Adaptive Hybrid Privacy Preserving Machine Learning Across Heterogeneous Domains

Loubna Ali
Faculty of Computer Science and Informatics
Berlin School of Business and Innovation (BSBI)
Berlin, Germany

Noufal Issa
Department of Applied Informatics
Faculty of Mathematics, Physics and Informatics
Comenius University, Bratislava, Slovakia

Shkelqim Hajrulla
Computer Engineering Department
Epoka University
Tirana, Albania

Taylan Demir
Department of Mathematics
Çankaya University
Ankara, Turkey

ABSTRACT
The rapid adoption of machine learning across diverse application domains has intensified concerns regarding data privacy. Although numerous privacy-preserving techniques have been proposed, their effectiveness is typically evaluated within isolated domains, which limits the understanding of their generalizability. This paper investigates the domain-dependent behavior of privacy-preserving machine learning through a comprehensive cross-domain empirical study. A hybrid privacy framework is implemented that combines suppression, generalization, and perturbation techniques to protect sensitive information while maintaining data utility. The framework is evaluated across five heterogeneous domains, namely healthcare, finance, social media, cybersecurity, and Internet of Things (IoT) environments, spanning six real-world datasets. The experimental results demonstrate that the effectiveness of privacy-preserving mechanisms varies significantly across domains. Structured and network-based datasets, such as financial and cybersecurity data, maintain high predictive performance under privacy constraints, whereas text-based social data experiences noticeable performance degradation. A complementary trade-off analysis, expressed through utility-retention and a privacy–utility efficiency measure, further quantifies these differences. The findings highlight the limitations of one-size-fits-all privacy solutions and emphasize the need for adaptive, domain-aware privacy strategies in real-world machine learning applications.

General Terms
Security, Machine Learning, Privacy, Algorithms.

Keywords
Privacy-Preserving Machine Learning, Cross-Domain Analysis, Hybrid Privacy Methods, Data Anonymization, Privacy–Utility Trade-off, IoT Security, Cybersecurity, Social Data Privacy.

1. INTRODUCTION
The widespread adoption of machine learning (ML) across critical domains such as healthcare, finance, cybersecurity, and Internet of Things (IoT) systems has significantly increased concerns regarding data privacy. Machine learning is increasingly integrated into healthcare decision support systems and personalized analytical platforms, where protecting sensitive user data is particularly important [1]. Modern ML models rely on large-scale datasets that often contain sensitive information, including personal identifiers, financial transactions, behavioral patterns, and network activity. The exposure of such data can lead to serious consequences, including identity theft, financial fraud, and unauthorized surveillance [2], [3], [4].

To mitigate these risks, a wide range of privacy-preserving techniques have been developed, including data anonymization, perturbation-based methods, differential privacy, and cryptographic approaches such as secure multi-party computation [3], [5], [6]. Among these, data transformation techniques—particularly suppression, generalization, and perturbation—remain widely used due to their simplicity, efficiency, and compatibility with standard machine learning pipelines [7], [8]. These techniques aim to reduce the risk of re-identification while preserving the utility of the data for predictive modeling.

Despite significant progress in this area, most existing studies evaluate privacy-preserving mechanisms within a single application domain. For instance, privacy techniques are often tested independently in healthcare datasets [7], financial transaction systems [9], or network intrusion detection datasets [10], [11]. However, real-world machine learning systems are increasingly deployed across heterogeneous environments, where data characteristics differ substantially. Structured tabular data, textual data, and network traffic data exhibit fundamentally different statistical properties and feature representations, which can directly influence the effectiveness of privacy-preserving techniques.

This raises an important yet underexplored research question: do privacy-preserving mechanisms behave consistently across different domains, or is their effectiveness inherently domain-dependent? Addressing this question is crucial, as the assumption of a one-size-fits-all privacy solution may lead to suboptimal performance in certain domains. Recent works have highlighted the trade-off between privacy and utility, demonstrating that stronger privacy guarantees often come at the cost of reduced model performance [12], [13]. However, these studies rarely investigate how this trade-off varies across fundamentally different types of data.

This paper addresses this gap by conducting a comprehensive cross-domain empirical study of privacy-preserving machine learning. A hybrid privacy framework is implemented that integrates suppression for direct identifiers, generalization for quasi-identifiers, and perturbation for numerical attributes. This unified approach allows the consistent application of privacy transformations across multiple datasets while

maintaining compatibility with standard machine learning models.

The proposed framework is evaluated across five heterogeneous domains: healthcare, finance, social media, cybersecurity, and IoT environments. To ensure a rigorous analysis, both model utility and privacy impact are evaluated using multiple metrics, including accuracy, F1-score, information loss, privacy score, and execution time. To the best of the authors' knowledge, this study is among the first to systematically investigate the domain-dependent behavior of privacy-preserving machine learning across heterogeneous datasets.

The main contributions of this paper are summarized as follows:

- A cross-domain evaluation of privacy-preserving machine learning is conducted across six distinct real-world datasets spanning five application domains.
- A unified hybrid privacy framework that combines suppression, generalization, and perturbation techniques is implemented.
- The privacy–utility trade-off is analyzed using multiple quantitative metrics across heterogeneous data environments.
- It is demonstrated that the effectiveness of privacy-preserving mechanisms is strongly domain-dependent, with significant variation between structured, network, and text-based data.

The experimental results presented in this study show that the proposed hybrid approach preserves predictive performance effectively in structured and network-based datasets, such as financial and cybersecurity data. In contrast, its performance degrades in text-based social datasets, highlighting the limitations of uniform privacy strategies. These findings emphasize the need for adaptive, domain-aware privacy mechanisms in practical machine learning applications.

2. RELATED WORK

Privacy-preserving machine learning has been studied extensively in recent years, with a wide range of techniques proposed to protect sensitive data while maintaining model utility. Existing approaches can generally be categorized into three main groups: anonymization-based methods, perturbation-based techniques, and advanced privacy-preserving frameworks such as federated learning and cryptographic solutions.

Anonymization-based approaches, such as k-anonymity, l-diversity, and t-closeness, aim to prevent re-identification by modifying or generalizing quasi-identifiers in the dataset. While these methods provide a structured way to protect privacy, they often suffer from significant information loss and may not scale effectively to high-dimensional data [14], [15]. Furthermore, these techniques are typically designed for tabular data and are less effective when applied to complex data types such as text or network traffic.

Perturbation-based methods introduce noise into the data to obscure sensitive information while preserving statistical properties. These approaches are widely used due to their simplicity and efficiency, particularly in large-scale systems. However, the level of noise must be carefully calibrated, as excessive perturbation can degrade model performance, while insufficient noise may fail to provide adequate privacy protection [16], [17]. This trade-off between privacy and utility remains a central challenge in privacy-preserving machine learning.

More recently, advanced frameworks such as federated learning and secure computation have gained attention. Federated learning enables decentralized model training without sharing raw data, thereby reducing privacy risks [18]. Similarly, homomorphic encryption and secure multi-party computation allow computations to be performed on encrypted data, providing strong privacy guarantees [19]. Despite their advantages, these approaches often introduce significant computational overhead and complexity, limiting their applicability in resource-constrained environments such as IoT systems.

Several studies have explored hybrid approaches that combine multiple privacy techniques to balance their respective strengths and weaknesses. For example, combined anonymization and perturbation frameworks have been proposed to improve privacy protection while maintaining data utility [20], [21]. Similarly, hybrid models for privacy-preserving data publishing have been investigated, demonstrating improved resistance to re-identification attacks [22]. However, these works primarily focus on improving privacy mechanisms within a single domain and do not evaluate their effectiveness across different types of data.

In parallel, research on privacy-preserving machine learning has also been conducted in domain-specific contexts. For instance, privacy techniques have been applied in healthcare data sharing, financial fraud detection, and network intrusion detection systems [23], [24]. While these studies provide valuable insights, they are typically limited to specific datasets and do not address the broader question of how privacy mechanisms perform across heterogeneous environments.

Despite the extensive literature, a key limitation remains: most existing works assume that a given privacy-preserving technique can be applied uniformly across different domains. There is a lack of systematic analysis examining whether the same privacy mechanism can maintain an optimal balance between privacy and utility across fundamentally different data types, such as structured tabular data, textual data, and network traffic data.

This gap motivates the present study, which differs from existing work by providing a cross-domain empirical evaluation of a unified hybrid privacy framework. Unlike prior approaches that focus on method development within a single domain, this work investigates how privacy-preserving mechanisms behave across multiple real-world environments. To better highlight the differences between existing privacy-preserving approaches and the proposed method, a comparative summary is provided in Table 1.

Table 1. Comparison with Existing Methods

Method	Technique Type	Domain Applicability	Strengths	Limitations
Data Perturbation [16]	Perturbation	Tabular	Preserves statistical properties	Limited to numerical data
k-anonymity [7]	Generalization	Tabular	Strong identity protection	High information loss
l-diversity [14]	Enhanced anonymization	Tabular	Protects attribute disclosure	Vulnerable to skewness
Federated Learning [18]	Distributed learning	Multi-domain	No raw data sharing	High computational cost
Differential Privacy [3]	Noise injection	Multi-domain	Formal privacy guarantees	Utility degradation
Proposed Hybrid Method	**Hybrid (Suppression + Generalization + Perturbation)**	**Cross-domain**	**Balanced privacy–utility trade-off**	**Less effective for text data**

As shown in Table 1, existing methods are typically designed for specific data types or domains, whereas the proposed framework provides a unified solution across multiple domains.

3. METHODOLOGY

3.1 Problem Definition

Let a dataset D consist of feature–label pairs, where each feature vector belongs to a d-dimensional real space and each label denotes the corresponding target variable. The dataset may contain sensitive attributes, including direct identifiers, quasi-identifiers, and numerical features.

$$D = \{(x_i, y_i)\},\ \ i = 1, \ldots, n,\ \ \ x_i \in \mathbb{R}^d$$

The objective of this work is to transform the original dataset D into a privacy-preserved version D', such that sensitive information is protected while the predictive performance of machine learning models trained on D' remains as close as possible to that of models trained on D. Formally, two competing goals are pursued:

- **Privacy preservation:** minimize disclosure risk.
- **Utility preservation:** maximize predictive performance.

3.2 Overview of the Proposed Framework

The proposed framework applies a hybrid privacy transformation consisting of three complementary mechanisms: suppression for direct identifiers, generalization for quasi-identifiers, and perturbation for numerical attributes. The composite transformation function is defined as:

$$D' = T(D) = P(G(S(D)))$$

where $S(\cdot)$ is the suppression function, $G(\cdot)$ is the generalization function, and $P(\cdot)$ is the perturbation function. This formulation ensures the sequential application of privacy mechanisms: identifiers are first removed, quasi-identifiers are then generalized, and numerical features are subsequently perturbed.

3.3 Suppression of Direct Identifiers

Direct identifiers are attributes that uniquely identify individuals or entities, such as user IDs and IP addresses. The suppression operation replaces the values of identifier attributes with a constant token:

$$x'_{ij} = \text{“suppressed”}\ \ if\ j \in X\_id\ ;\ \ x_{ij}\ \ otherwise$$

This ensures that no direct linkage between records and real-world entities can be established.

3.4 Generalization of Quasi-Identifiers

Quasi-identifiers are attributes that may indirectly reveal identity when combined, such as age, gender, time, or protocol. For numerical quasi-identifiers, binning (interval mapping) is applied:

$$x' = \lfloor x / b \rfloor \cdot b$$

where b is the bin size (e.g., $b = 10$) and $\lfloor \cdot \rfloor$ denotes floor division. This maps each value to the lower bound of its interval. For categorical quasi-identifiers, values are replaced with a generalized representation (x' = “generalized”), which reduces attribute granularity and limits re-identification risk.

3.5 Perturbation of Numerical Features

To protect numerical attributes while preserving their statistical properties, Gaussian noise is added to each numerical value:

$$x' = x + \varepsilon,\ \ \ \varepsilon \sim N(0, \sigma^2)$$

In the implementation, the standard deviation of the noise is proportional to the feature variability:

$$\sigma = \alpha \cdot std(x),\ \ \ \alpha = 0.01$$

where α is a small scaling factor. This ensures that large-scale features receive proportionally larger noise while the overall data distribution is preserved.

3.6 Machine Learning Pipeline

After the privacy transformation is applied, both the original and transformed datasets are processed using the same machine learning pipeline. Data preprocessing includes handling missing values and encoding categorical features using label encoding or one-hot encoding. Large datasets are subsampled to ensure computational efficiency. A Random Forest classifier is then trained, and performance is measured on a held-out test set. Using an identical pipeline for both versions guarantees that any observed difference in performance is attributable to the privacy transformation rather than to modeling choices.

3.7 Evaluation Metrics

To analyze the privacy–utility trade-off, utility, privacy, and efficiency metrics are employed. Utility is assessed using accuracy, precision, recall, and F1-score. Privacy is quantified using information loss and a privacy score, and efficiency is measured by execution time.

Information Loss (IL) measures the average difference between the original and transformed features:

$$IL = (1/m) \cdot \Sigma_j \, dist(x_j, x'_j), \quad j = 1, \ldots, m$$

Privacy Score (PS) measures the proportion of modified values:

$$PS = (1/m) \cdot \Sigma_j \, I(x_j \neq x'_j), \quad j = 1, \ldots, m$$

where $I(\cdot)$ is an indicator function and m is the number of attributes. Execution time is recorded for both the original and privacy-preserved pipelines to quantify computational overhead.

3.8 Cross-Domain Experimental Design

The framework is evaluated across five domains that exhibit distinct data characteristics: healthcare (structured tabular data), finance (transactional data), social media (text-based data), cybersecurity (network traffic data), and IoT (device communication data). This design enables the analysis of how privacy-preserving mechanisms behave under different data characteristics, directly supporting the study's objective of identifying domain-dependent effects.

4. EXPERIMENTAL SETUP AND RESULTS

4.1 Datasets

To evaluate the proposed framework, experiments were conducted on six publicly available datasets representing heterogeneous real-world domains. The healthcare domain is represented by the Heart Disease UCI dataset, which contains clinical and demographic attributes used for heart disease prediction [25]. The financial domain is represented by the Credit Card Fraud Detection dataset, a highly imbalanced dataset with anonymized features derived using principal component analysis [9]. The social domain is represented by the Sentiment140 dataset, a large-scale collection of labeled tweets for sentiment analysis [26]. The cybersecurity domain is represented by the UNSW-NB15 dataset, a modern intrusion detection dataset containing both normal and malicious network traffic [10]. The IoT domain is represented by the IoT-23 collection, which captures benign and malicious IoT network behaviors [27]; two distinct capture scenarios from this collection, denoted IoT-1 and IoT-3, are used as independent evaluation settings within the IoT domain. These datasets were selected to represent diverse data types, including structured tabular data, high-dimensional anonymized data, textual data, and network traffic data.

4.2 Experimental Configuration

A unified experimental pipeline was applied across all datasets to ensure consistency. A Random Forest classifier was used with an 80% training and 20% testing split. Preprocessing included handling missing values and encoding categorical variables, and large datasets were subsampled for efficiency. Each dataset was evaluated under two conditions: the original dataset with no privacy transformation, and the privacy-preserved dataset with the hybrid method applied.

4.3 Evaluation Metrics

To assess the privacy–utility trade-off, the metrics defined in Section 3.7 were used: accuracy, precision, recall, and F1-score for utility; information loss and privacy score for privacy; and execution time for efficiency.

4.4 Results

Table 2 reports accuracy, precision, recall, F1-score, information loss, privacy score, and execution time for both the original and privacy-preserved pipelines across all six datasets. These results provide the basis for the detailed per-metric analysis presented in the following subsections.

Table 2. Performance Comparison Across Datasets

Dataset	Method	Accuracy	Precision	Recall	F1-score	Information Loss	Privacy Score	Execution Time (s)
Healthcare	Original	0.663	0.637	0.663	0.649	0.000	0.000	0.141
Healthcare	Proposed	0.565	0.532	0.565	0.545	0.139	0.200	0.144
Finance	Original	0.999	0.999	0.999	0.999	0.000	0.000	0.534
Finance	Proposed	0.999	0.999	0.999	0.999	0.000	1.000	1.120
Social	Original	0.657	0.657	0.657	0.657	0.000	0.000	0.724
Social	Proposed	0.497	0.247	0.497	0.330	0.400	0.400	0.231
Cybersecurity	Original	0.924	0.924	0.924	0.924	0.000	0.000	0.396
Cybersecurity	Proposed	0.909	0.909	0.909	0.907	0.058	0.943	1.286
IoT-1	Original	0.997	0.997	0.997	0.997	0.000	0.000	0.223
IoT-1	Proposed	0.995	0.995	0.995	0.995	0.300	0.619	0.355
IoT-3	Original	0.999	0.999	0.999	0.999	0.000	0.000	0.202
IoT-3	Proposed	0.999	0.999	0.999	0.999	0.300	0.619	0.273

As shown in Table 2, the proposed method maintains high predictive performance in structured and network-based datasets while achieving varying levels of privacy across domains.

4.5 Accuracy and F1-score Analysis

The classification performance is illustrated in Figure 1 and Figure 2.

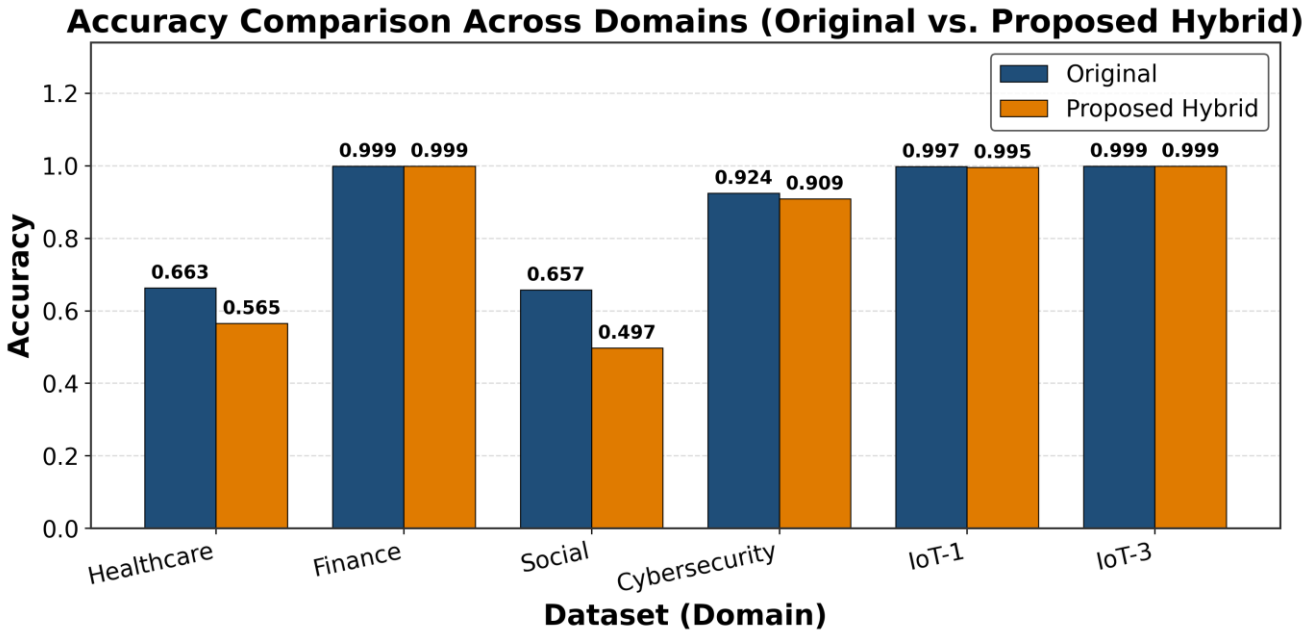

Fig 1: Accuracy comparison across datasets.

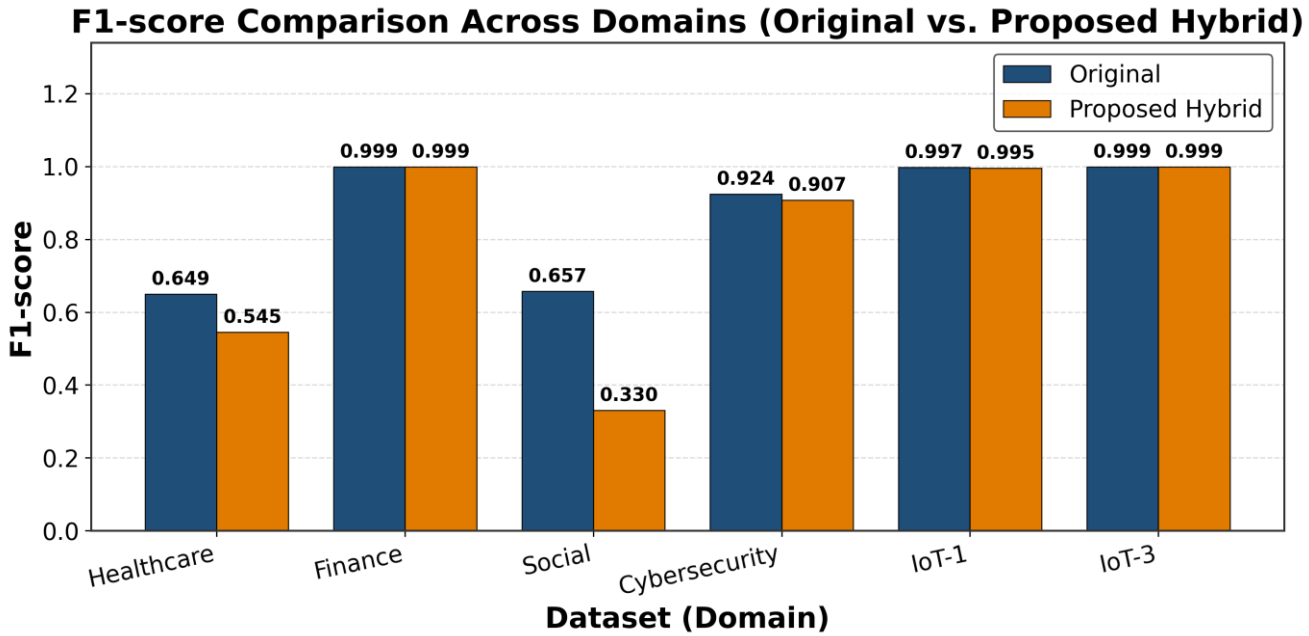

Fig 2: F1-score comparison across datasets.

The results show that the proposed hybrid privacy method preserves model performance in structured and network-based datasets. In the financial dataset, performance remains unchanged (accuracy ≈ 0.999), indicating robustness to privacy transformations. Similarly, both IoT scenarios maintain accuracy above 0.995. In the cybersecurity dataset, only a slight reduction is observed (0.924 → 0.909), which remains acceptable given the achieved privacy level. The healthcare dataset shows a moderate decrease (0.663 → 0.565), suggesting sensitivity to feature modification. The social dataset exhibits the largest degradation, where accuracy drops from 0.657 to 0.497 and the F1-score from 0.657 to 0.330, indicating that privacy transformations significantly affect text-based data. The disparity between the accuracy and F1-score reductions in the social dataset further suggests that minority-class performance is disproportionately affected once privacy transformations are applied to textual features.

4.6 Privacy and Information Loss Analysis

The privacy impact is illustrated in Figure 3 and Figure 4.

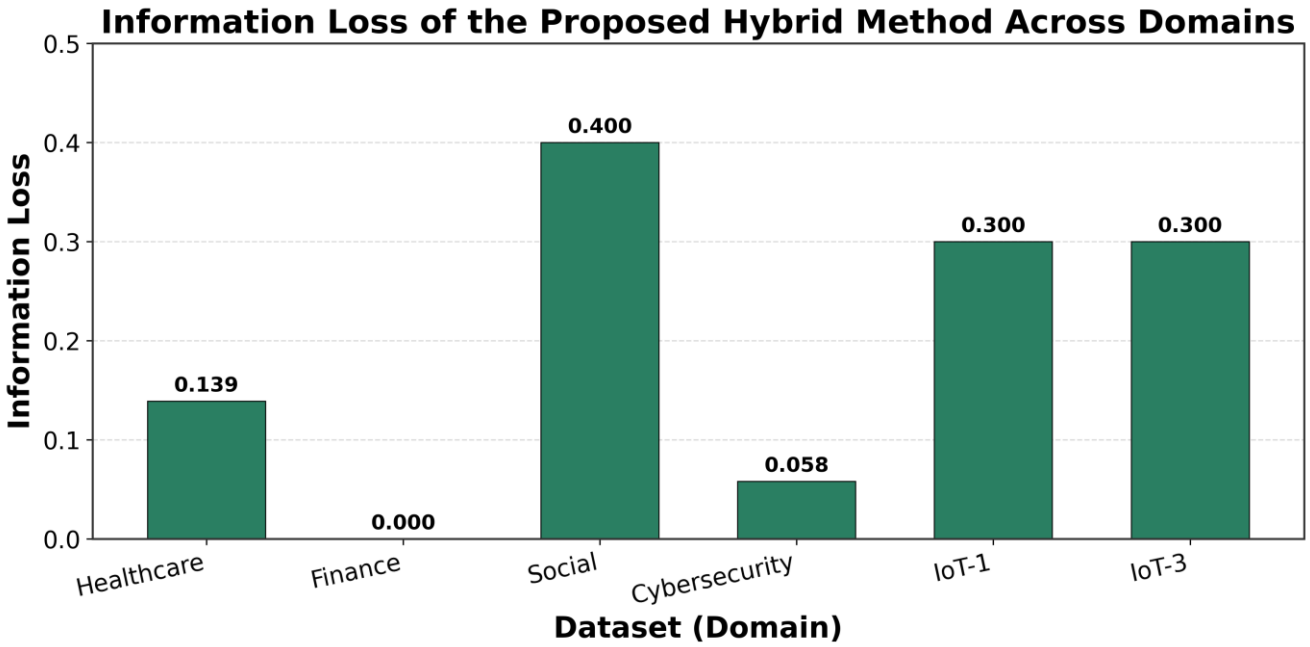

Fig 3: Information loss across datasets.

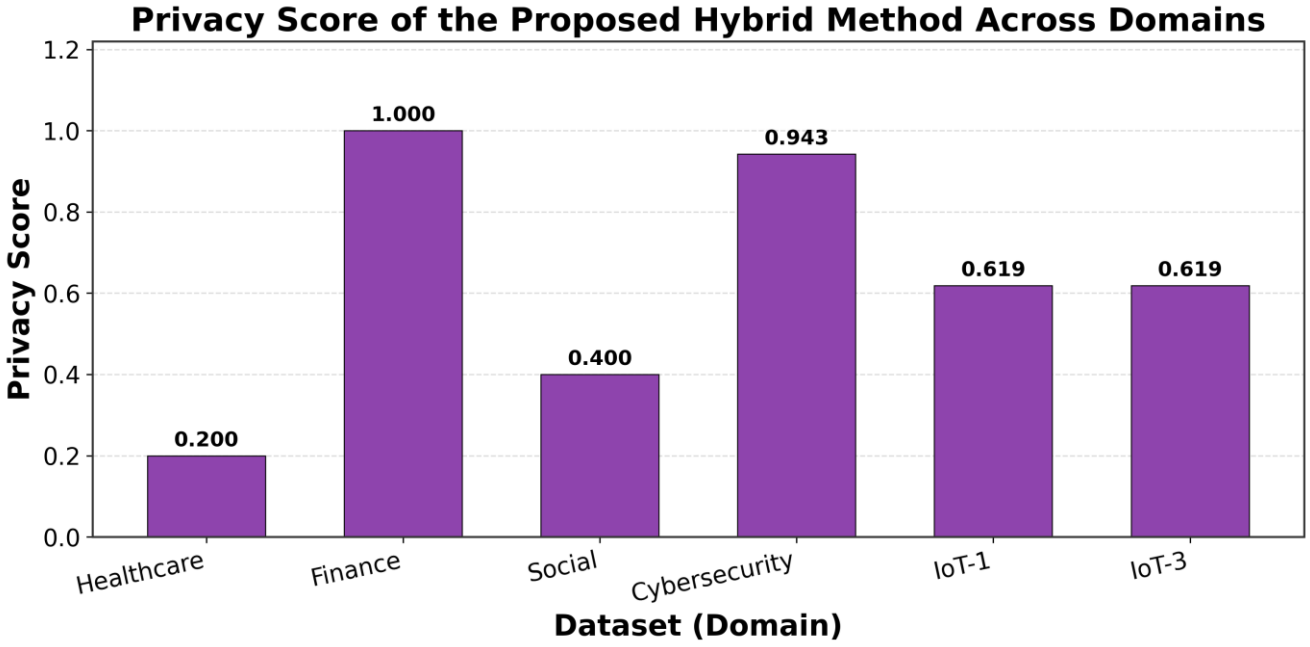

Fig 4: Privacy score across datasets.

The financial dataset achieves the highest privacy score (1.000) with negligible information loss, owing to its already anonymized features. The cybersecurity dataset also demonstrates strong privacy protection (0.943) with minimal information loss (0.058). In contrast, both IoT scenarios show moderate privacy levels (0.619) and higher information loss (≈ 0.300), reflecting the complexity of network data. The healthcare dataset achieves a lower privacy level (0.200), indicating limited transformation of attributes. The social dataset presents both high information loss (0.400) and a moderate privacy score (0.400), confirming that stronger transformations are required for text data, which in turn negatively impacts utility.

4.7 Execution Time Analysis

The computational overhead is shown in Figure 5.

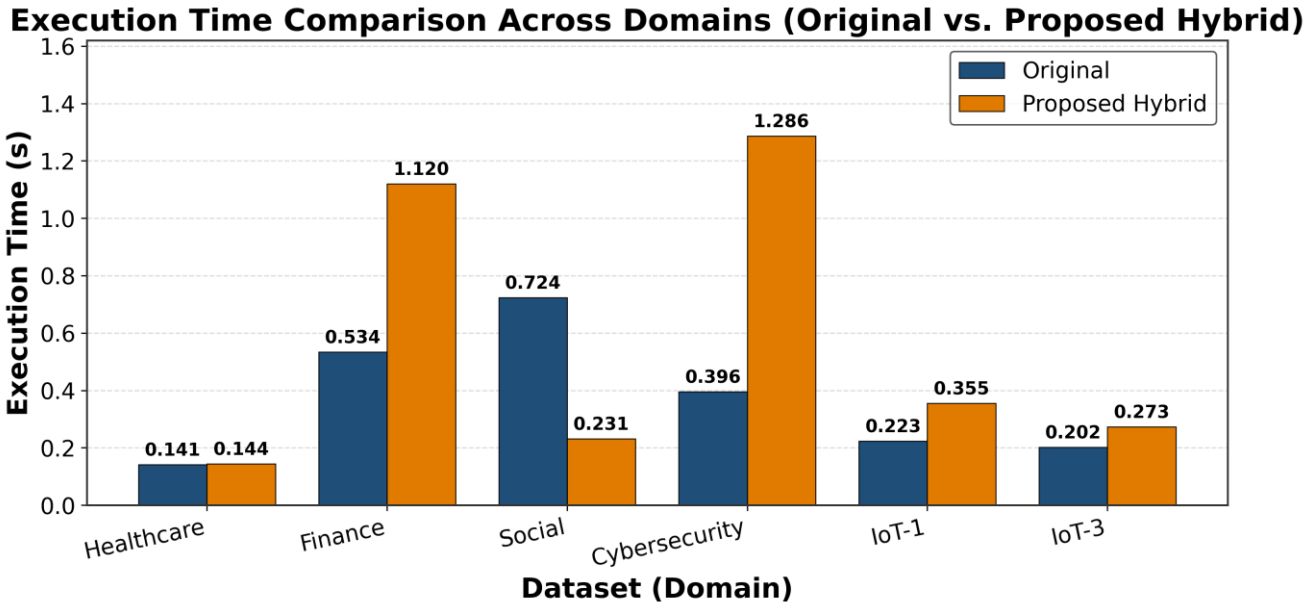

Fig 5: Execution time comparison across datasets.

The results indicate that the privacy-preserving process increases execution time, particularly in the financial and cybersecurity datasets. The overhead introduced by the transformation ranges from a negligible 1.02× in healthcare to 3.25× in cybersecurity, with the financial and IoT domains exhibiting overheads of 2.10× and 1.35–1.59×, respectively (see Table 3). Interestingly, the social dataset records a reduced execution time (0.32× of the original), which is attributed to the simplified feature representation produced by the aggressive generalization required for text data. In absolute terms, all execution times remain below 1.3 seconds, confirming that the framework is computationally lightweight and suitable for practical deployment.

4.8 Privacy–Utility Trade-off Analysis

To provide a more comprehensive evaluation, the relationship between privacy protection and predictive utility is examined quantitatively. For each domain, the utility retention is computed as the ratio between the performance of the privacy-preserved model and that of the original model, expressed as a percentage. In addition, a composite privacy–utility efficiency (PUE) measure is defined as the product of the privacy score and the F1-score retention, providing a single indicator that rewards mechanisms achieving strong privacy with minimal utility loss. Table 3 summarizes these derived measures across all evaluated domains.

Table 3. Derived Privacy–Utility Trade-off Measures Across Domains

Domain	Accuracy Retention (%)	F1 Retention (%)	Privacy Score	Information Loss	Time Overhead (×)	Privacy–Utility Efficiency
Healthcare	85.2	84.0	0.200	0.139	1.02	0.168
Finance	100.0	100.0	1.000	0.000	2.10	1.000
Social	75.6	50.2	0.400	0.400	0.32	0.201
Cybersecurity	98.4	98.2	0.943	0.058	3.25	0.926
IoT-1	99.8	99.8	0.619	0.300	1.59	0.618
IoT-3	100.0	100.0	0.619	0.300	1.35	0.619

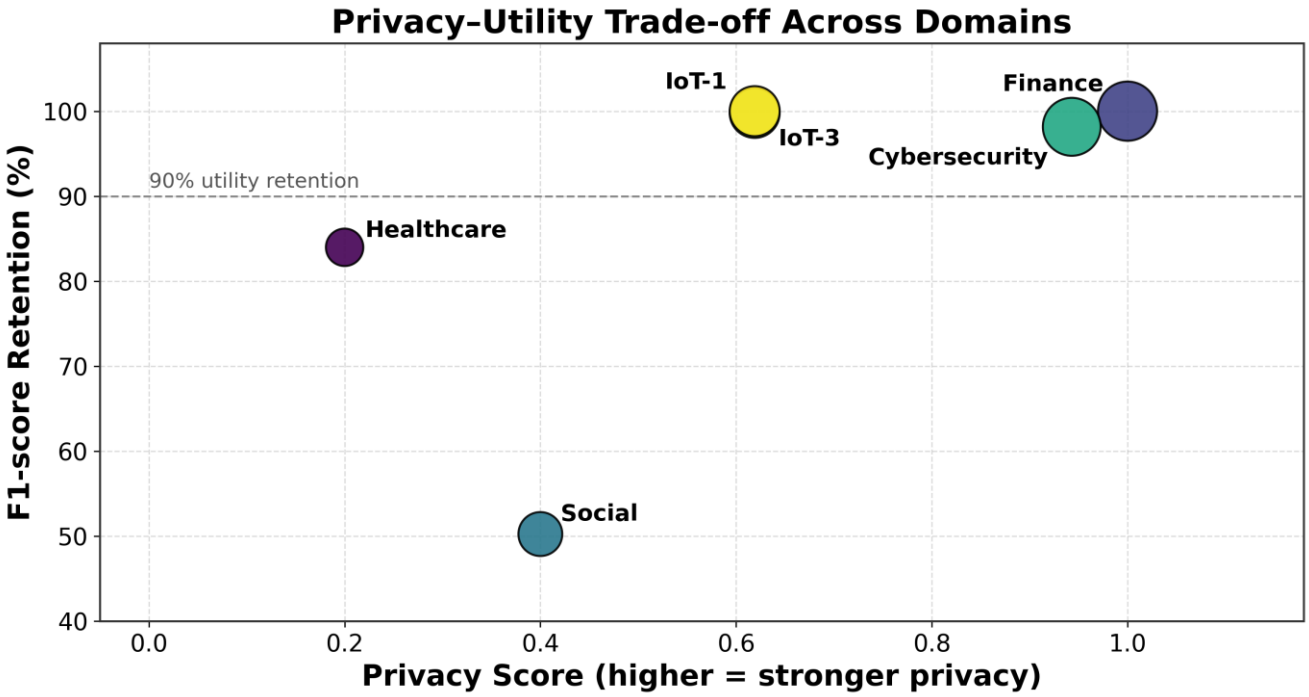

Fig 6: Privacy–utility trade-off across domains.

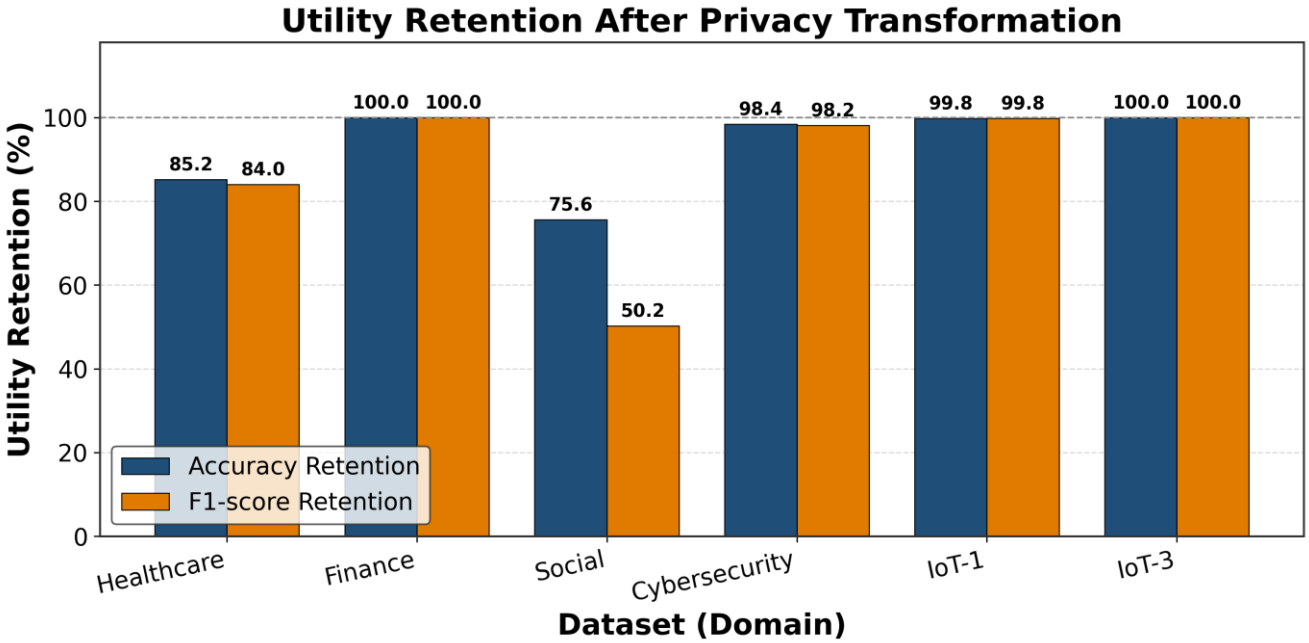

Fig 7: Utility retention after privacy transformation.

As illustrated in Figure 6, the domains form three distinct clusters in the privacy–utility space. The financial and cybersecurity domains occupy the most favorable region, combining high privacy scores (1.000 and 0.943, respectively) with utility retention above 98%. The two IoT scenarios form an intermediate cluster, retaining nearly all utility (≈ 99.8–100%) at a moderate privacy level (0.619). In contrast, the healthcare and, most notably, the social domains fall below the 90% utility-retention threshold, with the social domain retaining only 50.2% of its F1-score.

Figure 7 presents the accuracy and F1-score retention side by side, revealing that the gap between the two metrics is negligible for structured and network data but pronounced for the social domain, where the F1-score is far more sensitive to privacy transformation than accuracy. This indicates that, for imbalanced text-based data, accuracy alone can mask substantial degradation in minority-class performance, and that the F1-score is a more reliable indicator of utility under privacy constraints.

The privacy–utility efficiency values in Table 3 yield a clear overall ranking of the domains: Finance (1.000) > Cybersecurity (0.926) > IoT-3 (0.619) ≈ IoT-1 (0.618) > Social (0.201) > Healthcare (0.168). This ordering confirms that the same hybrid mechanism produces markedly different outcomes depending on data characteristics, and that strong privacy and high utility are simultaneously achievable only in specific

domains. The consistency of the two IoT scenarios further indicates that the observed behavior is stable across independent captures within the same domain rather than an artefact of a single dataset.

4.9 Key Findings

The experimental results reveal a consistent pattern across domains. Structured data, represented by the financial dataset, achieves both high utility and high privacy. Network data, represented by the cybersecurity and IoT datasets, achieves a balanced trade-off. Text data, represented by the social dataset, suffers significant performance degradation. These findings clearly demonstrate that the effectiveness of privacy-preserving mechanisms is strongly domain-dependent.

5. DISCUSSION

The results presented in Section 4 provide strong evidence that the effectiveness of privacy-preserving machine learning is not uniform across different data domains. This section interprets these findings and compares them with the approaches discussed in Section 2, highlighting both the strengths and limitations of the proposed hybrid framework.

The proposed method achieves high performance in structured and network-based datasets, particularly in the financial and cybersecurity domains. In the financial dataset, the model maintains near-perfect accuracy and F1-score after the privacy transformations are applied, indicating that the data representation is inherently robust to suppression, generalization, and perturbation. This observation is consistent with prior work on perturbation-based methods, which showed that the statistical properties of numerical data can be preserved under controlled noise injection [16], [17]. However, unlike these earlier studies, which focus on a single dataset, the current work demonstrates that this robustness is highly dependent on the domain.

When compared with anonymization-based techniques such as l-diversity and t-closeness [14], [15], the proposed hybrid framework provides a more favorable balance between privacy and utility. Traditional anonymization methods often result in significant information loss, particularly in high-dimensional datasets. In contrast, the results obtained in this study show minimal information loss in the financial and cybersecurity datasets while maintaining high predictive performance. This suggests that combining multiple techniques within a unified framework can overcome some of the limitations associated with individual methods.

The performance observed in the cybersecurity and IoT datasets further supports this conclusion. Recent studies have also highlighted the effectiveness of hybrid machine learning approaches in cybersecurity prediction tasks, particularly in vulnerability analysis and threat modeling [28]. In these domains, the proposed method achieves high accuracy with only minor degradation after the privacy transformation. Network traffic data typically consists of aggregated numerical features, which are less sensitive to perturbation and generalization. This behavior aligns with observations in intrusion detection research, where statistical features remain informative even after transformation [10]. At the same time, the moderate information loss observed in the IoT datasets indicates that some trade-off is unavoidable when sensitive attributes are protected. Such trade-offs are particularly relevant in wireless sensor and IoT deployments, where stringent energy and network-lifetime constraints further motivate lightweight privacy processing [29].

In contrast, the results obtained for the social dataset reveal a significant limitation of the proposed approach. The substantial drop in performance, particularly in the F1-score, indicates that text-based data is highly sensitive to privacy transformations. Unlike structured numerical data, textual data relies heavily on semantic relationships and feature representations. Transformations such as generalization and perturbation disrupt these relationships, leading to a loss of meaningful information. This issue is not adequately addressed in most existing studies, which tend to evaluate privacy mechanisms on structured datasets. Therefore, the results presented in this work extend previous findings by demonstrating that privacy-preserving techniques cannot be directly transferred to text-based domains without significant performance degradation.

Another important aspect highlighted by the results is the privacy–utility trade-off. In domains such as finance and cybersecurity, high privacy scores are achieved with minimal impact on model performance, indicating a favorable trade-off. This is particularly encouraging for data-intensive financial applications such as fraud detection and bank-stability prediction, where machine learning is increasingly applied to sensitive customer data [30]. However, in the social dataset, achieving even moderate privacy requires substantial modification of the data, resulting in a significant loss of utility. This observation is consistent with the general principle that stronger privacy guarantees often come at the cost of reduced model performance [8], but it also shows that the severity of this trade-off varies across domains.

From a practical perspective, these findings challenge the common assumption that a single privacy-preserving strategy can be applied across different applications. The results clearly demonstrate that the effectiveness of privacy mechanisms depends on data characteristics, including feature type, dimensionality, and representation. Structured numerical data is more resilient to transformation, whereas unstructured textual data is significantly more sensitive.

Overall, the discussion confirms that the proposed hybrid framework is effective in several domains but also exposes its limitations, particularly for text-based data. More importantly, it highlights a key insight that is not sufficiently addressed in the existing literature: privacy-preserving machine learning exhibits domain-dependent behavior. This insight provides a strong foundation for future work on adaptive and domain-aware privacy mechanisms.

Despite the promising results, this study has some limitations. The proposed framework relies on fixed privacy parameters, which may not be optimal for all datasets. In addition, the evaluation is limited to a single machine learning model (Random Forest), and results may vary with other models. Furthermore, the framework shows reduced effectiveness on text-based data, indicating the need for specialized approaches for unstructured data.

6. CONCLUSION

This paper presented a comprehensive cross-domain empirical study of privacy-preserving machine learning. A hybrid privacy framework combining suppression, generalization, and perturbation techniques was proposed and evaluated across six heterogeneous datasets spanning healthcare, finance, social media, cybersecurity, and IoT environments. The experimental results demonstrate that the proposed approach is effective in preserving predictive performance in structured and network-based datasets while achieving meaningful levels of privacy protection. In particular, the financial and cybersecurity datasets maintain high accuracy and F1-score even after the

privacy transformations are applied, indicating a favorable balance between privacy and utility.

The findings clearly demonstrate that privacy-preserving machine learning cannot be treated as a domain-independent solution, as its effectiveness varies significantly depending on data characteristics. In text-based social data, the application of the same privacy transformations leads to significant information loss and a substantial degradation in model performance. This highlights that privacy techniques cannot be uniformly applied across different data types without considering their underlying characteristics. The privacy–utility efficiency analysis quantifies this variation and shows that strong privacy and high utility are simultaneously achievable only in specific domains.

From a practical perspective, the proposed framework offers a simple and efficient solution that can be integrated into standard machine learning pipelines. At the same time, the results suggest that future systems should incorporate adaptive and domain-aware mechanisms capable of adjusting privacy parameters based on data characteristics. Future work will focus on extending the proposed framework by introducing adaptive privacy strategies, exploring domain-specific transformations, and integrating advanced techniques such as differential privacy and federated learning. In addition, further research is needed to develop effective privacy-preserving methods for text and other unstructured data types. In conclusion, this study highlights the importance of considering domain characteristics in the design of privacy-preserving machine learning systems and provides valuable insights for the development of more robust and context-aware privacy solutions.

7. REFERENCES

[1] Ali, Y. and Ali, L. (2026) ‘A Decision Support Framework for Meal Healthiness: Machine Learning Analysis with Expert Nutritional and Goal-Oriented Interpretation’.

[2] Shokri, R., Stronati, M., Song, C. and Shmatikov, V. (2017) ‘Membership inference attacks against machine learning models’, IEEE Symposium on Security and Privacy, pp. 3–18.

[3] Dwork, C. (2008) ‘Differential privacy: A survey of results’, International Conference on Theory and Applications of Models of Computation, pp. 1–19.

[4] Ali, L., Mathieu, H. and Biennier, F. (2006) ‘Monitoring and managing distributed networks using mobile agents’, Proceedings of the 2nd International Conference on Information and Communication Technologies (ICTTA). IEEE, vol. 2, pp. 3377–3382.

[5] Abadi, M., Chu, A., Goodfellow, I., McMahan, H.B., Mironov, I., Talwar, K. and Zhang, L. (2016) ‘Deep learning with differential privacy’, Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, pp. 308–318.

[6] Ali, L. (2008) Gestionnaire d’infrastructure distribuée. PhD thesis, Institut National des Sciences Appliquées de Lyon.

[7] Sweeney, L. (2002) ‘k-anonymity: A model for protecting privacy’, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10(5), pp. 557–570.

[8] Fung, B.C.M., Wang, K., Chen, R. and Yu, P.S. (2010) ‘Privacy-preserving data publishing: A survey of recent developments’, ACM Computing Surveys, 42(4), pp. 1–53.

[9] Dal Pozzolo, A., Caelen, O., Johnson, R.A. and Bontempi, G. (2015) ‘Calibrating probability with undersampling for unbalanced classification’, IEEE Symposium Series on Computational Intelligence, pp. 159–166.

[10] Moustafa, N. and Slay, J. (2015) ‘UNSW-NB15: A comprehensive data set for network intrusion detection systems’, Military Communications and Information Systems Conference (MilCIS), pp. 1–6.

[11] Ali, L., Jaber, M., Chaari, S. and Biennier, F. (2007) ‘Context-aware infrastructure to support distributed industrial services’, Proceedings of the IEEE International Conference on Emerging Technologies and Factory Automation (ETFA). IEEE, pp. 716–719.

[12] Geyer, R.C., Klein, T. and Nabi, M. (2017) ‘Differentially private federated learning: A client level perspective’, arXiv preprint arXiv:1712.07557.

[13] Beaulieu-Jones, B.K., Wu, Z.S., Williams, C., Lee, J., Bhavnani, S.P., Byrd, J.B. and Greene, C.S. (2019) ‘Privacy-preserving generative deep neural networks support clinical data sharing’, Circulation: Cardiovascular Quality and Outcomes, 12(7).

[14] Machanavajjhala, A., Kifer, D., Gehrke, J. and Venkitasubramaniam, M. (2007) ‘l-Diversity: Privacy beyond k-anonymity’, ACM Transactions on Knowledge Discovery from Data, 1(1), pp. 1–52.

[15] Li, N., Li, T. and Venkatasubramanian, S. (2007) ‘t-Closeness: Privacy beyond k-anonymity and l-diversity’, IEEE International Conference on Data Engineering, pp. 106–115.

[16] Agrawal, R. and Srikant, R. (2000) ‘Privacy-preserving data mining’, ACM SIGMOD International Conference on Management of Data, pp. 439–450.

[17] Kargupta, H., Datta, S., Wang, Q. and Sivakumar, K. (2003) ‘On the privacy preserving properties of random data perturbation techniques’, IEEE International Conference on Data Mining, pp. 99–106.

[18] McMahan, H.B., Moore, E., Ramage, D., Hampson, S. and Arcas, B.A. (2017) ‘Communication-efficient learning of deep networks from decentralized data’, Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, pp. 1273–1282.

[19] Acar, A., Aksu, H., Uluagac, A.S. and Conti, M. (2018) ‘A survey on homomorphic encryption schemes: Theory and implementation’, ACM Computing Surveys, 51(4), pp. 1–35.

[20] Medford, R.J., Saleh, S.N., Sumarsono, A., Perl, T.M. and Lehmann, C.U. (2020) ‘Anonymization and perturbation methods for healthcare data privacy’, Journal of Biomedical Informatics, 102.

[21] Ali, L. and Biennier, F. (2005) ‘Integration of security requirements in virtual enterprises’, APMS’05, Washington, United States.

[22] Alabdulatif, A., Khalil, I. and Yi, X. (2021) ‘Privacy-preserving data publishing: A survey on recent developments’, Journal of Network and Computer Applications, 190.

[23] Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H.R., Albarqouni, S. and Cardoso, M.J. (2020) ‘The future of digital health with federated learning’, npj Digital Medicine, 3(119).

[24] Veale, M., Binns, R. and Edwards, L. (2018) ‘Algorithms that remember: Model inversion attacks and data

protection law', Philosophical Transactions of the Royal Society A, 376.

[25] Dua, D. and Graff, C. (2019) UCI Machine Learning Repository. Available at: http://archive.ics.uci.edu/ml.

[26] Go, A., Bhayani, R. and Huang, L. (2009) Twitter Sentiment Classification using Distant Supervision. Stanford University.

[27] Garcia, S., Parmisano, A., Erquiaga, M.J. and Hofstede, R. (2020) The IoT-23 Dataset: A Labeled Dataset with Malicious and Benign IoT Network Traffic. Stratosphere Laboratory. Available at: https://www.stratosphereips.org/datasets-iot23.

[28] Issa, N., Gruska, D. and Ali, L. (2026) 'A hybrid GAM-based model for predicting vulnerability exploitation', in Cappiello, C., Hartig, O., Sellami, M. and Ouni, A. (eds.) Cooperative Information Systems. Lecture Notes in Computer Science, vol. 15535. Cham: Springer.

[29] Hajrulla, S., Ali, L. and Souliman, N. (2023) 'Normal distribution on energy saving problems for the wireless sensor network life on the vacation period', Journal of Natural Sciences & Mathematics (JNSM), 8.

[30] Farag, K., Ali, L., Mutai, N.C., Luqman, R., Mahmoud, A. and Krasniqi, N. (2025) 'Machine learning for predicting bank stability: The role of income diversification in European banking', FinTech, 4(2), 21. Available at: https://doi.org/10.3390/fintech4020021.