An Empirical Study of the Privacy-Utility Trade-off in Differentially Private Classifiers

Amar Kumar Mandal
Department of Administrative Sciences
Kadir Has University
Istanbul, Türkiye
amarkumar.mandal@stu.khas.edu.tr

S M Dedar Alam
Department of Administrative Sciences
Kadir Has University
Istanbul, Türkiye
sm.alam@stu.khas.edu.tr

Bimbo Lawrence Damitan
Department of Administrative Sciences
Kadir Has University
Istanbul, Türkiye
bimbo.damitan@stu.khas.edu.tr

Bisola Favour Adediji
Department of Administrative Sciences
Kadir Has University
Istanbul, Türkiye
bisolaadediji@stu.khas.edu.tr

Zivko Atanaskoski
Faculty of Computer Science and Engineering
Ss. Cyril and Methodius University
Skopje, North Macedonia
zivko.atanaskoski@finki.ukim.mk

Zorica Karapancheva
Faculty of Computer Science and Engineering
Ss. Cyril and Methodius University
Skopje, North Macedonia
zorica.karapancheva@finki.ukim.mk

Mila Dodevska
Faculty of Computer Science and Engineering
Ss. Cyril and Methodius University
Skopje, North Macedonia
mila.dodevska@finki.ukim.mk

Vesna Dimitrova
Faculty of Computer Science and Engineering
Ss. Cyril and Methodius University
Skopje, North Macedonia
vesna.dimitrova@finki.ukim.mk

***Abstract*—With the advancement of machine learning (ML) in healthcare, models trained on sensitive patient data face privacy risks. Traditional anonymization methods like k-anonymity are insufficient against adversaries with auxiliary knowledge, making Differential Privacy (DP) a necessary approach. While DP provides strong privacy guarantees, higher privacy levels can degrade predictive performance, which may compromise diagnostic reliability.**

This paper presents an empirical study of DP applied to classifiers on the Breast Cancer Wisconsin (Diagnostic) dataset across different privacy budgets. We compare standard (non-private) models with their differentially private counterparts. Results show that the DP random forest outperforms logistic regression by at least 15% under strong privacy ($\varepsilon < 1.0$), providing guidance for selecting an optimal privacy budget.

***Index Terms*—Differential Privacy, Privacy-Preserving Machine Learning, DP-SGD, Privacy-Utility Trade-off, Federated Learning, Random Forest, Logistic Regression**

I. Introduction

The integration of Machine Learning (ML) in high-stakes domains such as healthcare and finance has introduced significant privacy risks. While these models excel at identifying diagnostic patterns, they are susceptible to data leakage via attack vectors such as membership inference, model inversion, and gradient leakage [1], [2]. These vulnerabilities are exacerbated in third-party cloud environments where data owners lack direct control over the underlying infrastructure [3].

The core challenge lies in the tension between model utility and data confidentiality. Namely, accurate ML requires capturing specific patterns from individual records, yet without formal safeguards, the model cannot distinguish between generalizable trends and sensitive private information. This creates a critical privacy-utility trade-off. Traditional security measures often fail to address these ML-specific threats [1], leading researchers toward mathematically rigorous solutions like Differential Privacy (DP).

DP offers a formal framework to bound the impact of any single individual's data on the model's output [4]. By ensuring that the inclusion or exclusion of a specific record does not drastically alter the model's inference, DP provides a safeguard against skilled adversaries. The implementation of DP is governed by the privacy budget ε and smaller values of ε yield stronger privacy but introduce more noise, typically reducing model accuracy [5]–[7]. This trade-off is particularly sensitive in healthcare, where diagnostic reliability is as vital as patient privacy. Motivated by this, our study conducts an empirical analysis of DP-enabled classifiers on the Breast Cancer Wisconsin (Diagnostic) dataset. We compare standard non-private models against differentially private counterparts, specifically evaluating the resilience of Logistic Regression and Random Forest architectures across various privacy budgets. Our results highlight the performance thresholds required to maintain diagnostic utility while adhering to strict privacy guarantees.

The remainder of this paper is structured as follows: Section II reviews related work in privacy-preserving ML, while

Section III establishes the mathematical frameworks and threat models underpinning Differential Privacy. The experimental methodology, including dataset characteristics and the implementation of private classifiers, is detailed in Section IV. In Section V, we present our empirical findings and evaluate the impact of varying privacy budgets on model performance. Finally, Section VI concludes the paper with a discussion on the implications of our results and potential future research.

II. Related Work

The rapid evolution of deep learning has introduced profound privacy risks, with research demonstrating that model inference and inversion attacks can extract sensitive medical or biological information from training datasets. Traditional defenses like k-anonymity and homomorphic encryption often struggle against differential attacks, necessitating the mathematically rigorous framework of Differential Privacy. By injecting controlled noise into datasets or query results, DP provides a formal barrier against information leakage [8]. However, the transition of DP from theory to critical domains like healthcare and finance requires a systematic evaluation of how this noise impacts decision-making and the intrinsic privacy-utility trade-off [9]–[11].

Healthcare is a primary beneficiary of DP due to the extreme sensitivity of patient records. Research into clinical applications has highlighted a significant utility cost. For instance, Suriyakumar et al. [12] observed a 35% drop in AUROC for mortality prediction under strict privacy budgets, while Rahman et al. [13] noted a decline in accuracy from 90% to 75% as ε was tuned from 1.0 to 0.1 in federated settings. Beyond direct model training, DP has been integrated into synthetic data generation via frameworks like PATE-GAN [14] and used to secure individual healthcare queries through Laplace noise and value rounding [15]. Notably, large-scale studies demonstrate that DP can achieve performance nearly comparable to centralized models even when protecting millions of patient records in federated environments [16], [17].

The application of DP extends into the financial sector for tasks such as fraud detection and credit scoring. Byrd and Polychroniadou [18] demonstrated that combining secure multi-party computation (SMPC) with DP allows models to maintain utility similar to non-private baselines at $\varepsilon \geq 5$. Similarly, hybrid models in healthcare finance have shown that while excessive noise increases Type-II errors in fraud detection, a privacy budget between 1.0 and 10.0 offers a functional balance for risk scoring [19]. This sensitivity to noise distribution is further supported by governmental applications, such as the US Census, where adopting discrete Gaussian mechanisms reduced variance by half compared to Laplacian noise, thereby increasing utility for the same level of privacy [20].

Despite these advancements, a pervasive challenge remains the "utility cost" inherent in formal privacy guarantees. DP has been shown to amplify data imbalance, disproportionately reducing accuracy for minority classes [21]. Furthermore, several studies argue that many practical DP implementations rely on high epsilon values ($\varepsilon > 14$) that may offer a false sense of privacy, rather than true protection [22], [23]. To bridge this gap, recent work has focused on architectural adaptations and improved privacy accounting via Rényi Differential Privacy (RDP) [24]. The consensus across current literature suggests that DP cannot be implemented as a "closed-box" solution. It requires transparent, systematic evaluation of how noise injection alters critical outcomes [12], [18], [23].

III. Theoretical Foundations

This section establishes the mathematical framework of Differential Privacy (DP), characterizes the noise mechanisms employed in our study, and describes the underlying privacy-utility trade-off in machine learning.

A. Differential Privacy (DP) Formalism

Differential Privacy provides a probabilistic guarantee that the output of a randomized algorithm $\mathcal{M}$ remains statistically consistent regardless of the inclusion of any single individual's record. Formally, for two neighboring datasets D and D' differing by at most one element, $\mathcal{M}$ satisfies (ε, δ)-DP if for all output subsets $S \subseteq \text{Range}(\mathcal{M})$ [4]:

$$P[\mathcal{M}(D) \in S] \leq e^{\varepsilon} \cdot P[\mathcal{M}(D') \in S] + \delta \tag{1}$$

The privacy budget ε quantifies the information leakage, i.e. smaller values represent a narrower bound and thus stronger privacy. The parameter δ represents the probability of a privacy breach, typically set to be cryptographically small ($\delta \ll 1/|D|$).

B. Global Sensitivity and Noise Mechanisms

The core of DP implementation is the calibration of noise to the global sensitivity Δf. Sensitivity measures the maximum change the presence of a single record can induce in the function's output across any two neighboring datasets:

$$\Delta f = \max_{D,D'} \|f(D) - f(D')\|_p \tag{2}$$

Depending on the model architecture and data type, we utilize different mechanisms to achieve the guarantee in Eq. 1:

- **Laplace Mechanism:** Suitable for real-valued outputs (e.g., Logistic Regression coefficients), adding noise sampled from $\text{Lap}(\Delta f/\varepsilon)$. It ensures pure ε-DP [4].
- **Gaussian Mechanism and DP-SGD:** To satisfy (ε, δ)-DP in iterative training, the Gaussian mechanism is integrated into the optimization process via Differentially Private Stochastic Gradient Descent (DP-SGD). In this approach, per-sample gradients are clipped to a maximum L_2 norm C to bound sensitivity, and Gaussian noise is added to the resulting average gradient before the weight update [25].
- **Exponential Mechanism:** Employed for non-numeric selection, such as determining optimal split-points in

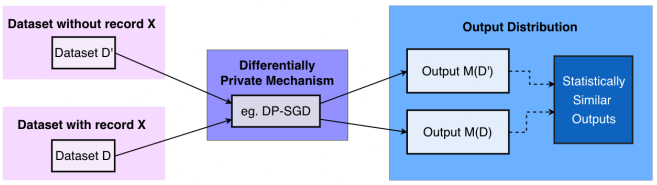

Fig. 1. Illustration of differential privacy

Random Forests. It selects outputs based on a utility score u, where the probability of selection is proportional to $\exp\left(\frac{\varepsilon u}{2\Delta u}\right)$ [8].

C. Threat Model: Membership Inference Attack (MIA)

While this study focuses on the empirical performance of DP models, the implementation of DP is motivated by the risk of Membership Inference Attacks. In an MIA scenario, an adversary attempts to determine if a specific patient's data was used to train a diagnostic model by observing prediction confidence or output labels. DP provides a formal defense against this threat by ensuring that the model's parameters do not over-index on individual records, maintaining the "indistinguishability" required for patient confidentiality in healthcare.

D. The Privacy-Utility Trade-off

The introduction of DP creates an inherent utility cost, governed by the "Fundamental Law of Information Recovery" [4]. In supervised learning, this is primarily driven by:

- **Noise Variance:** As ε decreases, the increased noise variance can obscure subtle diagnostic features, leading to lower classification accuracy.
- **Gradient Clipping:** To bound sensitivity in deep learning, gradients must be clipped. This can bias the model, potentially disproportionately affecting minority classes or outliers in medical datasets [21].
- **Advanced Accounting:** Repeated access to data accumulates privacy loss. Techniques like **Rényi DP (RDP)** are used to provide tighter bounds for this accumulation than standard composition, allowing for higher utility under the same ε target [24].

IV. Methodology

This section outlines the unified experimental pipeline developed to evaluate the privacy-utility trade-off across distinct model architectures, focusing on the *Breast Cancer Wisconsin (Diagnostic)* dataset [26].

A. Experimental Overview

The primary objective of our methodology is to quantify the performance degradation of classification algorithms when subjected to formal privacy constraints. We use two primary algorithms—*Logistic Regression* (LR) and *Random Forest* (RF)—to analyze how linear versus ensemble-based models respond to noise injection. The pipeline follows a four-stage process: dataset preprocessing, model training with DP integration, budget-varied evaluation, and comparative trade-off analysis.

B. Dataset and Preprocessing

The Breast Cancer Wisconsin (Diagnostic) dataset serves as our benchmark. It contains 569 instances with 30 numeric features. To ensure the mathematical integrity of our DP mechanisms, we performed the following:

1) **Feature Engineering:** Non-essential identifiers were removed, and the target diagnosis was encoded as Malignant (1) and Benign (0).
2) **Global Scaling:** We applied `StandardScaler` to achieve $\mu = 0$ and $\sigma = 1$. This is a critical step in DP. Without scaling, high-magnitude features would dominate the sensitivity Δf, requiring excessive noise that would render the model unusable.
3) **Reproducibility:** An 80/20 train-test split was applied using a fixed random seed (42).

C. Models and Algorithms

We compare standard (non-private) baselines against two private variants:

- **Standard LR and RF:** These models establish the performance ceiling ($\varepsilon = \infty$).
- **DP-LR (Objective Perturbation):** Privacy is achieved by modifying the objective function with controlled noise.
- **DP-RF:** Implemented via `diffprivlib`, this model uses the exponential mechanism for split selection and output perturbation for tree structures, providing pure ε-DP [27].

D. Evaluation Metrics

Beyond traditional metrics (Accuracy, Precision, Recall, F1), we utilize **Accuracy Loss (ACL)** [7] to quantify the "cost of privacy":

$$\text{ACL} = 1 - \frac{\text{Accuracy}(\mathcal{M}, \varepsilon)}{\text{Accuracy}(\mathcal{M}, \varepsilon = \infty)} \tag{3}$$

By normalizing the performance drop relative to the non-private baseline, the ACL serves as a standardized metric to quantify the specific 'utility tax' incurred by the injection of differential privacy noise.

V. Experiments and Results

Our assessment evaluates the resilience of classifiers as the privacy budget ε is restricted from a relaxed state ($\varepsilon = 10.0$) to a high-privacy mandate ($\varepsilon = 0.1$). To establish a definitive performance ceiling, we first evaluated the classification accuracy of the Random Forest and Logistic Regression models in a non-private setting. These baseline results, presented in Table I, represent the models' performance with no DP applied

Non-private models achieve near-perfect metrics due to the high linear separability of the Wisconsin dataset, establishing

TABLE I
BASELINE PERFORMANCE OF NON-PRIVATE MODELS

Model	Accuracy	Precision	F1	Recall
Standard RF	0.97	1.00	0.96	0.92
Standard LR	0.96	0.97	0.96	0.92

TABLE II
COMPARATIVE UTILITY PERFORMANCE. DP-RF VS. DP-LR

Epsilon (ε)	**DP-RF**			**DP-LR**		
	ACL	F1	Recall	ACL	F1	Recall
0.1	0.132	0.765	0.738	0.354	0.616	0.632
0.4	0.114	0.790	0.762	0.337	0.636	0.656
0.8	**0.079**	0.831	0.762	0.301	0.668	0.684
1.0	0.105	0.789	0.714	0.263	0.704	0.721
2.0	0.088	0.821	0.762	0.263	0.704	0.721
10.0	0.105	0.784	0.690	**0.127**	0.832	0.835

a utility "upper bound". However, this fidelity risks memorizing patient outliers, increasing vulnerability to membership inference attacks and necessitating a privacy-utility trade-off.

While the baseline models establish the upper bound of diagnostic performance, they offer no protection against membership inference attacks. To quantify the resilience of these classifiers under formal privacy constraints, we subjected both architectures to a range of privacy budgets (ε). Table II details the resulting performance degradation, capturing the "utility tax" required to ensure that individual patient records remain mathematically indistinguishable within the training set.

The empirical data in Table II reveals a stark contrast in architectural resilience. At the highest privacy setting ($\varepsilon = 0.1$), the Accuracy Loss (ACL) for Logistic Regression is nearly three times greater than that of the Random Forest, suggesting that the objective perturbation in linear models significantly warps the decision boundary. Conversely, the ensemble nature of the Random Forest provides a structural buffer; by utilizing a majority-voting scheme across privatized trees, it maintains a 15% utility advantage over its linear counterpart. The introduction of DP mechanisms induces a "utility tax" across all architectures. As shown in Table II, the utility gap is most pronounced in the high-privacy regime ($\varepsilon = 0.1$). Here, DP-LR accuracy suffers significantly, with an Accuracy Loss (ACL) of 0.354. In contrast, DP-RF maintains an F1-score of 0.765, suggesting that ensemble models are inherently better at isolating diagnostic signals from noise.

A critical observation in Fig. 2 is the non-linear utility recovery as ε increases. DP-RF reaches its "inflection point" at $\varepsilon = 0.8$, where it achieves its lowest ACL of 0.079. Beyond $\varepsilon = 2.0$, DP-RF enters a saturation phase where further budget relaxation yields diminishing returns. Conversely, DP-LR shows a linear recovery but requires a budget of $\varepsilon = 10.0$ to reach the utility benchmarks that DP-RF achieves at much stricter levels.

Figures 3 and 4 visually contrast the performance of standard models against their differentially private counterparts at a privacy budget of $\varepsilon = 1.0$. As illustrated in Fig. 3, Logistic Regression experiences a drastic reduction in Precision and F1-score, indicating that the linear decision boundary is highly sensitive to the injected noise. Conversely, Fig. 4 shows that the Random Forest maintains significantly higher Precision

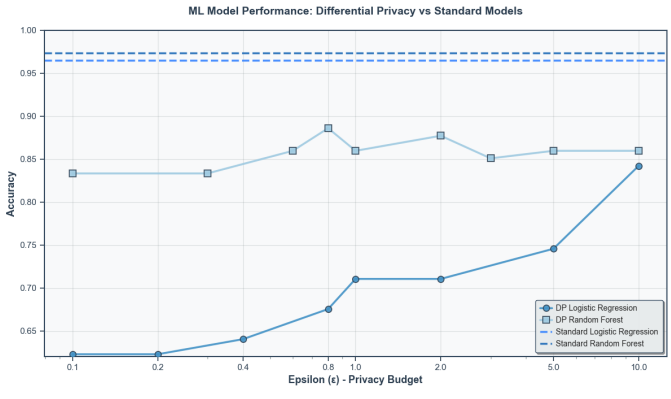

Fig. 2. Impact of Privacy Budget (ε) on Model Accuracy, highlighting the saturation point of ensemble models.

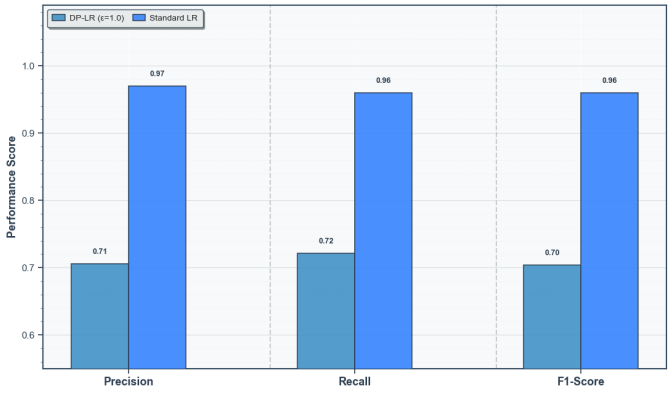

Fig. 3. Standard vs DP Logistic Regression ($\varepsilon = 1.0$)

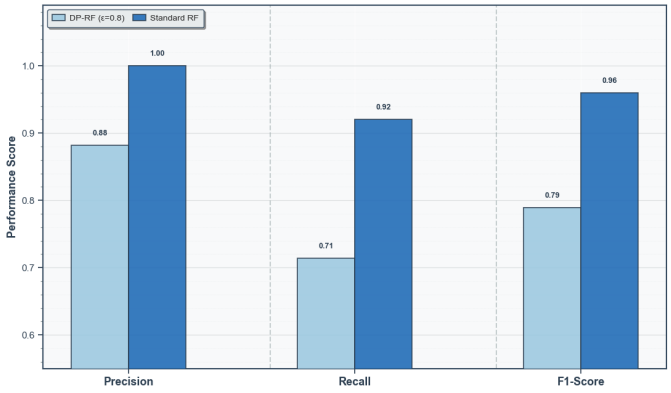

Fig. 4. Standard vs DP Random Forest ($\varepsilon = 1.0$)

(0.89) despite the privacy constraints. This suggests that the ensemble architecture of the Random Forest is more resilient to the blurring effect of Differential Privacy, preserving diagnostic utility more effectively than the linear model.

VI. Discussion, Conclusion and Future Work

A. Architectural Resilience and Noise Impact

Our empirical findings suggest that model architecture is as critical as the privacy budget (ε) in determining the feasibility of AI-driven diagnostics. The consistent 15% utility gap observed between DP-RF and DP-LR indicates that non-linear ensemble methods are structurally more resilient to the noise injection required for formal privacy. This resilience likely stems from the "output perturbation" mechanism in the DP-RF implementation; by applying the exponential mechanism to select split points, the model maintains the hierarchical structure of the data more effectively than the "objective perturbation" used in DP-LR, which directly warps the linear decision boundary.

While these findings have relevant clinical implications for balancing false positives and missed diagnoses, the technical evidence suggests that ensemble models aggregate multiple "weak learners" in a way that allows the fundamental clinical signal to survive perturbation at individual decision nodes.

B. Data Sensitivity and Scaling Limitations

A significant constraint identified in this study is the relationship between dataset size and the "privacy tax." With a sample size of $N = 569$, each individual record in the Wisconsin dataset possesses a relatively high impact on the model, leading to higher global sensitivity (Δf). As ε decreases, the magnitude of the Laplacian or Gaussian noise must increase to mask these influential data points. While larger datasets might "dilute" this sensitivity and potentially narrow the Accuracy Loss (ACL), our results prove that even at smaller scales, selecting a robust architecture like Random Forest can mitigate the worst effects of privacy-preserving noise.

C. Conclusion

This paper quantified the privacy-utility trade-off for breast cancer diagnosis using a standardized benchmark. We conclude that DP-RF is the superior model for privacy-preserving diagnostic tasks, outperforming DP-LR by 15% under strict privacy mandates where $\varepsilon \leq 1.0$. We have demonstrated that while Differential Privacy (DP) inherently introduces a utility degradation, the severity of this "tax" is not uniform across architectures. Based on the identified inflection points where accuracy begins to stabilize, we recommend a privacy budget range of $\varepsilon \in [0.8, 2.0]$ to achieve an optimal balance between patient confidentiality and predictive reliability.

D. Future Work

Building on these conclusions, future research will pursue three primary technical avenues:

1) **Adaptive Budget Allocation:** We intend to explore mechanisms where the privacy budget is dynamically allocated based on feature importance. By spending more of the ε budget on "high-signal" features (e.g., tumor thickness) and less on auxiliary data, the utility gap could be further reduced.
2) **Multi-modal DP Frameworks:** Future studies will extend this evaluation to multi-modal datasets that combine tabular FNA features with raw medical imagery. This will test the scalability of ensemble-like resilience within high-dimensional deep learning architectures.
3) **Hybrid Privacy Protocols:** We aim to investigate the integration of Differential Privacy with Federated Learning. This would facilitate multi-institutional model training where data remains on local servers, providing a dual layer of protection through decentralization and formal privacy guarantees.

Acknowledgment

This work was partially supported by the Faculty of Computer Science and Engineering at Ss. Cyril and Methodius University in Skopje, as well as the European Union through the ERASMUS MUNDUS framework under the CyberMACS project (Project #101082683, https://cybermacs.eu/).

References

[1] F. Khalid, M. A. Hanif, S. Rehman, and M. Shafique, "Security for machine learning-based systems: Attacks and challenges during training and inference," in *Proceedings of the International Conference on Frontiers of Information Technology (FIT)*, 2018, pp. 327–332.

[2] W. Wei and L. Liu, "Gradient leakage attack resilient deep learning," *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 303–316, 2022.

[3] R. Gupta and A. K. Singh, "A differential approach for data and classification service-based privacy-preserving machine learning model in cloud environment," *New Generation Computing*, vol. 40, no. 3, pp. 737–764, 2022.

[4] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Foundations and Trends in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–487, 2013.

[5] X. Tang, L. Zhu, M. Shen, and X. Du, "When homomorphic cryptosystem meets differential privacy: Training machine learning classifier with privacy protection," 2018, arXiv:1812.02292.

[6] K. Kuźniewski, K. Matusiewicz, and P. Sapiecha, "The high-level practical overview of open-source privacy-preserving machine learning solutions," *International Journal of Electronics and Telecommunications*, 2022.

[7] B. Zhao, M. A. Kaafar, and N. Kourtellis, "Not one but many tradeoffs: Privacy vs. utility in differentially private machine learning," in *ACM SIGSAC Conference on Cloud Computing Security Workshop*, 2020, pp. 15–26.

[8] Z. Shen and T. Zhong, "Analysis of application examples of differential privacy in deep learning," *Security and Communication Networks*, 2021.

[9] N. Ponomareva *et al.*, "How to dp-fy ml: A practical guide to machine learning with differential privacy," *Journal of Artificial Intelligence Research*, vol. 77, pp. 1113–1201, 2023.

[10] S. Wassan, L. Liudajun, H. Ying, H. Dongyan, and P. Fei, "Federated learning and differential privacy: Machine learning and deep learning for biomedical image data classification," *Digital Health*, vol. 11, 2025.

[11] D. Byrd and A. Polychroniadou, "Differentially private secure multi-party computation for federated learning in financial applications," 2020. [Online]. Available: https://arxiv.org/abs/2010.05867

[12] V. M. Suriyakumar, N. Papernot, A. Goldenberg, and M. Ghassemi, "Chasing your long tails: Differentially private prediction in health care settings," in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 2021, pp. 723–734.

[13] M. M. Rahman, M. S. Yeasmin, L. Kabir, T. Abedin, M. H. Uddin, M. Mahmud, A. Rahman, M. Nur-E-Alam, and M. Rokonuzzaman, "Balancing privacy and performance: Federated learning with differential privacy for real-time, resilient healthcare ai," *Diyala Journal of Engineering Sciences*, pp. 1–21, 2025.

[14] M. Giuffrè and D. L. Shung, "Harnessing the power of synthetic data in healthcare: innovation, application, and privacy," *NPJ digital medicine*, vol. 6, no. 1, p. 186, 2023.

[15] R. Subramanian, "Differential privacy techniques for healthcare data," in *2022 International Conference on Intelligent Data Science Technologies and Applications (IDSTA)*. IEEE, 2022, pp. 95–100.

[16] O. Choudhury, A. Gkoulalas-Divanis, T. Salonidis, I. Sylla, Y. Park, G. Hsu, and A. Das, "Differential privacy-enabled federated learning for sensitive health data," *arXiv preprint arXiv:1910.02578*, 2019.

[17] A. Sadilek, L. Liu, D. Nguyen, M. Kamruzzaman, S. Serghiou, B. Rader, A. Ingerman, S. Mellem, P. Kairouz, E. O. Nsoesie *et al.*, "Privacy-first health research with federated learning," *NPJ digital medicine*, vol. 4, no. 1, p. 132, 2021.

[18] D. Byrd and A. Polychroniadou, "Differentially private secure multi-party computation for federated learning in financial applications," in *Proceedings of the first ACM international conference on AI in finance*, 2020, pp. 1–9.

[19] J. Ara, M. Roy, and S. H. Swarnali, "Differential privacy and federated learning for secure predictive modeling in healthcare finance," *International Journal of Research and Innovation in Applied Science*, vol. 10, no. 9, pp. 745–759, 2025.

[20] C. L. Canonne, G. Kamath, and T. Steinke, "The discrete gaussian for differential privacy," *Advances in Neural Information Processing Systems*, vol. 33, pp. 15 676–15 688, 2020.

[21] E. Bagdasaryan, O. Poursaeed, and V. Shmatikov, "Differential privacy has disparate impact on model accuracy," *Advances in neural information processing systems*, vol. 32, 2019.

[22] A. Blanco-Justicia, D. Sánchez, J. Domingo-Ferrer, and K. Muralidhar, "A critical review on the use (and misuse) of differential privacy in machine learning," *ACM Computing Surveys*, vol. 55, no. 8, pp. 1–16, 2022.

[23] J. Domingo-Ferrer, D. Sánchez, and A. Blanco-Justicia, "The limits of differential privacy (and its misuse in data release and machine learning)," *Communications of the ACM*, vol. 64, no. 7, pp. 33–35, 2021.

[24] I. Mironov, "Renyi differential privacy," in *IEEE Computer Security Foundations Symposium*, 2017.

[25] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, 2016, pp. 308–318.

[26] W. Wolberg, O. Mangasarian, N. Street, and W. Street, "Breast Cancer Wisconsin (Diagnostic)," UCI Machine Learning Repository, 1993, dOI: 10.24432/C5DW2B.

[27] S. Fletcher and M. Z. Islam, "Differentially private random decision forests using smooth sensitivity," *Expert systems with applications*, vol. 78, pp. 16–31, 2017.