CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	HEALTH-DP: a framework for health data de-anonymization risk assessment and mitigation with differential privacy
Authors	Aguelal, Hamza;Shafiq, Akasha;Palmieri, Paolo
Publication date	2026
Original Citation	Aguelal, H., Shafiq, A. and Palmieri, P. (2026) 'HEALTH-DP: a framework for health data de-anonymization risk assessment and mitigation with differential privacy', Proceedings of the 12th International Conference on Information Systems Security and Privacy (ICISSP 2026), Marbella, Spain, 4-6 March, Volume 1, pp. 201-212.
Type of publication	Conference item
Rights	© 2026, SCITEPRESS – Science and Technology Publications, Lda. - https://creativecommons.org/licenses/by-nc-nd/4.0/
Download date	2026-10-05 11:48:24
Item downloaded from	https://hdl.handle.net/10468/18568

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

HEALTH-DP: A Framework for Health Data De-Anonymization Risk Assessment and Mitigation with Differential Privacy

Hamza Aguelal*[a], Akasha Shafiq*[b] and Paolo Palmieri[c]
Computer Science & IT, University College Cork, Cork, Ireland
{*h.aguelal, a.shafiq, p.palmieri*}*@cs.ucc.ie*

Keywords: De-Anonymization, Differential Privacy, Anonymization, Risk Assessment, Health Data, Privacy Enhancement Technologies.

Abstract: Privacy protection is a significant challenge in the computation of personal data, especially when data (e.g. health-related) is considered sensitive under relevant regulations. Although anonymization is widely applied, adversaries can still de-anonymize data through sophisticated attacks. Risks are particularly severe for health datasets, such as genomics or physiological data, due to their inherent uniqueness. Differential privacy (DP) has emerged as a strong privacy-preservation technique. However, current approaches to its implementation remain theoretical (and thus not directly linked to actual risks) or specific to a single context, and lack inclusive pathways for different stakeholders in the medical environment. This paper presents a comprehensive framework to address these limitations, combining a systematic study of re-identification attacks and practical risk assessment with DP implementation. The framework incorporates the parties' roles, threat pre-assessment, known attacks and DP integration. An adaptive mitigation strategy within a structured flow and logical process ensures wide coverage of different requirements. Furthermore, we validate the framework by applying central DP (CDP) to a heart-attack prediction dataset as an initial case study for a future broader end-to-end implementation. The framework provides a roadmap for implementing DP based on evaluating re-identification risks and data governance requirements, and gives stakeholders actionable guidance for safer data use.

1 INTRODUCTION

Recent pandemics and the rapid generation of healthcare data highlight the need for large-scale collection, sharing, and robust privacy. Healthcare systems are experiencing advancements in clinical research, quality improvement, and public reporting. All while ensuring patients' data privacy (records, signals, and images), relying on Privacy Enhancement Technologies (PET), particularly Anonymization. However, the traditional techniques (e.g identifier removal) or classical anonymity are susceptible to attacks, specifically de-anonymization, using available auxiliary information, latent identifiers and data linkage. Differential Privacy (DP) shields against de-anonymization by adding random noise to data, and recent work demonstrates application across multiple settings, query and noisy databases publishing, synthetic data generation, and privacy-preserving learning. Regardless of this progress, the transition from DP as a mechanism to a practice in healthcare remains challenging.

A persistent gap exists in prior contributions, focusing on a specific DP implementation, such as Laplace, Gaussian, or Differentially Private Stochastic Gradient Descent (DPSGD), within a mechanism-first and narrow context. While surveys and case studies combine implementation and challenges, they fall short of deploying processes and lack a general risk-driven, attack-pre-analysis with threat modelling integration. We put forward our work that considers those aspects. To the best of our knowledge, the first study to offer an inclusive privacy framework that scales to different modalities and DP outputs, as we highlight in our exploratory use case, is discussed further in the paper. In a related context, studies focused on specific modalities, such as Electrocardiogram (ECG) (Ghazarian et al., 2024) provided a high-level recipe for that specific type but miss a comprehensive architecture that addresses de-anonymization combined with guidance on DP in health. Existing work provides little or no guidance on selecting the appropriate DP approach when datasets differ in character-

[a] https://orcid.org/0009-0002-9409-4206
[b] https://orcid.org/0000-0002-2133-3547
[c] https://orcid.org/0000-0002-9819-4880
*These authors contributed equally.

Aguelal, H., Shafiq, A. and Palmieri, P.
HEALTH-DP: A Framework for Health Data De-Anonymization Risk Assessment and Mitigation with Differential Privacy.
Paper published under CC license (CC BY-NC-ND 4.0)
In *Proceedings of the 12th International Conference on Information Systems Security and Privacy (ICISSP 2026) - Volume 1*, pages 201-212
ISBN: 978-989-758-800-6; ISSN: 2184-4356
Proceedings Copyright © 2026 by SCITEPRESS – Science and Technology Publications, Lda.

istics and use cases and on integrating DP within a healthcare structure to match real-life requirements. This paper addresses these limitations by presenting a risk-based, governance-aware DP framework. It aligns with NIST guidance (Gary Howarth, 2025) that provides practitioners with the DP's prerequisites to adopt our proposed workflow. In contrast to the previous works, we cover the risk analysis, sensitivity and pipeline the flow into the DP strategies to mitigate against health de-anonymization. Our study, guided by pre-assessments, threat analysis, and separation of stakeholders' roles, couples the flow of deployment of DP aligning with institutional governance, detailed in section 4. Building on the re-identification literature, we further develop a basic proof-of-concept application using medical records. This practical example is based on selecting central DP (CDP) as the suitable method of noise addition and publishing aggregate statistics for this type of physiological information. In summary, our main contributions are:

- Introduce a practical privacy framework for healthcare, including real-world constraints beyond theoretical DP.
- Link de-anonymization risk assessment and DP mitigation considering the specific data nature and exposure, to inform privacy-flows choices.
- Bridge stakeholder perspectives by defining roles, responsibilities, and mapping communication.
- Demonstrate the framework using a baseline case study on health records data, assessing CDP deployment for aggregate statistics.

The remaining sections are arranged as follows: section 2 covers the background knowledge and related works to our framework. section 3 outlines the de-anonymization landscape and DP as mitigation, including a mapping between the two. section 4 details the HEALTH-DP framework alongside its conception, and presents our case study. Comparison and discussion in section 5 for the key insights and highlight strengths and challenges linked with our proposition. Finally, the conclusion 6 summarizes contributions and future research directions.

2 BACKGROUND AND RELATED WORKS

Health data is exposed to privacy breaches, with potential substantial regulatory consequences (e.g., GDPR (European Union, 2016)). Data protection is required, particularly anonymization: removing or modifying sensitive information so that personal data is non-identifiable. Evidence shows that anonymized data face threats (Newaz et al., 2021). Furthermore, the removal or aggregation of personally identifiable information (PII) or identifiers (IDs) (e.g. name), as well as the classical anonymization methods (Zhou et al., 2008) like k-anonymity, l-diversity, t-closeness, were not sufficient in practice (Fung et al., 2010), as shown in the literature (Narayanan and Shmatikov, 2006; Barbaro et al., 2006).

De-anonymization relies on exploiting the anonymization failure, using modern analytics and data availability. The risk arises when pattern inference in PII or QIDs (quasi-identifiers: non-unique IDs, e.g. ZIP code) is used. Systematic reviews (Al-Azizy et al., 2016; Henriksen-Bulmer and Jeary, 2016) catalogued a wide range of attempts to confirm threats on different modalities. However, the healthcare represents a unique and sensitive case, and features richness making it naturally identifiable; as a result, the re-identifiability of data is a factual threat, as detailed in reviews (Emam et al., 2011), (Aguelal and Palmieri, 2025) that point out vulnerabilities and group de-anonymization attacks, with proven risks in many works (e.g. $>95\%$ (S Shringar-pure, 2015)). While regulations recommend anonymization, the state of the art proves that the traditional approach cannot reliably prevent the risks in practice. This motivates our work to advocate for advanced privacy models, specifically DP, which we elaborate on in the subsequent sections.

2.1 Differential Privacy (DP)

We adopt standard (ε)-DP and its common relaxation (ε,δ)-DP to bound worst-case disclosure when implementing DP. Further, we discuss deployment models:
Differential Privacy. Any randomized algorithm M is said to be differentially private if for two dataset D and D' differ only in a single record, and S is the subset of all possible outputs of M (Dwork, 2006)

$$\frac{Pr[M(D)\varepsilon S]}{Pr[M(D')\varepsilon S]} \leq e^{\varepsilon}$$

M ensures the output invariance of the dataset by adding or removing a person. The probability of a change in the output is bounded by e^{ε}. Another DP variant is (ε,δ)-DP, which represents the information leakage as δ. This value is set as a tiny number, usually $\frac{1}{N}$(size of the D)

$$Pr[M(D)\varepsilon S] \leq e^{\varepsilon} Pr[M(D')\varepsilon S] + \delta$$

There are other relaxed variants of DP such as $(\varepsilon,\delta,\gamma)$-DP (Hall et al., 2013), Reny differnetial privacy (RDP) (Mironov, 2017).

- Central differential privacy (CDP): Users send data to a trusted aggregator to analyze and publish the noisy results. The noise is added to the output of the original query, such as

$$\text{Query} = \text{True result} + \text{Noise}$$

- Local differential privacy (LDP): Each individual adds noise to the data before sending it to an untrusted aggregator to aggregate data.

Sensitivity defines the maximum amount of effect on the query's output on the addition or removal of an individual from a dataset.

Privacy Budget ε regulates the amount of noise. A higher value of ε indicates less noise and more utility, whereas a smaller value designates more noise and less utility, providing higher privacy. The privacy-utility trade-off refers to the balance between the two concepts.

Noise Mechanism. Some of the most commonly used mechanisms are as follows:

- Randomized Response: Utilizes a coin flip to perturb each individual's data. This approach works best for surveys (Wang et al., 2016).
- Laplace Mechanism: Most commonly used to introduce noise into data. Drawn from a Laplace distribution with a mean of 0 and sensitivity $\Delta s/\varepsilon$ to add to a query's Q response (Dwork et al., 2006). When using ε-DP, the noise is introduced and only applies to numerical data.
- Gaussian Mechanism: Based on noise-injection drawn from a Gaussian distribution, achieves (ε, δ)-DP for numerical data. Scaled from a mean of 0 according to the Δs. δ defines the information leakage probability (Liu, 2018). DP libraries can be implemented in various settings. [1]

2.2 Related Work

Multiple studies have explored DP in healthcare from diverse perspectives. However, a smaller set offers operational guidance for deploying DP (beyond purely theoretical surveys). The study in (Moqurrab et al., 2023) proposed an e-health framework serving as a utility-aware anonymized publishing, but lacks concrete de-anonymization. In parallel, (Dyda et al., 2021) integrated DP within a COVID-19 reporting platform, demonstrating the feasibility of DP deployment. Additionally, (Lyu et al., 2025) proposed an edge-based DP framework (DPSDG), but both approaches lack scalability in their evaluations. Zhang and Li (Zhang and Li, 2022) propose a publishing pipeline, but as a single-mode publisher without governance or risk. In wearable devices, (Saifuzzaman et al., 2022) reviews DP applications, and (Liu et al., 2024) provides a broader survey of DP in medical data mining and challenges. Whereas (Subramanian, 2022) analyzed mechanisms' impact on data utility, provided results, but similarly lacks general guidance. In the same category, we find (Zia et al., 2020) that examined DP randomization in medical datasets, limited to Laplace noise. For the physiological domain, (Ghazarian et al., 2024) implemented CDP on ECG, focusing on a single modality with no threat assessment or pre-evaluation.

Most studies remain modality/mechanism-specific architectures, rarely connecting empirical attacks and directed flow to DP parameter selection. However, comprehensive reviews, deployment, and open challenges studies served as a baseline composition. Our work introduces a deployment-friendly framework that operationalizes when, where, and how to apply DP across different healthcare modalities with de-anonymization considerations. We combine an attack-prioritized approach with a role-based flow for selecting suitable actions.

3 HEALTHCARE DE-ANONYMIZATION AND DIFFERENTIAL PRIVACY

Our study incorporates the de-anonymization risks and DP as protection encompassing: Data types, adversarial capabilities and DP models. We synthesize attacks and elaborate on DP mitigation, not as an exhaustive taxonomy, but as a foundation for HEALTH-DP framework. We translate re-identification risk into role-aware choices for when, where and how to apply DP and ground it in the threat patterns.

3.1 Modalities and Vulnerabilities

Medical data is considered one of the most challenging for anonymization due to its rich personal nature and availability. The health sector is also a prime target for adversaries (IBM Security, 2023). The literature indicates that attacks often exploit modality-specific vulnerabilities; therefore, the risk varies. We highlight the substantial data-specific vulnerabilities:

- **Genomics.** Near-unique SNP (Single Nucleotide Polymorphism) patterns and vast auxiliary data (genealogy sites, beacons), in addition to phenotypes for face patterns.

[1] A summary of DP libraries can be found (OpenMined, 2022).

- **Medical Records.** Rich in QIDs, diagnostic sparsity, and hidden person-specific attributes.
- **Electrocardiogram.** Stable morphological features (PQRST) act as biometric fingerprints.
- **Wearable Devices Data.** Continuous multisensor time series (BVP/EDA/ACC/TEMP), could leak person-specific rhythms and routines.
- **Medical Images.** High-dimensional patterns (brain, histology textures, DICOM)

Beyond the data characteristics, failures also arise from insufficiently generalized QIDs, inadequate governance or data-sharing practices. Anonymization is both technical and governance challenge, and we advocate for both in our architecture.

3.2 Differential Privacy in Practice

DP application in healthcare is to be grounded in practice, following threat recognition and considering different modalities based on the presence of PII/QIDs, rather than isolated implementation. Noise is the main mechanism for DP, and can be introduced using CDP and LDP settings, along with the choice, the application point needs to be fixed such as:

- **Original Data (Non-Interactive, LDP)**, individual's data is perturbed before leaving the source, ensuring anonymization before analysis as in fig. 1. Used for untrusted recipients and when continuous collection is required, it mitigates linkage from raw QIDs and auxiliary information.

Figure 1: Releasing Noisy Database using Local DP.

- **Selected Attributes (LDP)**, noise can be added to only selected attributes before sending them to the server, only perturbing the attributes driving identifiability and used when specific attributes (QIDs, e.g., age, heart rate or location) are flagged as the main attack surface. Thus reducing the re-identification risk without reducing the utility.
- **Aggregate Data (Interactive, CDP)**, noise is added to the output of the aggregate functions, such as count, mean, and sum.

  see fig. 2. In such cases, linkage and inference are dominant threats, and CDP is applied when population insights are required to avoid per-record exposure.

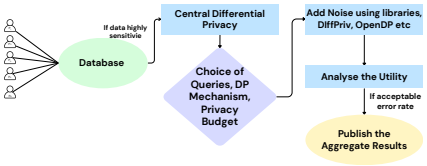

Figure 2: Releasing Aggregate Statistics using CDP.

- **Using Machine Learning**, models introduce noise during the training process. DP-SGD adds noise to the gradient, preventing model from learning specific information about training data. The method is practical when risk is directly bound to the influence of a single patient on the model output. Detailed application and noise addition points are presented in fig. 3. For instance, a classifier can be shared after DP-SGD, along with a report on the associated privacy loss.

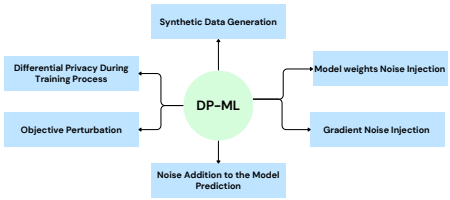

Figure 3: Differential Private Machine Learning.

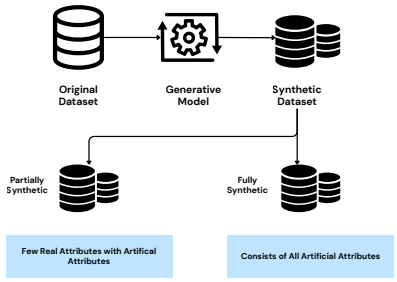

Figure 4: Synthetic Data Generation.

- **Synthetic Data using DP**, generates a synthetic dataset, mimic the original data features. However, it should not relate to any individual in the original data. A practical approach when data analysis and microdata publishing have a high identifiability risk, such as mimicking genome data. The fig. 4 shows high-level intuition for synthetic data generation.

3.3 Attack Landscape and DP Mitigation

Re-identification attacks have evolved as quickly as sophisticated analytic techniques to exploit the ML/Deep Learning (DL) advancements (foremost CNNs for imaging and RNNs for genomics). These

techniques, amongst others, are used in different methods (e.g. inference or linkage) to exploit feature-related threats. Across the published studies, four major classes emerge in the medical field and we studied DP and provided mitigation fitting for each method:

- **Inference Attacks:** Exploit correlations to infer sensitive features without explicit linkage. For example, inferring phenotypes in (Liu et al., 2018), or leak diagnoses in EHR (Pedarla et al., 2023).

  DP Mitigation: If auxiliary data exists and QIDs drive linkage, prefer CDP with aggregate releases; do not publish per-record data.

- **Linkage Attack:** Match IDs or QIDs with auxiliary sources such as public registries. In health, this could lead to patients ID reveal, as in genome beacons or ECG (Ghazarian et al., 2022).

  DP Mitigation: When microdata is shared under high identifiability, synthetic datasets should be preferred, e.g. small cohorts and rare conditions.

- **Membership Attack:** Reveal whether an individual's data was used in a de-identified dataset, or in training a model (Shokri et al., 2017). MIA (Membership Inference attacks), is relevant since DP offers a direct mitigation by restricting the inference of the data's presence in a dataset.

  DP Mitigation: If the artefact is a model, apply DP during training, such as DP-SGD.

- **Correlation and Stylometry Attack:** Exploit textual notes, metadata, usage patterns and stylometrics (Ji et al., 2020) to link across datasets.

  DP Mitigation: Use LDP to introduce noise if trust in the aggregator is limited or collection is on device level and data have persistent patterns.

Through our study of the techniques powering the attacks, we compare how data types are susceptible to de-anonymization, and link them with DP strategies. Adversaries exploit a mix of ML/DL-based, statistical or probabilistic methods considering modalities specifications. We summarized our analysis in table 1 with attack orientation and table 2 for DP aspects, to map contexts with the risks and techniques, provide success rates from literature. We mapped the DP application to the related type and risk to guide the choice of implementation, see table 2. Additionally, we highlight a few DP implementations; however, the scope of these is limited and lacks risk governance.

A simplified mapping from empirical breaches to implementation and governance serves as a foundation for our approach. In fig. 5, we aligned the state of the art from the de-anonymization aspects into the HEALTH-DP framework (details in fig. 7).

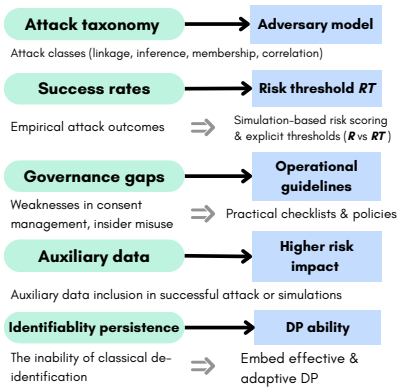

Figure 5: De-Anonymization Insights into the Framework.

4 HEALTH-DP FRAMEWORK

HEALTH-DP framework fig. 7 brings together practical systematic de-anonymization and use of DP, shaped into a modular process. Conceived to support healthcare professionals, including researchers and privacy specialists, in evaluating privacy risks and implementing DP strategies. We ensured that the architecture was not designed in abstraction but directly fed by empirical studies presented in sections 3 and 3.2, creating a practically deployable process.

4.1 Conception and Core Component

We focused on four primary considerations (fig. 6) in our framework development, first by drawing on de-anonymization attacks systematic literature and simulations of mechanisms to reflect threats.

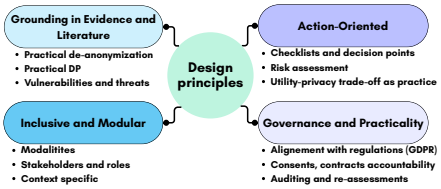

Figure 6: Leading Design Principles for DP Framework.

Furthermore, we considered various modalities and roles to translate theoretical aspects into a practical pipeline that healthcare professionals can apply. Additionally, we built the structure from connected components through phases: Scoping and data collection, to assessment and DP initiation and evaluation.
Data Modalities and Stakeholders: Health data is variable in types, with own specifications, meaning data-specific risk must be addressed. Data is generated and processed by diverse stakeholders, including health and information system staff, researchers and privacy officers. We separate each layer's responsi-

Table 1: De-Anonymization Techniques and Vulnerabilities by Health Data Type.

Data Modality	Key Vulnerabilities	Common Attack Techniques	Success example
Genomic Data	SNP patterns, auxiliary data (beacons)	Likelihood ratio tests, Bayesian inference, graph models, RNN	>95% membership (S Shringar-pure, 2015)
Electronic Health Records (EHR)	QIDs and sparse diagnoses, notes with stylistic identifiers	Probabilistics, auxiliary dbs, stylometric analysis, ML classifiers and features processing	Up to 70% membership in (Shokri et al., 2017)
ECG Signals	Biometric waveforms (PQRST morphology)	Signal processing (fiducial features, DTW alignment), deep CNN/LSTM	∼97% as high linkage rate in (Ghazarian et al., 2022)
Wearable Sensor Streams	Unique daily patterns and routines	Temporal pattern (e.g. DTW variants), clustering, multi-modal DL	Naïve attack avg: 28.9% (Lange et al., 2025)
Medical Images	High-dimensional structures (e.g. brain), texture signatures (MRIs), anatomical patterns,	CNN, transfer learning with pre-trained models, clustering on anatomical feature vectors	94–97% same-person MRIs (Ravindra and Grama, 2021)

Table 2: Differential Private sharing of Health Data Type.

Data Modality	Risk	DP Models/Focus	Dp Implementations
Genomic Data	Genetic uniqueness, SNP structure	Task Specific model, synthetic data/ CDP, DP-Synth	Aggregate statistics (Tramèr et al., 2015; Humbert et al., 2014), GWAS reproducibility (Jiang et al., 2022)
EHR	Linkage Inference	Aggregate only (e.g., count, average) /CDP	Aggregate Statistics, DP QIDs classification scheme (Zhang and Li, 2022)
ECG Signals	Biometric Morphology	DP features, Aggregation, DP Synthetic/ CDP, DP-ML, DP-Synth	Aggregate statistics (Ghazarian et al., 2024), Sythetic ECG Generation (Habiba et al., 2021)
Wearable Devices	Persistent patterns	Selected Attributes/LDP	Real time data collection (Hu et al., 2019), One dim heart rate collection (Kim et al., 2018)
Medical Images	High Dimensional Linkage	DP-Model, synthetic DP/ GAN-DP, DP-SGD	Medical Image classifcation with DP (Ziller et al., 2021), PRIVIMAGE: DP synthetic image generation (Li et al.,)

bilities between stakeholders, including shared ones.

Adversary Models: We distinguish between three models, relying on profile characteristics and attack strategies. We tailored our risk assessment to specific threats, considering attackers potential.

- **External Adversary:** Rely on public or auxiliary sources such as genealogy databases, beacons or registries, and mount linkage or inference attacks.
- **Insider Adversary:** Exploit privileged access or weak governance, including staff or personnel.
- **Advanced ML/DL Adversary:** Use sophisticated analytical capabilities in feature learning, or generative models in biometric re-identification, reconstructions or attribute inference.

Risk Assessment Layer: We quantify the risk through a sequence of tasks, shared mainly between the researcher and the privacy officer (See table 3). Data-specific vulnerabilities to be checked through targeted simulations (e.g., beacon queries for genomics). Attack bench feeds accuracy checks, then normalized into a risk score R compared to a predefined policy-driven threshold (RT), and used to trigger actions and the type of DP to apply. Let $\mathcal{A}$ indicate the set of attacks evaluated. For each attack $a \in \mathcal{A}$, we define the corresponding empirical risk component R_a as: $R_a = AttackSuccess_a$, where AttackSuccess can be accuracy (linkage, identification), membership inference, top-k identification rate, or likelihood score.

Table 3: Risk Assessment Layer: Methods and Outputs.

Step	Examples	Output
Identifiability	k-anonymity, entropy, linkage probability	QIDs risk profile, identifiability score
Modality Analysis	Genomics: beacon queries; ECG: DTW/CNN fingerprints; Imaging: feature/face linkage	Vulnerability map per modality
Attack Simulation	Linkage, membership, inference, prediction	Accuracy of re-identification attempts
Risk Scoring	Normalize results into global R	Risk score compared to threshold
Decision	Governance check, DP application if $R > RT$	Approve or DP-triggered re-assessment

We emphasise that R is not an intrinsic scalar property, but rather estimated from the assumed adversary threat model. Computed as a combination of the individual attack risks:

$$R = \sum_{a \in \mathcal{A}} w_a \cdot R_a \tag{1}$$

where w_a is an application domain-specific weight re-

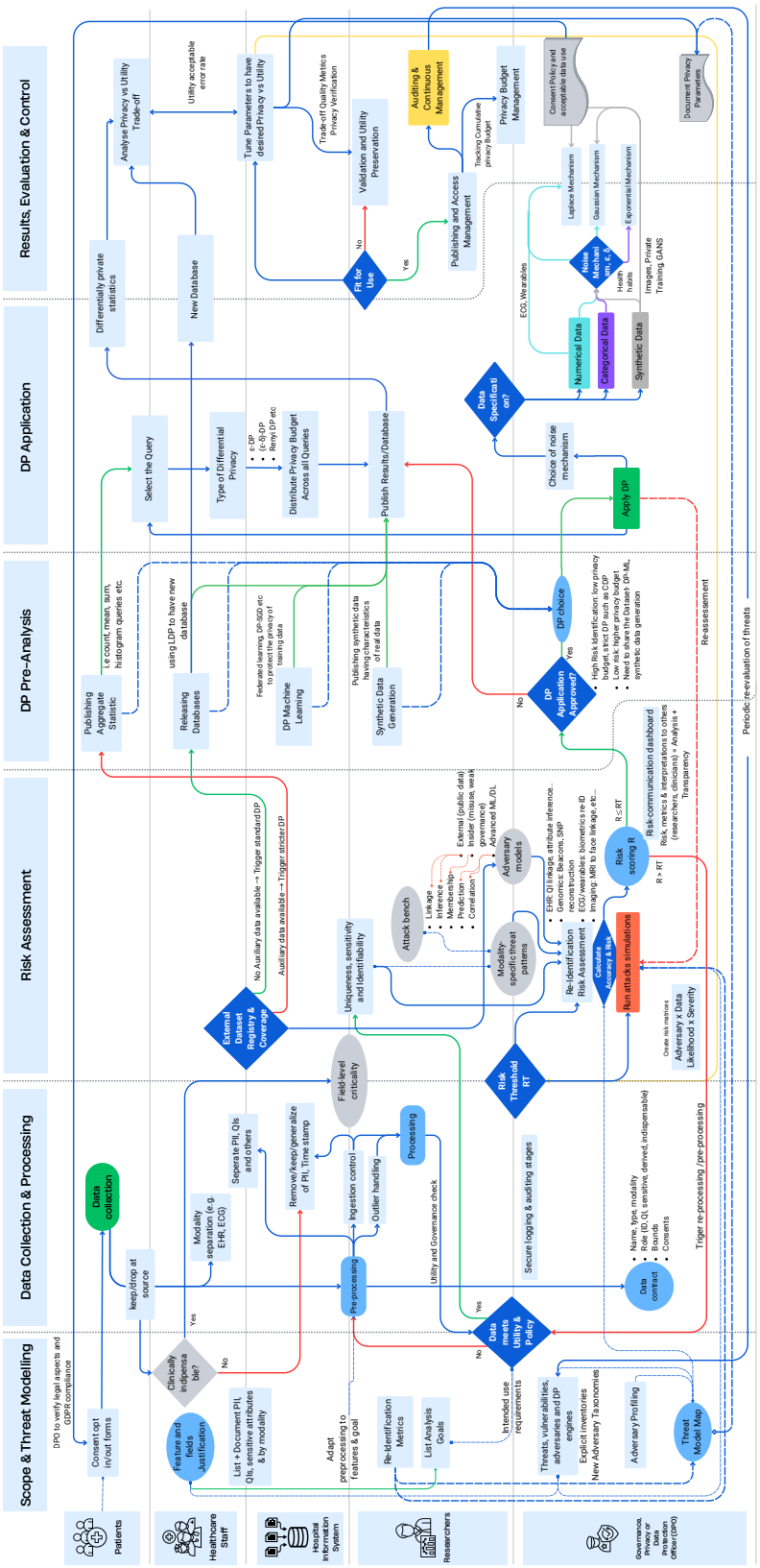

Figure 7: HEALTH-DP Framework.

flecting the relevance of attack a.

Privacy Preserving DP Layer: Privacy preservation using DP requires a specific approach. Foremost, decide on the type of setting to use based on the risk R, non-interactive and interactive, as discussed in section 3.2. For the non-interactive setting, follow:

DP Pre-Analysis: Based on the risk and utility requirement decide on the following parameters

1. Selection of queries to be published, such as count, histograms, to share helpful information.
2. Choose a variant of DP such as ε-DP, (ε, δ), (other DP variants discussed in section 2.1).
3. Sensitivity calculation, to analyze how the output is affected by the individual addition or removal.
4. Selection of the noise mechanism and privacy parameter, for the desired privacy vs utility trade-off.

Budget Management Module: For the management of information sharing and control

- Distribute ε over all the queries.
- Once the privacy budget is exhausted, no further queries will be allowed.

Application and Control: For the interactive setting, the data holder can publish the noisy database publicly using LDP, to allow analysis on the dataset

- Using LDP, noise is introduced to each individual's data.
- Adjust privacy parameters for desired privacy vs utility trade-off, publish new noisy database.
- Using DP for ML to prevent against MIA. Minimizing information leak while training by adding controlled noise during the training process (Abadi et al., 2016), (Nicolas Papernot, 2021), (Ponomareva et al., 2023).
- Differentially private synthetic using GANs. Generating synthetic data for health is exigent due to its high sensitivity and utility requirements. (Murtaza et al., 2023) provide a state-of-the-art synthetic data generation for the healthcare domain.

Processes and Flow: Data preparation is followed by threat modelling, leading to risk assessment, which is supplemented with prior scope studies and DP pre-analysis and application. Results are evaluated against governance checks and re-assessment to ensure the continuous privacy and auditing.

Operational Guidelines: We provide guidelines as practical checklists and governance logic table 4 wrap up the key points before processing health data.

We summarized a simplified version in fig. 8 for a high-level accessibility of the overall pipeline.

Table 4: Operational Checklist for Framework Application.

Step	Checklist item
Scope & Threats	Consents, PII/QIDs and vulnerabilities listed, fields justified, adversaries potential
Risk Assessment	Data checks, vulnerability mapping, attack pre-studies
DP Application	Privacy budgets, mechanisms and libraries; utility and use
Governance & Release	Data contracts; DPO standards and thresholds; policies, auditing and continuous monitoring

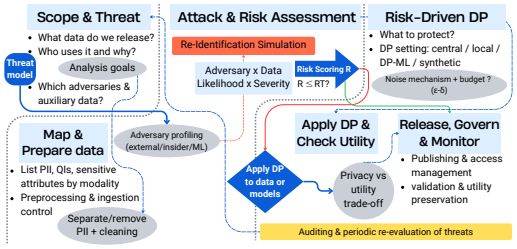

Figure 8: HEALTH-DP Framework - Simplified Overview.

4.2 Framework Application Case Study

This section presents our groundwork use case, applying CDP to the synthetic heart attack dataset (kaggle, 2024), which consists of data related to heart health. We consider a setting where a patient arrives at the emergency department with chest pain, and we implement DP following our pipeline:

1. **Collection and Scope:** During triage and evaluation, the hospital collects the routine variables required for diagnosing heart disease through a questionnaire. The DPO classifies fields to PII, QID, sensitive and clinically indispensable and decides not to release record-level data. The intended use of data is policy/quality reporting and exploratory research using aggregate statistics.
2. **Pre-Analysis:** Bound sensitivities by defining ranges based on feasible physiological limits for aggregates, e.g., DP age histograms, DP average heart rate. Define the desired utility objectives.
3. **Analysis Goals:** Stakeholders need population-level statistics and hospital governance limits the budget to restrict the workload to aggregate queries under CDP, avoiding records publication.
4. **Threat Model and Risk:** We assume an external adversary holding auxiliary data from prior visits and attempting de-anonymization (case of released records), QIDs and physiologies capture the patient's cardiac morphology so the linkage risk is non-negligible, meaning choice of CDP.

5. **DP Design and Implementation:** This section presents use case of applying CDP following algorithm 1. We evaluated noise injection under Laplace and Gaussian, demonstrating the privacy vs utility trade-off for each value of ε for $Q's$ implemented using `diffprivlib`, see figs. 9 to 11.

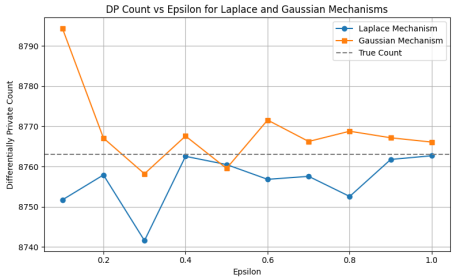

Figure 9: Count Query: Count of Patients in Dataset.

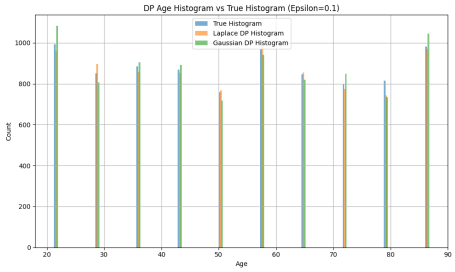

Figure 10: Histogram Query: Age Range of the patients.

Overall, our exploratory results show that:

- Utility Maintenance for basic aggregate statistics.
- Laplace and Gaussian mechanisms have different error profiles.
- Healthcare settings could incorporate the CDP pipeline for releasing aggregate statistics.
- The utility criteria must be evaluated by experts to analyze the acceptable error rate for each type of query and use case.
- The ε values in each graph show the effect of noise introduced to the particular mechanism at a particular ε. The smaller ε indicates more privacy and less utility, and vice versa.

While not exhaustive, our experiment demonstrates the feasibility of coupling re-identification concerns with DP deployment, providing guidance for practitioners in data publishing.

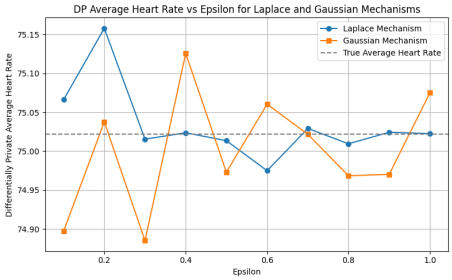

Figure 11: Average Query: Average Heart Rate of Patients.

Algorithm 1: CDP Data Release.

Input: Health dataset D;
Workload $W = \{Q_1, \ldots, Q_n\}$ (e.g. counts, means, histograms);
Total privacy budget $(\varepsilon_{\text{tot}}, \delta)$;
Sensitivity bounds $\{\Delta S\}$;
Noise type (Laplace or Gaussian)

Output: Private release $\widetilde{A} = \{Q_1(D), \ldots, Q_n(D)\}$;
Privacy report (per-query $(\varepsilon_i, \delta_i)$, total loss, utility diagnostics)

Step 1: Workload Preparation
For each $Q_i \in W$, define domain(Q_i), range, and compute sensitivity ΔS_i.

Step 2: Budget Allocation[2]
Distribute $(\varepsilon_{\text{tot}}, \delta)$ across queries using a policy (equal or utility-weighted split):
$\sum_i \varepsilon_i \le \varepsilon_{\text{tot}}, \quad \sum_i \delta_i \le \delta.$; `// Sequential composition`

for $i \leftarrow 1$ **to** n **do**
  if Q_i *numerical and* $\delta_i = 0$ **then**
    Apply **Laplace mechanism**:
    $\eta_i \sim \text{Lap}\left(0, \frac{\Delta S_i}{\varepsilon_i}\right)$;
    $\widetilde{A}(D) = Q_i(D) + \eta_i.$; `// ε-DP`
  else
    if Q_i *numerical and* $\delta_i > 0$ **then**
      Apply **Gaussian mechanism**:
      $\sigma_i \ge \sqrt{2\ln(1.25/\delta_i)} \cdot \frac{\Delta S_i}{\varepsilon_i}$;
      $\eta_i \sim \text{Gau}(0, \sigma_i^2)$;
      $\widetilde{A}(D) = Q_i(D) + \eta_i.$; `// (ε_i, δ_i)-DP`
    else
      Apply **Exponential mechanism**:
      Sample r with probability $\propto \exp\left(\frac{\varepsilon_i}{2\Delta S_i} S(D, r)\right)$;
      Publish $r.$; `// Non-numerical outputs`
  Post-Processing: Clip or normalize outputs to valid physiological ranges; apply consistency checks across related statistics. ;
  `// Preserves DP`

Step 3: Privacy Accountant
Aggregate per-query losses to report overall (ε, δ) under the chosen composition rule; verify budget $\le (\varepsilon_{\text{tot}}, \delta)$.

Step 4: Utility Diagnostics
Compute workload-level errors (e.g., mean absolute error on held-out aggregates) to document privacy–utility trade-offs.

Step 5: Release & Report[3]
Publish $\widetilde{A}$ with meta-summary: bounds, mechanisms, $(\varepsilon_i, \delta_i)$, total budget, composition rule, based on RT.

5 DISCUSSION

We introduce a risk-based deployment framework that operationalizes when, where and how to apply DP across healthcare data. The table 5 provides a qualitative comparison with the related work, high-

[2] Guidance on ε allocation can be found in (Dwork et al., 2019), (Hsu et al., 2014), (Lee and Clifton, 2011).

[3] Include the dataset's risk score $\mathcal{R}_{\text{assess}}$, privacy budget, and accountability rule in the release documentation.

lighting a gap in research for a similar end-to-end framework, thus the limited benchmarking.

5.1 Evaluation

Our study improves on works by (Moqurrab et al., 2023) that presents a two-phase privacy pipeline, but lack attack assessment grounds, and similarly in (Dyda et al., 2021) that provides a tool that embeds CDP using only histogram queries (our experiments demonstrated three different queries, see section 4.2) into their conceptual private reporting platform. For (Lyu et al., 2025), authors suggest an End-Edge-Cloud process flow but no attack-prioritized scoring for their DPSDG model. Our work also differs from literature surveys (e.g., (Liu et al., 2024)) that evaluate DP literature, and studies with one modality focus (Saifuzzaman et al., 2022) with no guidelines or frameworks. HEALTH-DP closes the literature gap by: suggesting attack-prioritized de-anonymization risk assessment to score risk level and map it to introduce pre-DP and DP, a comprehensive application following decision flows for suitable actions.

5.2 Strengths and Added Value

We developed a framework intending to bring a set of useful privacy aspects for the healthcare domain:

- **Evidence-Driven:** Based on systematic de-anonymization practical attacks, and grounded DP to create evidence-based guidelines.
- **Modular and Inclusive:** Integrate multiple roles, data types, in logical processes, including pre-assessment, DP application and evaluations.
- **Practical Orientation:** Going beyond theoretical or high-level strategies, for inclusive adoption with operational checklists to consider, strategy-oriented and a straightforward workflow.
- **Governance Integration:** Emphasize roles, contracts, and consents, acknowledging the organizational importance and separation of duties.
- **Bridging Research and Practice:** Translates state-of-the-art into guidance for healthcare.

5.3 Limitation and Challenges

While HEALTH-DP advances the state of practice in the face of de-anonymization, limitations and challenges remain:

- **Implementation Complexity:** Adversary profiling, attack simulations, and tuning DP require expertise and, in some cases, high computing.
- **Generality vs Specificity:** Combining modality-specific and threats is challenging, as some risks cannot be generalized (e.g. SNPs).
- **DP Trade-Off:** Privacy vs utility balancing is challenging while applying DP. Although it remains an open question, guidance can be found in [2](Hsu et al., 2014)(Nanayakkara et al., 2023).
- **Synthetic Data:** Synthetic data generation faces quality and evaluation complexity challenges requiring advanced analysis and assessment.
- **Governance:** Organizational processes are vital for data protection, yet the human factor remains a weak link that challenges governance.
- **Evaluation/Validation:** Our framework is prototypical, large-scale and requires multiple experiment settings for substantial design validation.

6 CONCLUSION

Privacy protection against de-anonymization requires more than standalone mechanisms. It demands a consolidated process that covers scoping and pre-assessment for DP deployment, while separating roles for a practical approach. Our study addressed this by introducing an inclusive privacy framework that comprehends de-anonymization risk assessment and DP mitigation, concise for the healthcare domain. Our proposal begins with threat modelling, data handling, sensitivity and attacks profiling to incorporate re-identification assessment, and aligns DP deployment with institutional governance and stakeholder responsibilities. An approach that advances the application of advanced anonymization from isolated techniques towards a logical pipeline. This layered guideline responds to prior work's shortcomings and often applies DP within a narrow modality context without providing general, scalable structures. We align technical measures with medical institutional formation, embed privacy constraints into decision processes and selection for each practice, and validate outcomes.

The preliminary case study on medical records demonstrated the CDP's ability to preserve aggregate-level insights while mitigating re-identification risks inherent in physiological data, justifying the choice of DP. Although our validation remains limited, we provide a foundation for HEALTH-DP adoption, and our results confirm the necessity of a risk-aware DP to advance research collaborations and innovation. Future research aims to expand the validation to a full practical deployment and to investigate adaptive DP to respond to changing threats and balance protection against clinical research de-anonymization attacks.

Table 5: Comparative Analysis of DP Frameworks and Models in Healthcare vs. HEALTH-DP.

Paper	**Type / Scope**	**DP model(s) & Output**	**Workflow**	**De-anonymization Assessment**	**ε Guidance**	**Governance / Roles**	**Notes vs HEALTH-DP**
(Dyda et al., 2021)	DP tool and dashboards for COVID Surveillance	CDP query release	Partial; Proof-of-concept orientation;	No attack-grounded pre-assessment	Query specefic ε and composability	Limited governance aspects	Uni-vision into COVID; no modality-aware threat phase, no role separation
(Zhang and Li, 2022)	DP publishing model	ACDP-Tree based on attribute correlation	Yes (Comprehensive publishing pipeline)	No	Considered budgeting and composition properties	No	Algorithmic guideline but no governance or attack lens
(Liu et al., 2024)	Survey and high-level application guideline	All	N/A	No	Discusses ε-utility privacy trade-offs	No	Sets baseline and guidance work, lacks scalable framework aspect
(Moqurrab et al., 2023)	E-health data publishing framework	Generalization + sampling framework	Yes (two-phase pipeline)	No	Generic guidline on ε and utility privacy trade-off	No	Inclusive framework aiming at utility awareness but no attack consideration or governance
(Ghazarian et al., 2024)	Modality-specific (ECG) guidance + case study	CDP query release on ECG analytics	Yes (six-step CDP workflow)	Partial (motivates ECG identifiability, no formal attacks assessment)	Practical guidance on choosing & splitting ε	No	ECG workflow, not multi-modality, no roles, pre-DP assessment not formalized
Lyu et al., 2025	Smart healthcare layered model	GAN model for Synthetic data	Yes (Three layers Cloud–Edge process)	No	Guided	No	Strong ε allocation for synthetic generation, lack risk layer & governance
HEALTH-DP (ours)	**Framework across modalities + case study**	**CDP/LDP; aggregates/synthetic**	**Yes:** end-to-end with role-based decision flow	**Yes:** attack-prioritized risk assessment	**Scalable parameters, budget tracking and composition**	**Yes:**	N/A

ACKNOWLEDGEMENTS

Funded by the European Union (EU), Grant Agreement no.101095717 (SECURED) and Taighde Éireann–Research Ireland Grant no.18/CRT/6222. Views and opinions expressed are those of the authors and do not necessarily reflect those of the EU or the Health and Digital Executive Agency.

This publication has emanated from research conducted with funding by Taighde Éireann - Research Ireland under Grant no.18/CRT/6223.

REFERENCES

Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., and Zhang, L. (2016). Deep learning with differential privacy. ACM,

Aguelal, H. and Palmieri, P. (2025). De-anonymization of health data: A survey of practical attacks, vulnerabilities and challenges. In *11th Int. Conf. ICISSP*.

Al-Azizy, D., Millard, D., Symeonidis, I., O'Hara, K., and Shadbolt, N. (2016). A literature survey and classifications on data deanonymisation. In *Risks and Security of Internet and Systems*.

Barbaro, M., Zeller, T., and Hansell, S. (2006). A face is exposed for aol searcher no. 4417749. *New York Times*.

Dwork, C. (2006). Differential privacy. In *Intl Colloquium on Automata, Languages and Programming*.

Dwork, C., Kohli, N., and Mulligan, D. (2019). Differential privacy in practice: Expose your epsilons! *Journal of Privacy and Confidentiality*, 9(2).

Dwork, C., McSherry, F., Nissim, K., and Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography*.

Dyda, A., Purcell, M., Curtis, S., Field, E., Pillai, et al. (2021). Differential privacy for public health data: An innovative tool to optimize information sharing while protecting data confidentiality. *Patterns*, 2(12).

Emam, K. E., Jonker, E., Arbuckle, L., and Malin, B. (2011). A systematic review of re-identification attacks on health data. *PLoS ONE*, 6(12):e28071.

European Union (2016). Regulation 2016/679 general data protection regulation. Journal of the European Union.

Fung, B. C., Wang, K., Chen, R., and Yu, P. S. (2010). Privacy-preserving data publishing: A survey of recent developments. *ACM Computing Surveys (Csur)*.

Gary Howarth, e. a. (2025). A community-driven differential privacy deployment registry. NIST Internal Report 8588 (Initial Public Draft).

Ghazarian, A., Zheng, J., and Rakovski, C. (2024). Privacy-preserving ecg data analysis with differential privacy: A literature review and a case study.

Ghazarian, A., Zheng, J., Struppa, D., and Rakovski, C. (2022). Assessing the reidentification risks posed by deep learning algorithms applied to ecg data. *IEEE Access*.

Habiba, M., Borphy, E., Pearlmutter, B. A., and Ward, T. (2021). Ecg synthesis with neural ode and gan models. In *Conf. on Electrical, Computer and Energy Technologies*, pages 1–6. IEEE.

Hall, R., Wasserman, L., and Rinaldo, A. (2013). Random differential privacy. *JPC*, 4(2).

Henriksen-Bulmer, J. and Jeary, S. (2016). Re-identification attacks—a systematic literature review. *International Journal of Information Management*, 36:1184–1192.

Hsu, J., Gaboardi, M., Haeberlen, A., Khanna, S., Narayan, A., Pierce, B. C., and Roth, A. (2014). Differential privacy: An economic method for choosing epsilon. In *2014 IEEE 27th Computer Security Foundations Symposium*, pages 398–410. IEEE.

Hu, Y., Ge, L., Zhang, G., and Qin, D. (2019). Research on differential privacy for medical health big data processing. In *20th Intl. Conf. on Parallel and Distributed Computing, Applications and Technologies*. IEEE.

Humbert, M., Ayday, E., Hubaux, J.-P., and Telenti, A. (2014). Reconciling utility with privacy in genomics. In *Proceedings of the 13th Workshop on Privacy in the Electronic Society*, pages 11–20.

IBM Security (2023). Cost of a data breach report 2023. Technical report, IBM Security.

Ji, S., Gu, Q., Weng, H., Liu, Q., Zhou, P., Chen, J., Li, Z., Beyah, R., and Wang, T. (2020). De-Health: All Your Online Health Information Are Belong to Us. In *2020 IEEE 36th Intl Conf on Data Engineering (ICDE)*, pages 1609–1620, Dallas, TX, USA. IEEE.

Jiang, Y., Ji, T., Li, P., and Ayday, E. (2022). Reproducibility-oriented and privacy-preserving genomic dataset sharing. *arXiv preprint arXiv:2209.06327*.

kaggle (2024). heart attack prediction. www.kaggle.com/datasets/iamsouravbanerjee/heart-attack-prediction-dataset [accessed on 2025-10-21].

Kim, J. W., Jang, B., and Yoo, H. (2018). Privacy-preserving aggregation of personal health data streams. *PLOS ONE*, 13(11):1–15.

Lange, L., Schreieder, T., Christen, V., and Rahm, E. (2025). Slice it up: Unmasking user identities in smartwatch health data. In *20th ACM Asia Conf. on Computer and Communications Security*, ASIA CCS '25, page 710–726, NY, USA.

Lee, J. and Clifton, C. (2011). How much is enough? choosing ε for differential privacy. In *Intl. Conf. on Information Security*, pages 325–340. Springer.

Li, K., Gong, C., Li, Z., Zhao, Y., Hou, X., and Wang, T. Privimage: Differentially private synthetic image generation using diffusion models with semantic-aware pretraining. In *33rd USENIX Security Symposium*.

Liu, F. (2018). Generalized gaussian mechanism for differential privacy. *IEEE Transactions on Knowledge and Data Engineering*, 31(4):747–756.

Liu, W., Zhang, Y., Yang, H., and Meng, Q. (2024). A survey on differential privacy for medical data analysis. *Annals of Data Science*, 11(2):733–747.

Liu, Y., Wan, Z., Xia, W., Kantarcioglu, M., and other (2018). Detecting the presence of an individual in phenotypic summary data. In *AMIA Annual Symposium Proceedings*, pages 760–769.

Lyu, M., Ni, Z., Chen, Q., and Li, F. (2025). Edge-DPSDG: An Edge-Based Differential Privacy Protection Model for Smart Healthcare. *IEEE Transactions on Big Data*.

Mironov, I. (2017). Rényi differential privacy. In *2017 IEEE 30th computer security foundations symposium (CSF)*, pages 263–275. IEEE.

Moqurrab, S. A., Naeem, T., Shoaib Malik, M., Fayyaz, A. A., Jamal, A., and Srivastava, G. (2023). Utilityaware: A framework for data privacy protection in e-health. *Information Sciences*, 643:119247.

Murtaza, H., Ahmed, M., Khan, N. F., Murtaza, G., Zafar, S., and Bano, A. (2023). Synthetic data generation: State of the art in health care domain. *Computer Science Review*, 48:100546.

Nanayakkara, P., Smart, M. A., Cummings, R., Kaptchuk, G., and Redmiles, E. M. (2023). What are the chances? explaining the epsilon parameter in differential privacy. In *32nd USENIX Security Symposium*.

Narayanan, A. and Shmatikov, V. (2006). How to break anonymity of the netflix prize dataset. *arXiv preprint cs/0610105*.

Newaz, A. I., Sikder, A. K., Rahman, M. A., and Uluagac, A. S. (2021). A survey on security and privacy issues in modern healthcare systems: Attacks and defenses. *ACM Trans. Comput. Healthcare*, 2(3).

Nicolas Papernot, A. G. T. (2021). Nist. https://www.nist.gov/blogs/cybersecurity-insights/how-deploy-machine-learning-differential-privacy.

OpenMined (2022). A survey of differential privacy frameworks.

Pedarla, L. P., Zhang, X., Zhao, L., and Khan, H. (2023). Evaluation of query-based membership inference attack on the medical data. In *2023 ACM Southeast Conf.*, page 191. Assoc. for Computing Machinery.

Ponomareva, N., Hazimeh, H., Kurakin, A., Xu, Z., Denison, C., McMahan, H. B., et al. (2023). How to dp-fy ml: A practical guide to machine learning with differential privacy. *Journal of AI Research*.

Ravindra, V. and Grama, A. (2021). De-anonymization Attacks on Neuroimaging Datasets. In *Proceedings of the 2021 Intl, Conf. on Management of Data*. ACM.

S Shringar-pure, C. D. B. (2015). Privacy Risks from Genomic Data-Sharing Beacons. *The American Journal of Human Genetics*, 97(5):631.

Saifuzzaman, M., Ananna, T. N., Chowdhury, M. J. M., Ferdous, M. S., and Chowdhury, F. (2022). A systematic literature review on wearable health data publishing under differential privacy. *Journal of Information Security*.

Shokri, R., Stronati, M., Song, C., and Shmatikov, V. (2017). Membership inference attacks against machine learning models. In *IEEE Symposium on Security and Privacy*.

Subramanian, R. (2022). Applications of differential privacy to healthcare. *Available at SSRN 4005908*.

Tramèr, F., Huang, Z., Hubaux, J.-P., and Ayday, E. (2015). Differential privacy with bounded priors: reconciling utility and privacy in genome-wide association studies. In *Proceedings of ACM SIGSAC conference on computer and communications security*.

Wang, Y., Wu, X., and Hu, D. (2016). Using randomized response for differential privacy preserving data collection. In *EDBT/ICDT Workshops*.

Zhang, S. and Li, X. (2022). Differential privacy medical data publishing method based on attribute correlation. *Scientific Reports*, 12(1):15725.

Zhou, B., Pei, J., and Luk, W. (2008). A brief survey on anonymization techniques for privacy preserving publishing of social network data. *SIGKDD Explor.*

Zia, M. T., Khan, M. A., and El-Sayed, H. (2020). Application of differential privacy approach in healthcare data–a case study. In *2020 14th intl. conf. on innovations in information technology (IIT)*, pages 35–39.

Ziller, A., Usynin, D., Braren, R., Makowski, M., Rueckert, D., and Kaissis, G. (2021). Medical imaging deep learning with differential privacy. *Scientific Reports*, 11(1):13524.