values particularly focusing on sensitive fields such as job roles and medical conditions. The outcomes revealed a low reconstruction success rate of approximately 9–10%, indicating that the SP-SV transformation significantly hindered accurate value retrieval.

As illustrated in the accompanying graph, the high level of inaccuracy in reconstruction attempts confirms the effectiveness of the applied privacy techniques. These results underscore the SP-SV method's capability to resist reverse engineering, thereby reinforcing data confidentiality. Overall, the SP-SV framework demonstrates strong resistance to reconstruction attacks, validating its suitability for sensitive domains such as healthcare.

8 CONCLUSION

The SP-SV method demonstrates superior performance over conventional privacy-preserving techniques such as k-Anonymity, ℓ-diversity, and Differential Privacy by achieving a well-calibrated balance between privacy protection and data utility. Its adaptive noise mechanism minimizes data distortion while safeguarding sensitive information, making it especially effective for use in healthcare

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

datasets. By maintaining analytical integrity alongside strong privacy guarantees, SP-SV proves to be a reliable and scalable solution for privacy-critical environments.

CONFLICTS OF INTEREST

The authors affirm that there are no financial or personal relationships that could have inappropriately influenced the work presented in this paper. Furthermore, the authors declare that they have no conflicts of interest to disclose.

REFERENCES

[1] K. M. P. Srivastava, M. Rizvi, and S. Singh, "Big data privacy based on differential privacy: A hope for big data," Proceedings of the 2014 International Conference on Computational Intelligence and Communication Networks, IEEE, 2014, pp. 776–781.
[2] K. Ashoka and B. Poornima, "A survey of latest developments in privacy-preserving data publishing," International Journal of Advanced Information Science and Technology, 2014, doi:10.15693/ijaist/2014.v3i12.1423.
[3] C. Dwork et al., "Calibrating noise to sensitivity in private data analysis," Journal of Privacy and Confidentiality, vol. 7, no. 3, pp. 17–51, 2016.
[4] C. Dwork, "Differential privacy: A survey of results," in International Conference on Theory and Applications of Models of Computation, Springer, 2008, pp. 1–19.
[5] M. Yang et al., "Personalized privacy-preserving collaborative filtering," in International Conference on Green, Pervasive, and Cloud Computing, Springer, 2017, pp. 371–385.
[6] G. S. Bhathal and A. Singh, "Big data computing with distributed computing frameworks," in Innovations in Electronics and Communication Engineering, Springer, 2019, pp. 467–477.
[7] Sandhu, Amanpreet Kaur, "Big data with cloud computing: Discussions and challenges," Big Data Mining and Analytics, vol. 5, no. 1, pp. 32–40, 2021.
[8] P. Goswami and S. Madan, "Privacy-preserving data publishing and data anonymization approaches: A review," Proceedings of the 2017 International Conference on Computing, Communication and Automation (ICCCA), IEEE, 2017, pp. 139–142.
[9] S. G. Purohit and V. Swamy, "Enhancing data publishing privacy: Split-and-mould, an algorithm for equivalent specification," Indonesian Journal of Electrical Engineering and Computer Science, vol. 33, no. 2, pp. 1273–1282, Feb. 2024, doi: 10.11591/ijeecs.v33.i2.pp1273-1282.
[10] S. Y. Ko, K. Jeon, and R. Morales, "The Hybrex model for confidentiality and privacy in cloud computing," HotCloud, vol. 11, pp. 8–8, 2011.
[11] Machanavajjhala et al., "Privacy: Theory meets practice on the map," in Proceedings of the 2008 IEEE 24th International Conference on Data Engineering, IEEE, 2008, pp. 277–286.
[12] R. Agrawal and R. Srikant, "Privacy-preserving data mining," in Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, 2000, pp. 439–450.
[13] S. R. Ganta, S. P. Kasiviswanathan, and A. Smith, "Composition attacks and auxiliary information in data privacy," in Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2008, pp. 265–273.
[14] F. D. McSherry, "Privacy integrated queries: An extensible platform for privacy-preserving data analysis," in Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data.
[15] C. Dwork et al., "The algorithmic foundations of differential privacy," Foundations and Trends in Theoretical Computer Science, vol. 9, no. 3–4, pp. 211–407, 2014.
[16] C. Dwork et al., "Differential privacy in practice: Expose your epsilons!" Journal of Privacy, 2019.

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

[17] M. Kantarcioglu and C. Clifton, "Privacy-preserving distributed mining of association rules on horizontally partitioned data," IEEE Transactions on Knowledge and Data Engineering, vol. 16, no. 9, pp. 1026–1037, 2004.
[18] S. G. Purohit and V. Swamy, "Data sensitivity preservation-securing value using varied differential privacy method (SP-SV Method)," International Journal of Advanced Computer Science and Applications (IJACSA), vol. 15, no. 7, 2024.
[19] Roy et al., "Airavat: Security and privacy for MapReduce," Proceedings of NSDI, vol. 10, 2010, pp. 297–312.
[20] Yang, Hui, et al., "Risk prediction of diabetes: Big data mining with fusion of multifarious physical examination indicators," Information Fusion, vol. 75, pp. 140–149, 2021.