User-Level Differential Privacy in Medical Machine Learning

Johannes Kaiser[1,*], Jakob Eigenmann[1,*], Daniel Rueckert[1,3, 4], and Georgios Kaissis[1,2]

[1]Chair for AI in Healthcare and Medicine, Technical University of Munich (TUM) and TUM University Hospital, Munich, Germany
[2]Institute for Machine Learning in Biomedical Imaging, Helmholtz Munich, Neuherberg, Germany
[3]Munich Center for Machine Learning (MCML), Munich, Germany
[4]Department of Computing, Imperial College London, UK
[*]Equal Contribution

Abstract

We address the challenge of ensuring user-level DP when individuals contribute *varying* numbers of data records to a dataset. While group privacy can be used to aggregate record-level budgets, it can be overly pessimistic and lacks flexibility when users contribute varying numbers of data points. We propose a method for accounting for arbitrary numbers of records per user while maintaining a *fixed per-user privacy guarantee* by leveraging individual privacy assignment. Experimentally, our method yields excellent utility comparable to record-level DP while providing a more meaningful/interpretable protection.

1 Introduction

DP [1] is designed to protect individuals by ensuring that the output of a randomized algorithm remains almost indistinguishable when applied to two neighboring databases. A paradigm is commonly considered in which every user's data corresponds to a single database row. However, in many domains, such as medical imaging, a single user's data can correspond to one or more rows, that is, more than one record. These records can come from different imaging modalities (e.g. radiography and magnetic resonance imaging) from follow-up scans of the same patient, different views, etc. Then, individuals contributing more data are at significantly higher privacy risk than those with fewer entries.

In this setting, the record-centric paradigm of protecting individual database rows is ill-fitting and must be translated to the user level by aggregating the row-level privacy budgets, usually through group privacy (or, equivalently, through adjusting the global sensitivity). To circumvent this complexity, many studies applying differential privacy to medical data adhere to the record-level formulation, that is, safeguarding individual *images* rather than the membership of the actual user/patient in the dataset [2–4]. However, protecting individual records does not translate to the same protection of individuals, as users contributing more data are at significantly higher privacy risk than users contributing less.

This misalignment highlights the need for a shift in focus: DP loses some of its interpretability when applied on the record rather than the user level. Moreover, prior research indicates that the privacy requirements of users are overwhelmingly based on a user-centric privacy perception and not a record-centric one [5–7].

In this work, we propose using *individual DP assignment* for enabling user-level DP (ULDP) machine learning. Our approach has a key benefit compared to group-privacy-based approaches, allowing us to easily work with variable numbers of records per patient. We demonstrate that on CheXpert [8] – a popular benchmark medical imaging dataset, our approach drastically outperforms a naïve record-level privacy implementation that is then extended to the user level *ex-post*. Our approach, moreover, achieves classification performance that is competitive with record-level privacy (that can be aggregated with group privacy to establish weaker user-level guarantees) while providing equivalent protection for all individuals.

2 Individual-DP based ULDP

2.1 User-level DP

To shorten definitions, WLOG, we will assume $D' \subsetneq D$ as necessary in the following. The definition of (ε, δ)-ULDP is equal to (ε, δ)-DP (approximate DP), except for the concept of neighboring datasets. Contrary to record-level neighboring datasets which differ in exactly one record, we define two datasets as user-level neighboring, iff $\exists u \in U_D : D \setminus D' = D_u$ with U_D being all users that contributed to D and $D_u \subseteq D$ being all records of user u in D. Intuitively, this means that two datasets are neighboring if they differ in exactly the contribution of one user, regardless of how many samples the user has contributed. This guarantees that the algorithmic results are indistinguishable whether any *user* is present or not.

The need for ULDP is easily recognizable in the context of federated learning (FL), where data is intended to be kept private on a client-level basis and regular DP algorithms are not easily applicable [9–11]. Lately, there has also been some work on ULDP for centralized learning [12, 13], with some mimicking FL [14], and others proposing different clipping and sampling-based approaches [15–17]. Our work proposes a new idea of achieving ULDP by adapting individual DP methods.

2.2 Individual DP

Individual DP (iDP) extends DP by furnishing individual privacy guarantees for each user as opposed to a dataset-level "blanket" guarantee.

Definition 2.1 (Individual Approximate DP, [18]). A randomized mechanism $\mathcal{M}$ satisfies $(\varepsilon_i, \delta_i)$ - individual (approximate) DP or $(\varepsilon_i, \delta_i)$-iDP if, for all pairs of neighboring databases D, D', with $D' \in \{D^{-i}, D^{+i}, D^{\sim i}\}$ (with $-i$, $+i$, $\sim i$ indicating remove, add and replace operations w.r.t. any database element z_i), all sets $S \subseteq \text{Range}(\mathcal{M})$ it holds that:

$$\Pr\left[\mathcal{M}(D) \in S\right] \leq \exp\left(\varepsilon_i\right) \Pr\left[\mathcal{M}(D') \in S\right] + \delta_i. \quad (1)$$

The study of iDP recognizes two distinct domains: **individual privacy accounting**, which entails tracking the *realized privacy budget* for each data point individually [18–20], and **individual privacy assignment**, which allows different users to choose "bespoke" privacy budgets. Our method leverages individual privacy assignments to render the contribution of a user/patient independent of the number of their contributed samples/records (i.e. medical images). Existing approaches to individual privacy assignment can be categorized into **sensitivity-based** and **subsampling-based** methods.

In **sensitivity-based approaches**, the amount of noise required to ensure DP depends on the algorithm's per-instance sensitivity to the underlying data. These methods thus bound the sensitivity of an algorithm for each instance (e.g. each record or each user) individually. This technique was first applied in [21], by using a *stretching mechanism* that scales individual data elements with a transformation matrix to adjust the sensitivity.

Sampling-based approaches [22] extend [23] to use user-specific sampling rates to train machine learning models with per-user privacy budgets by leveraging the subsampling theorem [24, 25]. Even though originally defined on groups sharing the same privacy budget, the method reduces to individual privacy when the group size is one.

2.3 Methods

For clarity, we will denote user-specific quantities with a superscript ($x^{(j)}$) and record-specific quantities with a subscript (x_i). We consider the setting of training a machine learning model on the data of N unique users, with each user contributing up to K images. The cardinality of the set of images contributed by the j^{th} user will be denoted $k^{(j)}$. The training of the machine learning model should satisfy DP with respect to some adjacency relation on the user database. For now, we will assume an equal target privacy budget for all users, but subsequently, we want to enable users to choose their specific budget based on their individual privacy perception. For simplicity, we will specify the target *user-level* privacy budget in terms of approximate DP, i.e. as an $(\widetilde{\varepsilon}, \widetilde{\delta})$-tuple. Although it is possible to assign a user-specific $(\widetilde{\varepsilon}^{(j)}, \widetilde{\delta}^{(j)})$, we assume the same user-specific budgets for all users for simplicity. Moreover, sharper accounting can be achieved by specifying the budget in terms of a privacy curve or collection of RDP budgets [26–28]. Recall that our aim is to design a technique that (1) allows users to contribute an arbitrary number of samples and (2) accounts precisely for the actual number of samples rather than an upper bound on the samples (K in the notation above).

Naïve User-Level Privacy We can naïvely achieve fixed user-level protection by defining a maximal number of contributed records per user $\bar{K} \leq K$ and splitting the privacy budget per record to yield $(\widetilde{\varepsilon}, \widetilde{\delta})$ with group privacy. This approach has two major problems, as (1) limiting the data elements per user to $\bar{K}$ removes $k^{(j)} - \bar{K}$ data elements of any user that provides more than $\bar{K}$ elements and (2) does not fully exploit the $(\widetilde{\varepsilon}, \widetilde{\delta})$ budget for users who contribute fewer than $\bar{K}$ records, and is thus usually suboptimal. Note that with $\bar{K} = 1$, this naïve approach reduces to "enforcing" record-level DP by discarding all but one record per user. This trivial strategy only achieves desideratum (1) above, and only in a crude way, i.e. by upper-bounding the number of records contributed per user.

Thus, we require a method that is able to also satisfy achieve desideratum (2) above. We will achieve this by (intuitively) "splitting" the user-level budget into $k^{(j)}$ record-level privacy sub-budgets $(\varepsilon_i^{(j)}, \delta_i^{(j)})$ using the group privacy property. Note again that we express this strategy in terms of approximate DP, but better bounds can be achieved through employing an alternative DP definition such as RDP or f-DP [29]. Alternatively, one can also calibrate to an operational privacy risk rather than a privacy budget as shown in [30]. Overall, we assign to each record a corresponding singular privacy sub-budget, which, when aggregated, yields a per-user target privacy budget. Training a neural network with this method is a $(\widetilde{\varepsilon}, \widetilde{\delta})$-user-level-DP algorithm.

We propose two strategies to achieve the aforementioned aim by leveraging individual privacy assignment.

Strategy 1: Sampling Adaptation Our first strategy directly interferes with the Poisson sampling process of DP-SGD. Instead of assigning equal sampling probabilities to every record, we increase the sampling rates of users with fewer records and decrease the sampling rates of users with higher amounts of records. The weighted average of the individual sampling rates is kept the same as for DP-SGD.

In practice, this is done by the *sampling with constant batch size* technique [22] which determines the record-level sampling probabilities that yield a desired batch size in expectation. In brief, this technique numerically or analytically determines the appropriate Poisson sampling probability q to yield an expected batch size of T and moreover determines the appropriate record-sampling probabilities $\{q^{(j)}\}_{j=1,\ldots,N}$ such that the resulting algorithm guarantees $\left(\varepsilon_i^{(j)}, \delta_i^{(j)}\right)$-record-level individual DP. The technique ensures that the algorithm satisfies $(\widetilde{\varepsilon}, \widetilde{\delta})$-user-level DP through the group privacy property.

Strategy 2: Clipping Adaptation In the neural network training setting described above, we now aim to equalize the *contribution* of each user (measured as an upper bound on the total ℓ_2-norm of their gradients over the duration of training). We sample from all of the users' records with probability $T/\sum_{j=1}^{N} k^{(j)}$. Now, a batch of T records is expected to contain more records from users who contributed more and fewer from users who contributed fewer records. Next, we specify a gradient ℓ_2 clipping bound per user that is inversely proportional to the number of records they hold. Thus, the overall contribution of any user is bounded by an ℓ_2 "norm budget". To determine the appropriate ℓ_2 gradient clipping bounds and Gaussian noise scale, we employ the *scale* technique of [22]. This technique numerically or analytically determines the appropriate Gaussian noise scale σ and the appropriate ℓ_2 clipping bounds $\{c^{(j)}\}_{j=1,\dots,N}$ such that the resulting algorithm satisfies $(\widetilde{\varepsilon}, \widetilde{\delta})$-user-level DP as above.

Note that with both strategies, if large discrepancies between the number of records per user exist (e.g. some users contribute tenfold the number of records compared to other users), it may still be beneficial to define a "cut-off" $\bar{K}$ for the maximum number of admissible records per user as in the naïve approach above. Otherwise, with a user contribution of many records, the user-level privacy budget gets fragmented – potentially reducing the overall extractable information.

We note that our current implementation technique of both strategies relies on Rényi DP for accounting and uses the approximate DP group privacy property. This has some limitations which will be addressed in future work: (1) It requires that certain technical conditions be met for analytically solving the aforementioned optimization problems (see [31]); (2) it relies on conversions between approximate DP and Rényi DP, which are imperfect [27]); (3) it uses the approximate DP group privacy property rather than the Rényi DP group privacy property.

Moreover, note that both aforementioned strategies are *data dependent* and we consider the number of records per user a public quantity. We believe that considering the number of samples per user to be public constitutes an acceptable privacy risk in most cases. Alternatively, we propose a data-independent alternative, which however has apparent weaknesses.

Alternative: User-Level Accounting As above, we train a neural network with Poisson-sampled DP-SGD and account for each user's varying number of records via Rényi group privacy and the use of a *privacy filter* [18, 32] for fully-adaptive composition. After the user-level privacy is reached *all* records of that user are removed (filtered) from training. This allows us to condition on the number of records per user while considering it a private quantity. However, filtering often leads to catastrophic forgetting of the dropped data [19]. Consequently, this strategy does not lead to sufficient utility of the neural network or underexploited privacy budgets (when early stopping) and is thus not explored further.

3 Empirical Evaluation

Dataset and Architecture We evaluate our method on CheXpert, a multi-label medical image classification dataset of 224 316 chest X-rays from 65 240 individuals with imbalanced labels extracted from radiology reports. Following [2], we restrict focus to the classes *Atelectasis, Cardiomegaly, Consolidation, Edema, and Pleural Effusion*. The dataset's number of records per user exhibits a somewhat heavy-tailed distribution with record counts between 1 and 92 per user, see Figure 1. This reflects the typical structure of medical data with consecutive visits, subsequent scans and multiple views of the same patient (frontal, lateral). We evaluate our method using a ResNet-18 [33] trained for a fixed number of iterations, ensuring all models receive the same number of gradient updates.

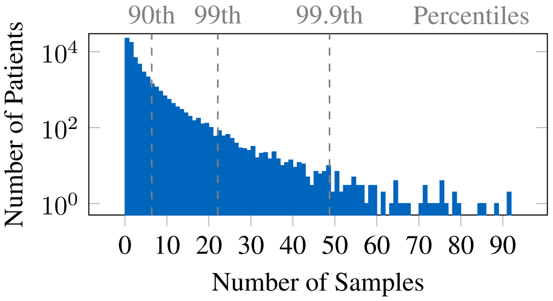

Figure 1: Histogram of the number of records per user/patient and corresponding percentiles of the whole dataset. Note the somewhat heavy-tailed nature of the distribution.

Comparison of Proposed Strategies Figure 2 contains results for the naïve and for the iDP-based user-level privacy and compares them to a non-private and a record-level DP baseline. As expected, all private algorithms perform worse in terms of ROC-AUC than the non-private baseline. For the naïve approach, with an increase in $\bar{K}$, the privacy budget allotted to **each** individual record decreases and the privacy budget is not fully exploited as discussed above. Increasing $\bar{K}$ means that more samples per user contribute to the model's training. Evidently, this does not compensate for most users' overly strict privacy guarantee, and consequently, the model's performance suffers.

Our proposed iDP-based strategies ensure that for any $\bar{K}$, the model satisfies $(\widetilde{\varepsilon}, \widetilde{\delta})$-DP for all users. As exemplified in Figure 2, the resulting models are comparable in performance to record-level DP whose per-record guarantees, which are dubiously meaningful and, when

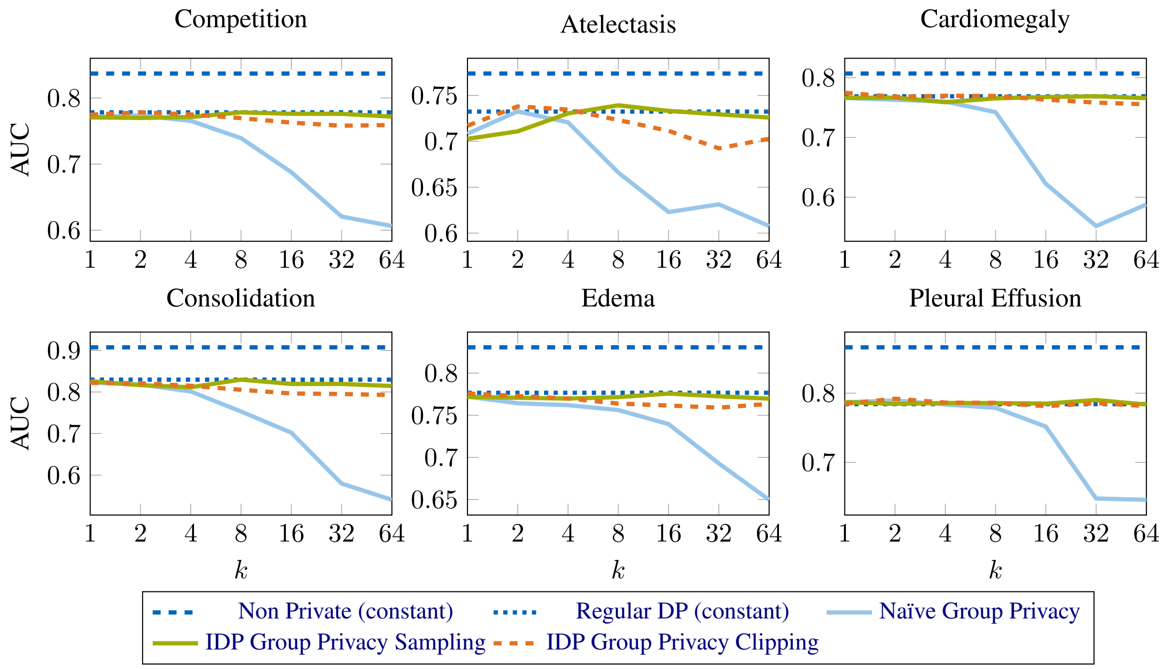

Figure 2: Empirical evaluation of the Naïve Group Privacy and IDP Group Privacy using Clipping and Sampling for 5 averaged runs of a ResNet-18 trained for 10 000 iterations with batch size 64 on CheXpert data with a user-level privacy budget $\widetilde{\varepsilon} = 8$ and user level $\delta = 10^{-5}$. *Competition* indicates the average across the AUC scores of all investigated classes. For higher $\bar{K}$, our proposed strategies outperform the naive approach, achieving ROC-AUC comparable to record-level DP while maintaining more meaningful user-level privacy guarantees.

accumulated across a user's records, yield far higher privacy risks. There is no substantial difference in performance between strategy 1 (Sampling Adaptation) and strategy 2 (Clipping adaptation). Moreover, the empirical performance of the iDP-based user-level privacy strategy is largely independent of $\bar{K}$. This provides more flexibility and reduces the risk of privacy budget fragmentation. In summary, our proposed iDP-based user-level privacy strategy allows training machine learning models while (1) incorporating *all available data* and (2) providing a $(\widetilde{\varepsilon}, \widetilde{\delta})$-*user-level-DP guarantee*.

4 Discussion and Conclusion

We propose iDP-based ULDP for machine learning settings where individuals contribute multiple records to a dataset, but it is desired to express privacy guarantees in terms of user-level DP. Our proposed method utilizes all available data while maintaining the flexibility to specify individual user-level privacy budgets.

Empirical results on the CheXpert dataset demonstrate that our approach achieves performance comparable to less private, non-user-centric, record-level DP. Moreover, it outperforms a naïve strategy to ULDP machine learning that entails removing data and suboptimal exploitation of the user-level privacy budget. Therefore, this indicates the feasibility of strong user-level privacy guarantees without compromising model performance.

We anticipate even larger improvements on datasets with higher inter-user variability, where the model can better benefit from the additional data diversity. Therefore, we plan to extend our experiments to more datasets such as MIMIC-CXR and BreakHis [34, 35] and FL scenarios.

Further, we aim to extend our work by comparing our proposed strategies to another ULDP baseline [14].

For simplicity, we trained with a uniform user-level privacy budget across users. However, our method naturally extends to individualized user-level privacy requirements. This flexibility is crucial, as prior research suggests that privacy perceptions vary across individuals, necessitating different privacy guarantees for different demographics [5–7]. By allowing contributors to control the number of records and their desired overall privacy level, our method facilitates truly person-centric privacy while maintaining seamless integration into existing DP machine learning frameworks.

Future research will extend our work into a DP active learning setting similar to [36], where data is labelled and added to the dataset solely when it is beneficial to training and investigating an informed distribution of the user-level privacy budget across a user's records [14].

In conclusion, we hope that our work serves as a first step toward designing DP machine learning from a more user-centric perspective.

References

[1] Cynthia Dwork. "Differential Privacy". In: *Automata, Languages and Programming*. Berlin, Heidelberg: Springer, 2006, pp. 1–12.

[2] Leonard Berrada et al. *Unlocking Accuracy and Fairness in Differentially Private Image Classification*. 2023. arXiv: 2308.10888.

[3] Zelun Luo et al. “Differentially Private Video Activity Recognition”. In: *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. 2024, pp. 6643–6653.

[4] Soroosh Tayebi Arasteh et al. “Securing Collaborative Medical AI by Using Differential Privacy: Domain Transfer for Classification of Chest Radiographs”. In: *Radiology: Artificial Intelligence* 6.1 (2023).

[5] Carlos Jensen, Colin Potts, and Christian Jensen. “Privacy Practices of Internet Users: Self-reports versus Observed Behavior”. In: *International Journal of Human-Computer Studies* 63.1 (2005), pp. 203–227.

[6] Bettina Berendt, Oliver Günther, and Sarah Spiekermann. *Privacy in E-Commerce: Stated Preferences vs. Actual Behavior**. SSRN Scholarly Paper. Rochester, NY, 2005.

[7] Humphrey Taylor. *Harris Interactive | The Harris Poll - Most People Are*. https://www.harrisinteractives.com/harris_poll/index-PID-365.html. 2003.

[8] Jeremy Irvin, Pranav Rajpurkar, and et al. “CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison”. In: *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence*. AAAI Press, 2019, pp. 590–597.

[9] Zhibo Wang et al. “Beyond Inferring Class Representatives: User-Level Privacy Leakage From Federated Learning”. In: *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*. Paris, France: IEEE Press, Apr. 2019, pp. 2512–2520.

[10] H. Brendan McMahan et al. *A General Approach to Adding Differential Privacy to Iterative Training Procedures*. Mar. 2019.

[11] Peter Kairouz et al. “Practical and Private (Deep) Learning Without Sampling or Shuffling”. In: *Proceedings of the 38th International Conference on Machine Learning*. PMLR, July 2021, pp. 5213–5225.

[12] Daniel Asher Nathan Levy et al. “Learning with User-Level Privacy”. In: *Advances in Neural Information Processing Systems*. 2021.

[13] Yuhan Liu et al. *Learning Discrete Distributions: User vs Item-Level Privacy*. 2021.

[14] Lynn Chua et al. *Mind the Privacy Unit! User-Level Differential Privacy for Language Model Fine-Tuning*. 2024.

[15] Badih Ghazi, Ravi Kumar, and Pasin Manurangsi. “User-Level Differentially Private Learning via Correlated Sampling”. In: *Advances in Neural Information Processing Systems*. Vol. 34. Curran Associates, Inc., 2021, pp. 20172–20184.

[16] Zachary Charles et al. *Fine-Tuning Large Language Models with User-Level Differential Privacy*. 2024.

[17] Daogao Liu and Hilal Asi. “User-Level Differentially Private Stochastic Convex Optimization: Efficient Algorithms with Optimal Rates”. In: *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*. PMLR, 2024, pp. 4240–4248.

[18] Vitaly Feldman and Tijana Zrnic. “Individual Privacy Accounting via a Rényi Filter”. In: *Advances in Neural Information Processing Systems*. Vol. 34. Curran Associates, Inc., 2021, pp. 28080–28091.

[19] Antti Koskela, Marlon Tobaben, and Antti Honkela. *Individual Privacy Accounting with Gaussian Differential Privacy*. 2022. arXiv: 2209.15596.

[20] Da Yu et al. *Individual Privacy Accounting for Differentially Private Stochastic Gradient Descent*. 2023. arXiv: 2206.02617.

[21] Mohammad Alaggan, Sébastien Gambs, and Anne-Marie Kermarrec. “Heterogeneous differential privacy”. In: *Journal of Privacy and Confidentiality* (2016).

[22] Franziska Boenisch et al. “Have It Your Way: Individualized Privacy Assignment for DP-SGD”. In: *Advances in Neural Information Processing Systems* 36 (2023), pp. 19073–19103.

[23] Zach Jorgensen, Ting Yu, and Graham Cormode. “Conservative or liberal? Personalized differential privacy”. In: *2015 IEEE 31St international conference on data engineering*. IEEE. 2015, pp. 1023–1034.

[24] Shiva Prasad Kasiviswanathan et al. “What can we learn privately?” In: *SIAM Journal on Computing* 40.3 (2011), pp. 793–826.

[25] Martin Abadi et al. “Deep learning with differential privacy”. In: *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*. 2016, pp. 308–318.

[26] Yu-Xiang Wang, Borja Balle, and Shiva Prasad Kasiviswanathan. “Subsampled rényi differential privacy and analytical moments accountant”. In: *The 22nd international conference on artificial intelligence and statistics*. PMLR. 2019, pp. 1226–1235.

[27] Borja Balle et al. *Hypothesis Testing Interpretations and Renyi Differential Privacy*. 2019. arXiv: 1905.09982.

[28] Georgios Kaissis et al. *Beyond the calibration point: mechanism comparison in differential privacy*. 2024. arXiv: 2406.08918.

[29] Jinshuo Dong, Aaron Roth, and Weijie J Su. "Gaussian differential privacy". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 84.1 (2022), pp. 3–37.

[30] Bogdan Kulynych et al. "Attack-aware noise calibration for differential privacy". In: *Advances in Neural Information Processing Systems* 37 (2024), pp. 134868–134901.

[31] Ilya Mironov, Kunal Talwar, and Li Zhang. *Rényi Differential Privacy of the Sampled Gaussian Mechanism*. 2019. arXiv: 1908.10530.

[32] Mathias Lécuyer. *Practical Privacy Filters and Odometers with Rényi Differential Privacy and Applications to Differentially Private Deep Learning*. 2021. arXiv: 2103.01379.

[33] Kaiming He et al. "Deep Residual Learning for Image Recognition". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 770–778.

[34] Alistair EW Johnson et al. "MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports". In: *Scientific data* 6.1 (2019), p. 317.

[35] Yassir Benhammou et al. "BreakHis based breast cancer automatic diagnosis using deep learning: Taxonomy, survey and insights". In: *Neurocomputing* 375 (2020), pp. 9–24.

[36] Kristian Schwethelm et al. *Differentially Private Active Learning: Balancing Effective Data Selection and Privacy*. 2025. arXiv: 2410.00542.