ISSN: 2093-5374 / E-ISSN: 2093-5382

# Improved Data Privacy with Differential Privacy in Federated Learning

Cina Mathew[1*], and Dr.P. Asha[2]

[1*]Sathyabama Institute of Science and Technology, Chennai, India. cinamma@gmail.com, https://orcid.org/0009-0006-8573-7379

[2]Professor, Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Chennai, India. ashapandian225@gmail.com, https://orcid.org/0000-0003-3046-8811

Received: May 10, 2024; Revised: July 15, 2024; Accepted: August 14, 2024; Published: September 30, 2024

## Abstract

Multiple users may train machine learning models cooperatively using Federated Learning (FL). There is a risk of malicious acquisition of participants' personal data due to the fact that traditional machine learning needs users to provide data for training. Through the use of federated learning, which involves moving the training process from a central server to terminal devices, users' data may be protected. Each participant keeps their dataset local and only exchanges model updates. This research proposed an innovative proposal for the medical industry's differentiated privacy approach for overcoming these problems. When several healthcare organizations work together to develop models that use different and extensive information, clinical applications may be greatly enhanced. Thus, the Local and Centre Differential Privacy (LCDP) on clinical datasets is a feature of our proposed approach. Reason being that the training data is the primary emphasis of the local model, while the machine learning model is the primary focus of the central model. We discover that the local model and the central model are linked in a unique way, changes in the original data lead to changes in the gradient, which in turn lead to changes in the model parameters. Based on this finding, our technique is better than prior central methods since it protects the data, gradient, and model all at once by bridging the gap between the two. Our system provides better privacy protections and even higher performance than some of the best previous central methods, which is an excellent outcome of rigorous evaluation.

**Keywords:** Federated Learning, Privacy, Attack Detection, Security, Data Leakage.

## 1 Introduction

Advances in computer power and data gathering techniques have been beneficial to machine learning in recent years, allowing the field to grow in scope and efficiency (Jordan & Mitchell, 2015). Massive data processing and model training are made possible by the high-performance computing resources made available by cloud computing platforms. Prediction, classification, and recommendation systems are just a few areas where machine learning benefits from diverse data sets. One example is the machine learning platform used by the biotech firm Berg (Fleming, 2018). to examine vast amounts of patient biological outcome data, such as lipids, metabolites, enzymes, and protein spectra. This platform helps to highlight important differences between healthy and unhealthy cells and to discover new cancer mechanisms. The

*Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA),* volume: 15, number: 3 (September), pp. 262-280. DOI: *10.58346/JOWUA.2024.I3.018*

*Corresponding author: Sathyabama Institute of Science and technology, Chennai, India.

goal of collecting vast and varied datasets for training in machine learning is to create more trustworthy models, which in turn requires users to submit their data to a third-party server (Zhou, 2016). The majority of this data come from devices on the borders, such glucose monitors, GPS trackers, and cell phones. On the other hand, personal information like medical records & travel plans are often included in this data (Liu et al., 2021). There may be serious consequences if this sensitive information fell into the wrong hands. Moreover, data sharing is strictly prohibited in several specific industries (De Cristofaro, 2021). Consequently, it is critical to ensure that users data remains private while doing machine learning (Jarin & Eshete, 2022). One distributed machine learning approach that Google suggested in 2016 is Federated Learning (McMahan et al., 2016; Udayakumar et al., 2023). By doing model training on local devices, it hopes to safeguard data privacy. The objective is to have data trained on edge devices, where users may adjust model parameters (gradients) based on the current environment (Mohandas et al., 2024). After each cycle, the cloud server updates the global model by aggregating the local gradients using algorithms like FedAvg (McMahan et al., 2017). In this process, participants only transmit their local gradient information to the server (Sindhusaranya et al., 2023; Udayakumar et al., 2023).

The original data is often delivered to a trusted central location, or "data center," before training in real-world situations, as depicted in Figure 1 (a). In the event that the "data center" cannot be trusted, local differential privacy (LDP) was suggested as a solution to provide a reasonable denial (Beimel et al., 2008; Peter et al., 2014). This method involves randomly distributing data before it is made public. Rather than concentrating on the final machine learning model, LDP prioritizes the privacy of communications between people and the 'server,' as seen in Figure 1 (b) (Di et al., 2018; Wang et al., 2019). However, LDP's privacy-preserving noise is consistently significant which limits predictive performance.

In order to address the issues highlighted before, this research examines the LCDP technique, which leads to improved differential privacy in the end model. In Figure 1, you can see how our technique stacks up against earlier perturbation methods. It is clear that our approach maintains the original data to a certain degree while concentrating on the final model. When compared to conventional central models, there is a significant reduction in the leaking of critical information, even if the data centre is hacked. In fact, it is standard practice in computer vision to introduce noise into raw data in order to protect individual privacy (Hill et al., 2016). Because of this, getting back to the original data is a problem (Agrawal & Srikant, 2000). As a result of pre-processing protections with noise, our technique outperforms conventional central models in terms of reliability. Along with that, we find that our input perturbation strategy bridges the gap between central and local differential privacy by perturbing not only the final model parameters but also the gradient.

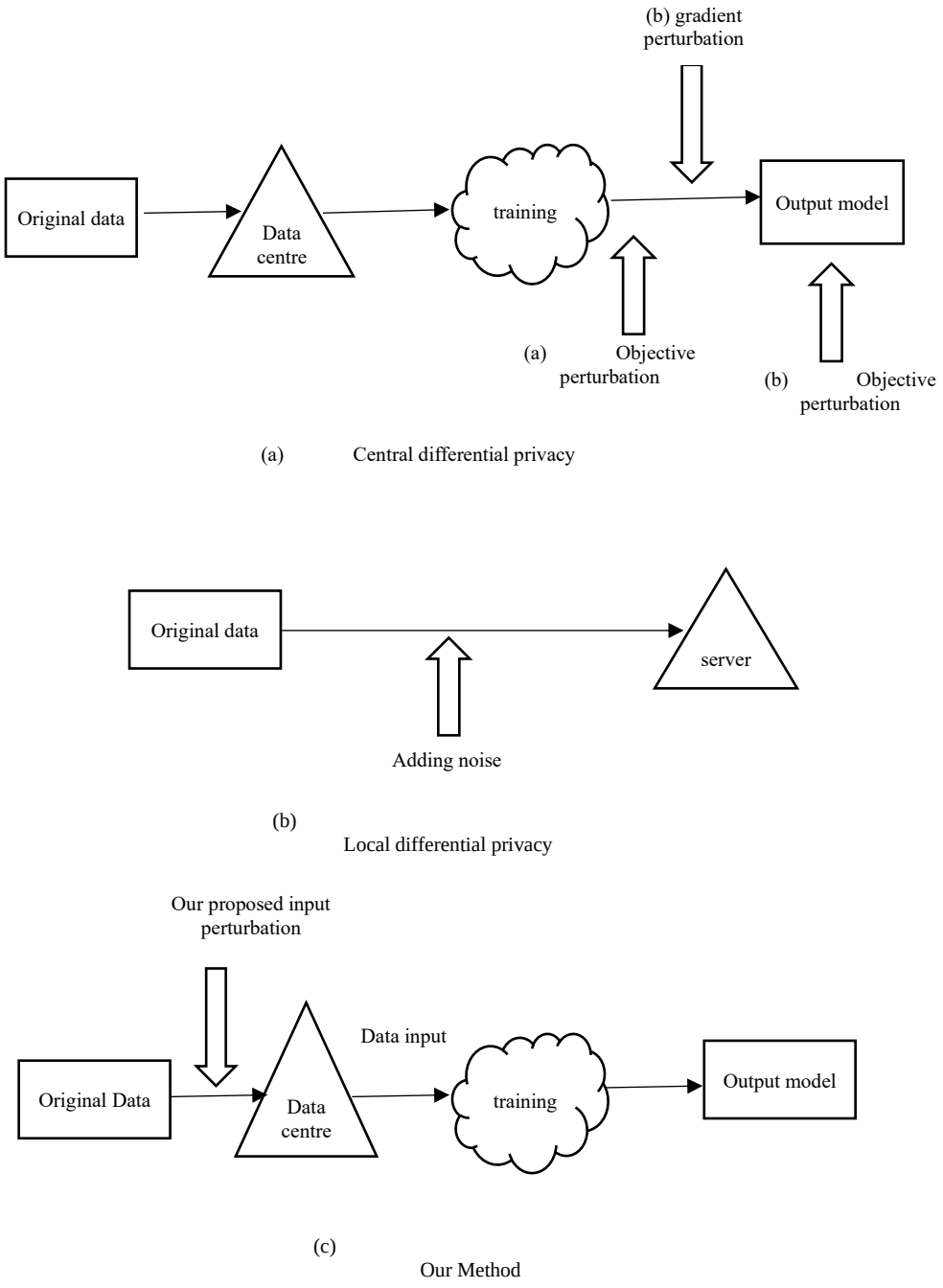



Figure 1: CDP and LDP - Differential Privacy Model

**Contribution of Our Proposed Model**

- For the purpose of evaluating the confidentiality of health records, we have created an assessment methodology we call LCDP.
- We ensure differential privacy on the final model and some type of privacy on the original data concurrently by bridging the gap between CDP and LDP Model.
- Our approach outperforms the current algorithms in the central context, according to thorough theoretical study and experimentation.

The following is the outline for the remainder of the paper. A brief summary of related work is provided in the related work section. In the section under "Preliminaries," the requirements are detailed. This section looks into the Proposed methodology and the necessary security measures, The sections dedicated to our proposed design and security analysis include extensive descriptions and analyses of the plan. Afterwards, experimental design and evaluation will be shown in the performance analysis section. The paper is summarized in the last section, the Conclusion.

## 2 Related Work

### 1) Importance of Privacy in Health Care

Data, models, or code may be exposed as a consequence of privacy attacks that occur during FL training (See Figure 2). To ensure the confidential and ethical treatment of protected health information (PHI), data privacy in healthcare settings is essential. Ensuring the security of confidential patient data and keeping trust in health systems requires the identification of specific threats, the implementation of strong mitigation strategies, the adoption of encryption technologies, and compliance with privacy regulations (Bobir et al., 2024). These regulations include laws like HIPAA in the US (Annas, 2003), GDPR in the EU (Voigt & Von Dem Bussche, 2017) and DISHA in India (Haidar & Kumar, 2021). The following are some important health-care scenarios that are related to privacy preservation and FL: (1) electronic health records (EHRs) that contain detailed information about a patient's medical history, procedures, diagnoses, treatments, medications, and other relevant data; (2) wearable devices that gather data about a patient's fitness, nutrition, and overall well-being that is directly linked to protected health information (PHI); (3) picture archiving and communication systems (PACSs) and biobanks (4) Billing and health insurance system; and (5) Medical policy (Odeh & Taleb, 2023).

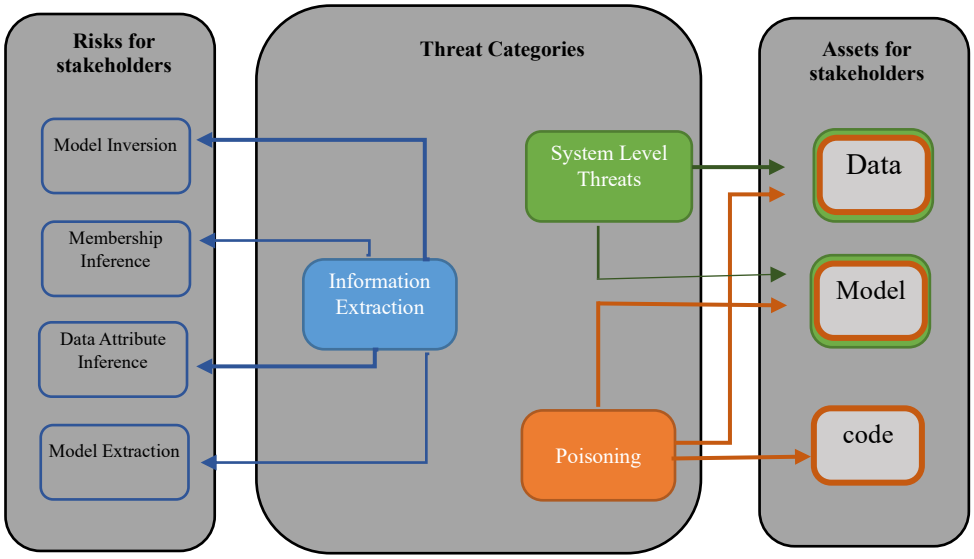



Figure 2: Categories of Threat in Medical Industry

Figure 2. Users in a federated learning system face various risks related to different types of privacy threats (Adnan et al., 2022; Malekzadeh et al., 2021; Ziller et al., 2021; Pfohl et al., 2019; Tayebi Arasteh et al., 2023). The green box represents "system-level threats" that aim to compromise the data and model, the orange box represents "poisoning" attacks that try to expose the secret data, and the blue box represents "information extraction" that aims to undermine model inverting, a membership reasoning, information characteristic inference, and the model extraction.

### 2) Differential Privacy Models in FL

By using the mitigation measures outlined below, we may safeguard data during training-by-training models during FL. On the other hand, these methods aren't fool proof against membership inference attacks, in which an attacker uses a trained model to deduce details about the training data that was discovered or learned (Dwork & Feldman, 2018). One popular method for reducing the extent to which a model learns specific inputs from the data during training is differentially private model training. It does this by including randomness into the training process, which allows it to learn from the data overall while keeping the impact of individual contributions. A privacy assurance associated with DP algorithms is the probability that any given data point may be identified. An algorithm is considered "differentially private" in a broad sense if its output cannot be used to determine the presence or absence of a specific data contribution in the dataset that was used to train the algorithm (Dwork & Feldman, 2018). The kind of contributions that DP often takes into account range from those of a single data record to those of whole collaborator databases. Users' privacy concerns in AI have led to the rise of DP training algorithms, despite the fact that DP ideas originated in data analytics (Abadi et al., 2016). It is possible to utilize DP algorithms in a federated environment either for training local model updates or for aggregating models at the global consensus level (Zhao et al., 2019). Each participating institution conducts local training using a DP training algorithm in local DP FL (Adnan et al., 2022). Here, a DP privacy guarantee is used to construct the local model updates that are communicated to aggregator. This may be a good thing to have if you don't trust the aggregator's administrators or operators to keep their infrastructure safe from privacy threats (Sadilek et al., 2021).

In the absence of pre-existing trust, the aggregator infrastructure may be secured by implementing the privacy options covered above. If you're looking to maximize usefulness while maintaining a certain amount of privacy, global DP FL is the way to go rather than local FL since it combines more data (all collaborator changes) before notifying (Liu et al., 2022). Deep learning (DL) has begun to gain popularity for medical applications, although using DP reduces "model utility," or the model's ability to generalize to new data, and increases computation (Pati et al., 2021). There is less model usefulness because of the extra noise that is added during training. There is also more computation because the training might need to use the base computing system differently (Lee & Kifer, 2021), which could mean that more rounds of training are needed. It is worth noting that DP training in FL may prevent some collaborators from using data quality checks. This is because private local model changes at the aggregator might hide signals that could cause problems. Articles that provide a survey of DP (Shen & Zhong, 2021; Demelius et al., 2023) might be useful for synthesizing the many methods, recommendations, and areas for further study. Nevertheless, more research is necessary to comprehend the costs and benefits linked to certain applications, such as the amount of utility loss that may be expected at a particular privacy level. According to current research on DP FL training in medicine (at $\Delta = 4$, for instance), federated training with DP may achieve scores that are within 5% of what would be possible without DP. The generalizability of these early findings will be better understood when further research is conducted across other datasets and model architectures. Another important factor to think about is the level of anonymity that a DP algorithm provides.

Here, we introduce noise into the data, which perturbs the gradient and leads to improved differential privacy on the metaheuristic parameter of the model, bridging the gap between central and local differential privacy. Compared to earlier central perturbation approaches, our method produces superior theoretical findings, as can be shown by an in-depth review.

# 3 Preliminaries

In this section, we will have a look at DP and Federated Learning (FL).

### 1) Federated Learning (FL)

FL allows users to train machine learning models in a collaborative environment (McMahan et al., 2017). There is a central server and $N$ participants, or clients, who each have their own private dataset. A key difference from the conventional centralized method is that data is not aggregated and sent to all participants at once; instead, each participant trains their model locally and updates their parameters by exchanging the trained model with the server. FL consists of several rounds: At the beginning of round 0, the server sends out models to all participants that have random parameters $\Theta_0$. Following this, $K$ out of $N$ participants are chosen at random for each round r. Each participant i updates their parameters by computing training gradients locally using their dataset $D_i$ and sending them to the server. This second step calculates the global parameters $\theta_r = \sum_{i=1}^{K} \theta_i / K$ and then distributes them to all $N$ participant for the following section. The model has been stopped using parameters $\theta_R$ after a certain number of iterations $(R)$. Various methods exist in FL to improve privacy. It is possible to encrypt the parameters of the participants using Homomorphic Encryption $(HE)$ so that the server can only decipher the aggregates (Bonawitz et al., 2017). While this helps reduce the probability of data leaks caused by global parameters, it is resource-intensive and does limit to protect against inference attacks on the aggregated output (Kairouz et al., 2021). We will next go across an alternative method that makes use of differentially private approaches.

### 2) Differential Privacy (DP)

As a statistical measure, DP protects against the information that an attacker may infer from a random algorithm's output. Adding noise to the algorithm gives an unconditional upper constraint on how much of an impact a single person can have on the result (Dwork & Roth, 2014).

**Definition 1**

**Differential Privacy.** A random process If for any two adjacent databases, $D_1$ and $D_2$, that vary in just one record, and for all potential outputs $S \in \text{Range}(A)$, M offers $(\varepsilon, \delta)$-differential privacy, which are given in Equation (1).

$$P[M(D_1 \epsilon A)] \leq e^{\varepsilon} P[M(D_2 \epsilon A)] + \delta \quad (1)$$

One measure of privacy loss is the ε parameter, which is also known as the privacy budget. Additionally, it manages the trade-off between privacy and utility, meaning that lower ε values signify greater privacy but probably lower utility as well. The δ parameter takes into consideration a (small) chance where the maximum limit ε is not applicable. It is the greatest change in the output owing to the addition or removal of a single record that determines sensitivity of the output, which in turn determines the quantity of noise required to achieve DP. Learn a dataset distribution using DP in ML while protecting individual records' privacy (Ji et al., 2014). For this research, we use LCDP, which stands for Local Central Differential Privacy, as opposed to PATE, which stands for Private Aggregation of Teacher Ensembles. To locate differentially private minima, DP-SGD (Abadi et al., 2016) use a noisy variant of stochastic gradient descent. This is accomplished by first defining the gradients' boundaries, and then using the "moments accountant" approach to introduce noise while monitoring the privacy budget. On

the other hand, PATE (Papernot et al., 2016) uses a student instructor architecture to protect training data.

### 3) DP in FL

As previously stated, there are two versions of DP that may be used in the context of FL: local and central (Geyer et al., 2017; McMahan et al., 2017).

#### Local Differential Privacy (LDP)

Each participant in an LDP session adds noise locally, which is a prerequisite for DP. The results of a random perturbation method M are sent to the server by each participant. The ε value ensures that altered outcome will safeguard an individual's data. That follows a formal definition (Duchi et al., 2013).

**Definition 2:** From Equation (2), Consider a collection of potential values, X, and a set of noisy values, Y. For every $x_1, x_2 \in X$ and every $y \in Y$, M is $(\varepsilon, \delta)$-locally differentially private ($\varepsilon -$ LDP) if:

$$P[M(x) = y] \leq e^{\varepsilon} P[M(x') = y] + \delta \quad (2)$$

Since participants train the model on their datasets using proposed model LCDP, we apply LDP in FL. Using this method, we may monitor the privacy budget's usage with the help of moments accountant. Algorithm 1 demonstrates the operation of LCDP in LDP.

**Algorithm 1: LDP in FL**

---

Main ():
Initialization: mode $\theta_0$
For all epoch r =1,2, ….do
$K_r \leftarrow$ Ķ samples are selected randomly
For all samples k ϵ Ķ do
$\theta_r^k \leftarrow$Optimizer (…) //This is done in parallel
end

$\theta_r \leftarrow \Sigma_{i=1}^{kr} \frac{nk}{n} \theta_r^k$ *//$n^k$ is the size of dataset-K*

end
return

Optimization (*S- norm, D- dataset, p-Probability, noise magnitude σ, learning rate η, Iteration E, loss*

*Function L(θ(x), y)*:
Initialization $\theta_0$
**for** *all local iteration I from 1 to* E do
**for** (*x, y)* ϵ *random data from dataset D with p* **do**
$g_i = \nabla_\theta L(\theta_i; (x, y))$
**end**

$$Temp = \frac{1}{PD} \Sigma i \in batch\ gi \in \min\left(1, \frac{s}{||gi||2}\right) + N(0, \sigma 2I)S$$

$$\theta_{i+1} = \theta_i - \eta(Temp)$$

***end***
*return* $\theta_E$

## Central Differential Privacy (CDP)

As a result of the server interfering with the FL aggregation function, participant-level DP is achieved using CDP. This ensures that the aggregation function's output cannot be distinguished from the fact that a particular participant is involved in the training process, with a probability limited by ∈. Here, users must have confidence in the server to 1) accurately update their models and 2) execute perturbation (noise addition, etc.) appropriately. Though confidence in the server is required, it is much less strong than giving the server actual facts. To contrast, compared to having data exposed, the privacy risk of deducing training set membership or attributes from model updates is far lower. On top of that, users in FL don't often disclose whole datasets for a reasons of efficiency concerns and potential legal or regulatory constraints.

The CDP method for FL, as described in (McMahan et al., 2017) and (Geyer et al., 2017), is used in this work and is shown in Algorithm 2. After participants' updates are clipped, the server aggregates them and adds Gaussian noise. Then it clips the l2 norm of the aggregate. Overfitting to updates from any participant is prevented by this. As mentioned in (Abadi et al., 2016), the moments accountant approach may be used to monitor the expenditure on privacy.

**Algorithm 2: CDP in FL**

Main ():
Initialization: Model $\theta_0$, Moment Accountant (ϵ, N) //*N is the* sample numbaes
**for** *all epoch r =1*, 2, … *do*
$C_r \leftarrow$ Participants are randomly selected with probability q
$P_r \leftarrow$moment_Accountant.get-privacy-spent () //*Returns the spent privacy budget(SPB)*
**if** $p_r > T$ // *If SPB>*
*Threshold, return current model*
**then**
return $\theta_r$
**end**
**for** *all samples k* $\epsilon$ $C_r$ do
$\Delta r_{k+1} \leftarrow$ Updated participants (*k, θ*) *// Parallel Process*
**end**
$S \leftarrow$*bound*
$Z \leftarrow$ *noise level*
$\sigma \leftarrow z\ S/q$
$\theta r_{+1} \leftarrow \theta_r + \Sigma_{i=1}^{Cr} \Delta_i^{r+1}/c_r + N(O, I\sigma^2)$
Moment_ Accountant.accumulate_spent_privacy($z$)
**end**
**return**

**Function** Updated participants ($k,\theta_r$):
$\theta \leftarrow \theta_r$
**for** *all local iteration i from 1 to E* ***do***
**for** *each batch b* $\epsilon$ *B* ***do***
$\theta \leftarrow \theta - \eta \nabla L(\omega; b$
$\Delta \leftarrow \theta - \theta_r$
$\theta \leftarrow \theta_0 + \Delta \min\ (1, \frac{S}{||\Delta||2})$
**end**
**end**
**return** $\theta - \theta_r$

# 4 Methodology

The Local Central Differential Privacy (LCDP) structure is described in depth in this section. It allows the involvement of different dataset aspects while guaranteeing data security and privacy without deleting data. By establishing both model and data privacy, LCDP bridges the gap between LDP and CDP models. Without introducing excessive risk from adding too much personal data into the Federated training process, the framework guarantees compliance with defined privacy criteria.

Data training in FL in the LCDP model is based on the assumptions about LDP and CDP. With CDP, everyone gives sensitive information to the data collector on the premise that they are trustworthy. Data collectors use algorithms that meet certain levels of differential privacy to answer to queries sent by data analysts. Data analysis for training is then provided the perturbed data after participant data has been pooled. The data collector's unreliability is more realistically assumed in an LDP context. The data collector transfers the differential privacy algorithm to each participant to ensure data privacy. After each participant applies the differential privacy method to their own data, they submit the modified data to the collector. In a similar vein, data collectors answer to queries launched by data analysts by using the data they have already gathered. While training data in FL, the proposed structure employs the modified perturbational central and local forms to generate models and ensure data privacy. Figure 3 shows the LCDP model's framework.

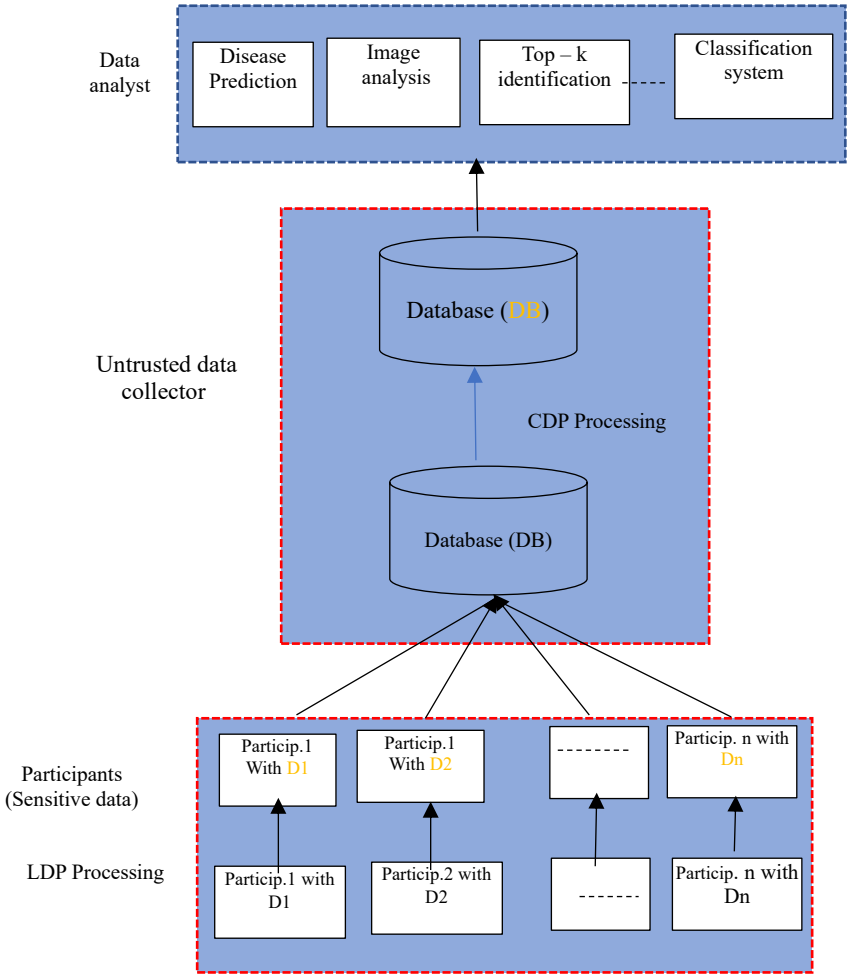



Figure 3: Proposed Model of LCDP for Enhanced Privacy in FL

### 1) Perturbation Mechanism

Here, we extensively analyse the drawbacks of conventional central perturbation techniques and the FL models presented in Section 3, before presenting our method's input perturbation proposal. The 'data centre,' shown in Figure 1, is constantly pre-populated with original data before model training. Following these three conventional perturbation techniques of central CDP leaves original data vulnerable, leading to the assumption that the 'data centre' can be trusted. In spite of this, "data centre" isn't exactly a "trust"-worthy establishment, what with all the potential monitoring and the fact that opponents are always planning to "take" the original data. Consequently, the model parameters are very vital, but the security of the original data instances is even more crucial. To provide differential privacy over communications (data exchanges) between people and the "data centre," LDP is an improved method of addressing the "untrusted data centre" issue. The performance will inevitably be more difficult than central models because to the enormous amount of noise introduced to the data. To address these issues, we provide a novel input perturbation technique that involves introducing noise into data instances and then using these "perturbed data instances" to train a machine learning model. This approach causes the objective function which is given in Equation (3).

$$\hat{L}(\theta) = \frac{1}{n}\sum_{i=1}^{n} l(\theta, x_i + z, y_i) \quad (3)$$

We represent the objective function of input perturbation as $\hat{L}(\theta)$ in order to differentiate it from the objective function without privacy concern, $L(\theta)$, in equation (3). For the purpose of differentiating between the original and perturbed data, the "noise adding" in (8) has already been executed. Attaining $(\epsilon, \delta)$ −differential privacy on the machine learning model while maintaining some kind of privacy on the original data is the main objective of our approach. This means that people's "true original data" is protected against certain types of attacks, even if the "data centre" isn't trustworthy or controlled, since the data that is "taken away" by attackers is accompanied by random noise. We vary from the local model in that we prioritize protections between people and the 'server,' and we do not address the privacy of the model parameters, even if our technique involves adding noise to the original data. Instead, we concentrate on the $(\epsilon, \delta)$ −differential privacy of the final model. In order to ensure the quality of the machine learning model, we compromise certain individuals' privacy in order to get better performance, in comparison to LDP and the input perturbation approach in (Odeh & Taleb, 2023). When contrasted with (Odeh & Taleb, 2023)., the cost really isn't that great. That is to say, we do our best to protect the confidentiality of the original data while still concentrating on maintaining outstanding performance. It is clear that our data is far less noisy than that of LDP and the prior input perturbation approach. Although our approach's privacy protections aren't as robust as those of LDP and the prior input perturbation method, they're still better than those of central methods. Algorithm 3 contains specifics about our procedure.

**Algorithm 3** Perturbation Method
**Require:** *D*- Dataset, learning rate $\alpha$, iteration T,
1: **function** INPUT PERTURBATION (T,$D, \alpha$)
2: All input samples$(x_i, y_i)$ in D, add noise z into D:
3: $(x_i, y_i) \leftarrow (x_i + z, y_i)$.
4: Perturbed data $(x_i + z, y_i)$ is denoted as 'New data'.
5: Perturbed data is trained using Equation (4)
6: **for** t = 0 to T – 1 **do**
7: $\theta_{t+1} \leftarrow \theta_t - \alpha \frac{1}{n}\Sigma_{i=1}^{n} \nabla L(\theta_t)$.
8: **end for**
9: return $\theta_r$.
10: **end function**

Algorithm 3 uses independently sampled random noise $z \sim R^d$ and elements $z_i \sim N(0, \sigma^2)$. It is evident from line 7 of Algorithm 1 that the gradient is affected by the noise introduced to the original data. In addition, we establish a connection between both local and central differential privacy by noting that our method introduces noise into the original data instances, which perturbs the gradient and, in turn, the model parameters. Input perturbation safeguards all three at once, providing a higher level of privacy than conventional central perturbation methods without compromising either theoretical or practical outcomes. We improve performance when compared to LDP and the earlier input perturbation approach, though at the cost of certain individuals' privacy.

### 2) Privacy of LCDP Model

Our proposed approach, input perturbation has privacy assurances set will be discussed in this section. We examine the $(\epsilon, \delta)$ −differential privacy of LCDP proposed model: perturbation method is described in in Algorithm 3 in this section. Here, calculates $(\epsilon, \delta)$ −differential privacy using the Gaussian model used in (Voigt & Von Dem Bussche, 2017) and the moment accountants used in (Jordan & Mitchell, 2015) in this research. It is also assumed in (Sindhusaranya et al., 2023) that $\ell(\theta, x, y)$ is equal to $\ell(y\theta^T x)$.

Statement 1. Assuming $\epsilon, \delta > 0$, and $\ell(\theta, x, y)$ is G-Lipschitz and $\Delta -$ strongly convex over θ, for any constant c, it is $(\epsilon, \delta)$-differential privacy in Algorithm 1. Equation (4) explains the details.

$$\sigma^2 = c\frac{G^2 T \log(1/\delta)}{n(n-1)\sqrt{\Delta}\varepsilon^2} \quad (4)$$

Compared to the gradient perturbation technique suggested in (Zhang et al., 2017), our approach adds about the same amount of noise to data instances. The discrepancy is the product of a constant and a factor of $(n-1)\sqrt{\Delta}/n$. Our noise bound outperforms the conventional gradient perturbation approach suggested in (Liu et al., 2021) by a factor of up to $n^4\frac{log(n)}{T}$. This outcome is feasible since our technique's noise bound is far better than LDP's. This is because LDP protects users' privacy between themselves and the "server," whereas proposed method focuses more on protecting the privacy of the final ML model. Our approach is comparable to gradient perturbation technique in that both share the same observation: local and central differential privacy are bridged by perturbing gradients, which in turn perturb original data. The outcome is that our proposed input perturbation approach accomplishes $(\epsilon, \delta)$ −DP on the final model by means of this interface. To provide a more trustworthy degree of privacy protection in the FL field, our solution concurrently maintains the privacy of the original data instances, the gradient, and the model parameters.

## 5 Result and Discussion

### 1) Dataset Description

In the current research, we evaluated the proposed model's performance using three separate medical datasets. This section makes use of the MIMIC-III, SEER cancer dataset as well as the Chronic Disease Indicators (CDC) dataset. Deidentified health records from more than forty thousand critical care admissions at Beth Israel Deaconess Medical Centre from 2001 to 2012 can be found in the massive, publicly available Medical Information Mart for Intensive Care (MIMIC-III). The breast cancer SEER dataset originated from the National Cancer Institute's SEER Program, which offers data on cancer rates

based on population-based estimates, and was last updated in November 2017. Invasive breast cancer in female patients diagnosed between 2000 and 2017 was the subject of the dataset. Patients' ages, races, ethnicities, cancer stages, tumor sizes, grades, and treatments are all part of the dataset. Important for public health practice, the 124 indicators provided by the CDC's Division of Population Health enable states, territories, and major metropolitan areas to consistently define, gather, and report data on chronic diseases. These indicators were developed through consensus and are accessible to all three levels of government.

The applications were simulated on a personal computer with 16.0 GB of RAM and an Intel(R) Core (TM) i3-12100 3.30GHz CPU, which was built in Python. Intel(R) Core (TM) i5-7200U CPUs @ 2.50GHz and 2.70GHz were used by the participants in the simulation using Lenovo 310s laptops. We used a learning rate of 0.01 and a batch size of 128 samples per group for gradient descent.

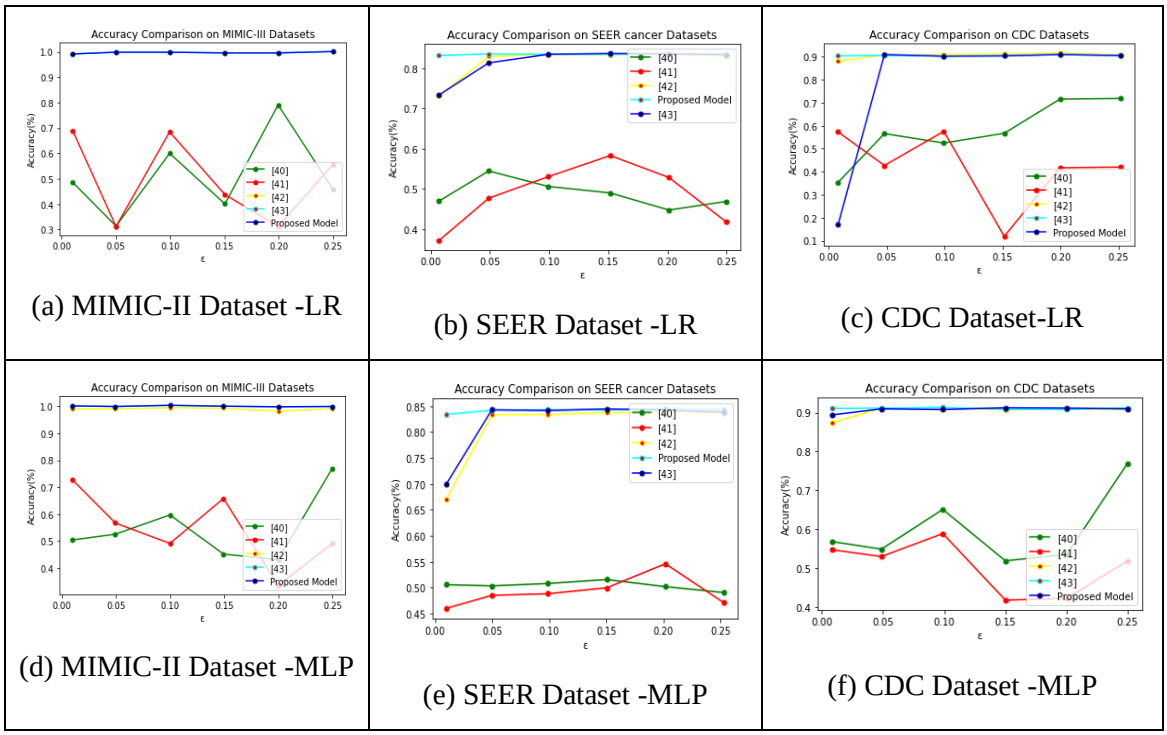
**2) Evaluation Metrics**

Accuracy and the optimality gap, which is represented as $L(\theta_{priv}) - L^*$, are indicators of performance. Performance on test data is represented by accuracy, whereas optimum gap indicates excess empirical risk on training data (Di et al., 2018). In the evaluation section, a DL model is represented by an MLP with an input layer and a hidden layer of the same size. At random, we choose both the training and testing sets. The cross-validation method is used to choose T and α in every trial. $\varepsilon$ the differential privacy budget, is set between 0.01 and 0.25, and its influence is assessed. The number of datasets determines δ, which may be considered a constant in the meanwhile (Ficek et al., 2021). To clarify, $d$ is less than p in deep learning models but equals p in logistic regression models.

Our proposed technique outperforms the gradient perturbation method from (Bassily et al., 2014) and the objective perturbation method from (Kifer et al., 2012) in terms of accuracy, as shown in Figure 4. Whether we're dealing with an LR or MLP model, our approach is almost identical to the output perturbation technique from (Zhang et al., 2017) and the gradient perturbation method from (Di et al., 2017) in terms of accuracy. But, when the method's gradient is enhanced by Gaussian noise, which has a huge variation (Bassily et al., 2014), The optimality gap has been measured by using Equation (5).

$$O\left(\frac{G^2 n^2 \log\left(\frac{n}{\delta}\right) \log\left(\frac{1}{\delta}\right)}{\epsilon^2}\right) \quad (5)$$

Figure 5 shows that, contrary to the theoretical analysis, our technique outperforms the other methods discussed above on most datasets, and its optimality gap is almost identical to that of the output perturbation method suggested in (Di et al., 2017). Furthermore, it is evident that our strategy achieves almost identical performance to the (Di et al., 2017). model without privacy concern in some cases, on both the LR and MLP models, since the optimality gap is close to zero on certain datasets. Moreover, due to its noise constraint, the optimality gap of the gradient perturbation approach suggested in (Bassily et al., 2014) changes rapidly, much as the accuracy in Figure 4. Figure 4 shows that the method's accuracy across ϵ varies dramatically.

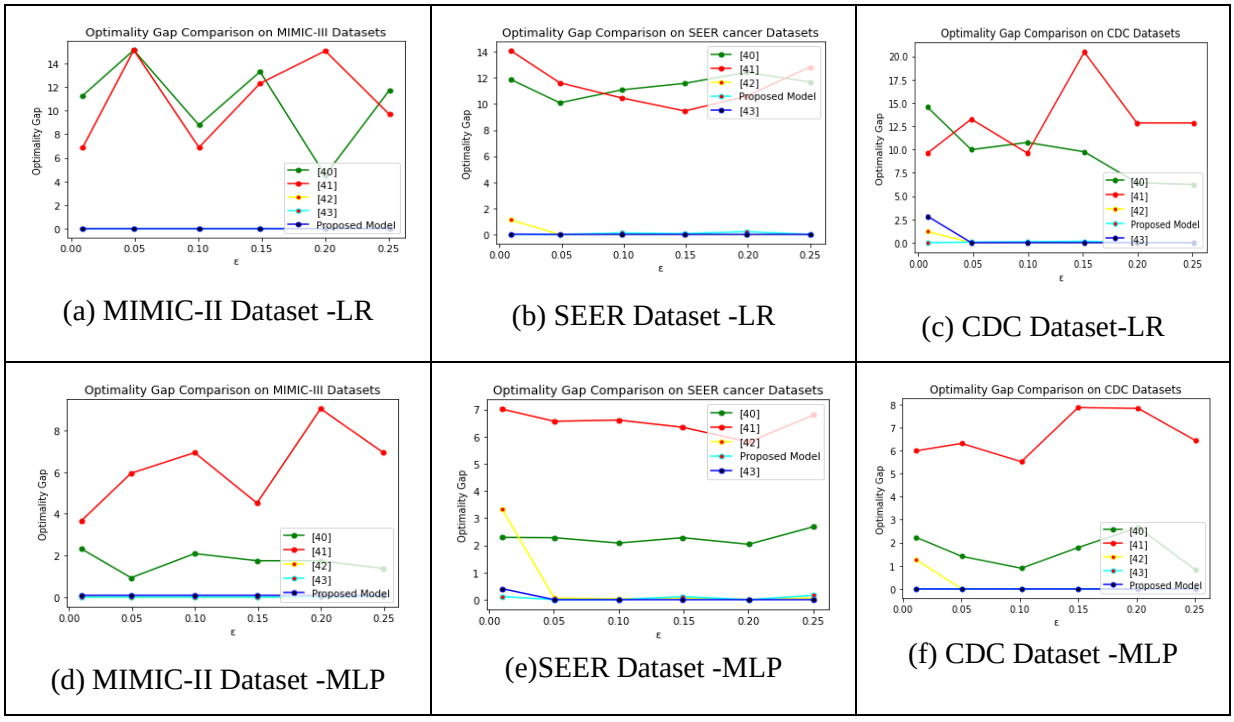



(a) MIMIC-II Dataset -LR (b) SEER Dataset -LR (c) CDC Dataset-LR

(d) MIMIC-II Dataset -MLP (e) SEER Dataset -MLP (f) CDC Dataset -MLP

Figure 4: Performance Comparison of Accuracy on Various Dataset Vs Classifiers

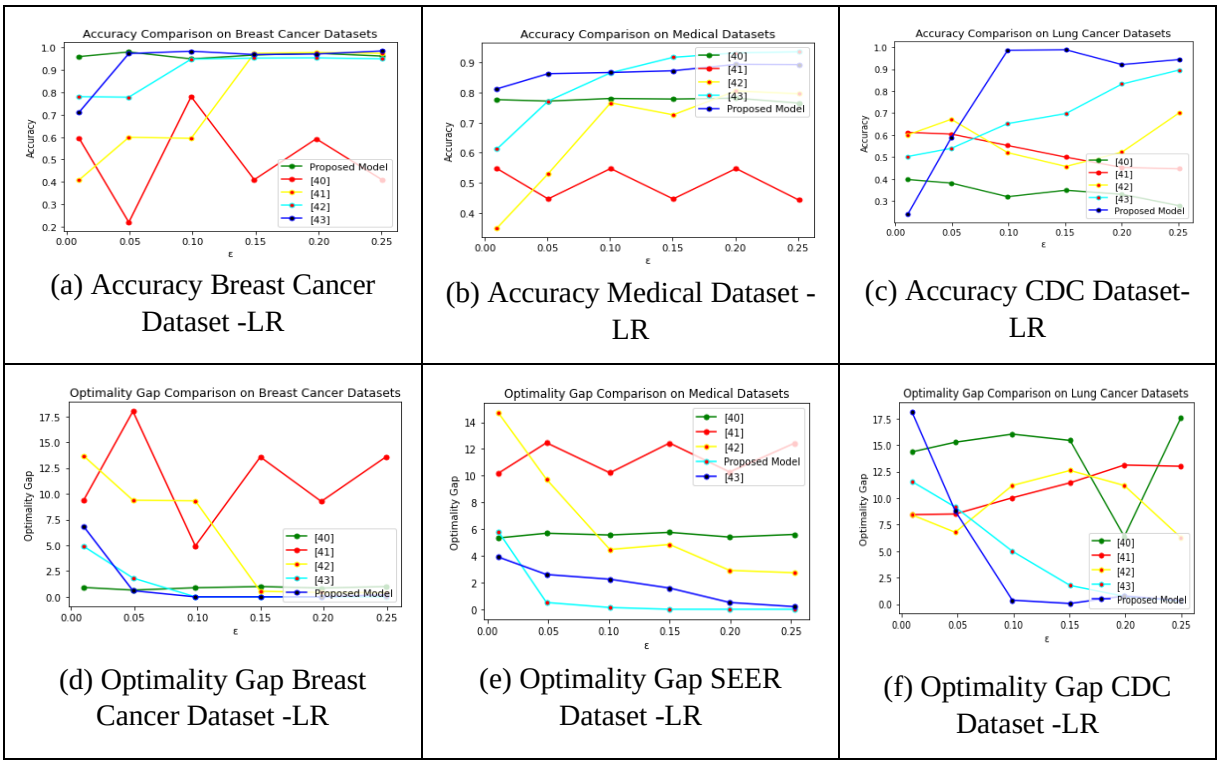
The accuracy and optimality gap are shown in Figure 6 on small datasets (with sizes fewer than 1000) when the logistic regression model is the only model that is deployed. Consistent with Figures 4 and 5, our technique seems to be advantageous in the majority of situations.




(a) MIMIC-II Dataset -LR (b) SEER Dataset -LR (c) CDC Dataset-LR

(d) MIMIC-II Dataset -MLP (e)SEER Dataset -MLP (f) CDC Dataset -MLP

Figure 5: Performance Comparison on Optimality Gap on Various Dataset and Classifiers (LR & MLP)




(a) Accuracy Breast Cancer Dataset -LR

(b) Accuracy Medical Dataset - LR

(c) Accuracy CDC Dataset-LR

(d) Optimality Gap Breast Cancer Dataset -LR

(e) Optimality Gap SEER Dataset -LR

(f) Optimality Gap CDC Dataset -LR

Figure 6: Performance Comparison on Accuracy and Optimality Gap on Various Dataset with Є

Based on our analysis of the experimental results, we can say that our method outperforms the gradient perturbation method from (Bassily et al., 2014). and the objective perturbation method from (Kifer et al., 2012). This is due to the initial's loose noise bound and the more restrictive objective perturbation method, respectively, although there are small differences in evaluation results across datasets. When compared to more conventional machine learning models, such as logistic regression, the experimental outcomes for the deep learning model (MLP) are quite comparable. Our solution offers greater privacy without sacrificing efficiency compared to earlier central methods, and it retains the original data, gradient, and final model all at once. It's a promising result.

### 3) Performance Evaluation of LCDP Model

The accuracy of the proposed model's classification on the SEER dataset employing NB classifiers is shown in Figure 7. It's easy to use the LCDP model with a wide variety of machine learning models and datasets. We have assumed that removing certain characteristics might enhance the model's performance at the data perturbation stage.

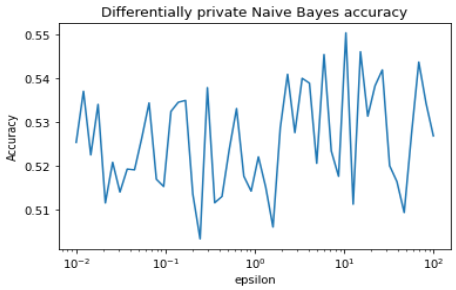



Figure 7: NB Classifier-based Accuracy Calculation on SEER Dataset

There is an increase in noise in the SEER medical datasets, as seen in Figure 8. Before sending data from a user to a model, this method introduces noise into the data at the source. Since the model only ever sees the noisy data, this strategy ensures that each individual's dataset contribution remains private.

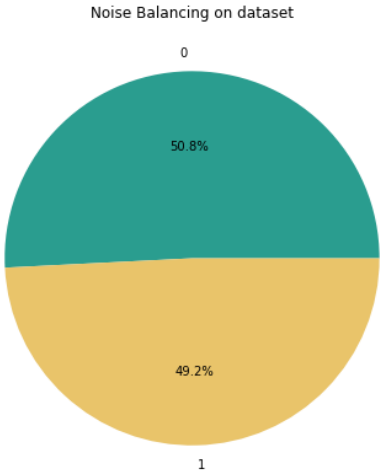



Figure 8: Noise Balancing on SEER Dataset

Figure 9 displays the proposed model-based data visualization with perturbations. By its very nature, LCDP is an innovation in the field of privacy protection. Strong privacy is accomplished in the LCDP model by encoding user data before transmitting it to an unreliable aggregator. The input dataset's security will be heightened by the perturbation.

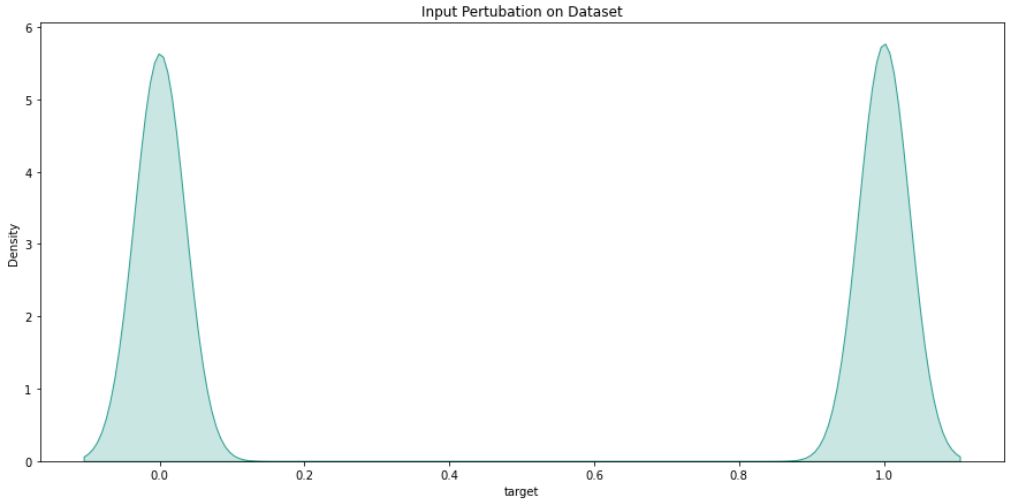



Figure 9: Input Perturbation on SEER Dataset

The specifics of our proposed model's aggregated features are shown in Figure 8. Multicollinearity, such as a highly correlated set of explanatory factors, is a potential issue with aggregation. The existence of multicollinearity is not always indicative of an inadequately described model, as should be pointed out. A more bell-shaped distribution seems to represent the aggregated characteristics. It is recommended to user to perturb features before adding them to a model and then track how well the model performs. Aggregated Feature Distribution on SEER Dataset shown in Figure 10.

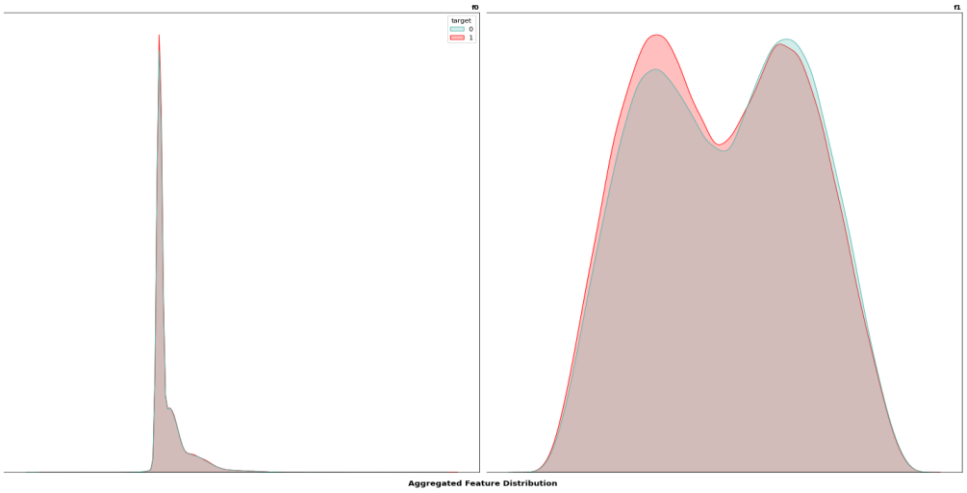



Figure 10: Aggregated Feature Distribution on SEER Dataset

## 6 Conclusion

In order to solve the problem of both central and local DP models in Federated training, this work introduces LCDP, a new privacy-preserving federated learning approach. The input perturbation approach is examined in this research by training the machine learning model using 'perturbed data,' which consists of original data instances that have been perturbed data with Gaussian noise. Enhanced DP on a final ML system along with a certain form of user data privacy are achieved by bridging the gap between central and LDP and by recognizing that input perturbation causes perturbation on the gradient. We are able to concurrently maintain the final machine learning model, the gradient, and the original data due to the "bridge concept of LDP and CDP". The experimental findings show that LCDP can handle different types of federated datasets, and when compared to other methods, our proposed approach performs better overall. In addition, future study will provide careful examination of the privacy maintained on user data using our technique, as well as ways to enhance user privacy.

## References

[1] Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016, October). Deep learning with differential privacy. *In Proceedings of the ACM SIGSAC Conference on Computer and Communications Security*, pp. 308-318. https://doi.org/10.1145/2976749.2978318

[2] Adnan, M., Kalra, S., Cresswell, J. C., Taylor, G. W., & Tizhoosh, H. R. (2022). Federated learning and differential privacy for medical image analysis. *Scientific Reports*, *12*(1), 1953. https://doi.org/10.1038/s41598-022-05539-7

[3] Agrawal, R., & Srikant, R. (2000). Privacy-preserving data mining. *In Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, 439-450.

[4] Annas, G. J. (2003). HIPAA regulations: a new era of medical-record privacy?. *New England Journal of Medicine*, *348*, 1486-1490, https://doi.org/10.1056/NEJMlim035027

[5] Bassily, R., Smith, A., & Thakurta, A. (2014). Private empirical risk minimization: Efficient algorithms and tight error bounds. *In IEEE 55th Annual Symposium on Foundations of Computer Science*, 464-473.

[6] Beimel, A., Nissim, K., & Omri, E. (2008). Distributed private data analysis: Simultaneously solving how and what. *In Advances in Cryptology–CRYPTO 2008: 28th Annual International Cryptology Conference, Santa Barbara, CA, USA,* 451-468.

[7] Bobir, A.O., Askariy, M., Otabek, Y.Y., Nodir, R.K., Rakhima, A., Zukhra, Z.Y., Sherzod, A.A. (2024). Utilizing Deep Learning and the Internet of Things to Monitor the Health of Aquatic Ecosystems to Conserve Biodiversity. *Natural and Engineering Sciences, 9*(1), 72-83.

[8] Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., & Seth, K. (2017). Practical secure aggregation for privacy-preserving machine learning. In *proceedings of the ACM SIGSAC Conference on Computer and Communications Security*, 1175-1191.

[9] De Cristofaro, E. (2021). A critical overview of privacy in machine learning. *IEEE Security & Privacy*, *19*(4), 19-27.

[10] Demelius, L., Kern, R., & Trügler, A. (2023). Recent Advances of Differential Privacy in Centralized Deep Learning: A Systematic Survey. *arXiv preprint arXiv:2309.16398*. https://doi.org/10.48550/arXiv.2309.16398

[11] Di, W., Marco, G., & Jinhui, X. (2018). Empirical risk minimization in non-interactive local differential privacy revisited. *In Advances in Neural Information Processing Systems*, 965–974.

[12] Di, W., Minwei, Y., & Jinhui, X. (2017). Differentially private empirical risk minimization revisited: Faster and more general. *In Advances in Neural Information Processing Systems*. 2722–2731.

[13] Duchi, J. C., Jordan, M. I., & Wainwright, M. J. (2013). Local privacy, data processing inequalities, and statistical minimax rates. *arXiv preprint arXiv:1302.3203*. https://doi.org/10.48550/arXiv.1302.3203

[14] Dwork, C., & Feldman, V. (2018). Privacy-preserving prediction. *In Conference on Learning Theory*, 1693-1702.

[15] Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, *9*(3–4), 211-407.

[16] Ficek, J., Wang, W., Chen, H., Dagne, G., & Daley, E. (2021). Differential privacy in health research: A scoping review. *Journal of the American Medical Informatics Association*, *28*(10), 2269-2276. https://doi.org/10.1093/jamia/ocab135

[17] Fleming, N. (2018). How artificial intelligence is changing drug discovery. *Nature*, *557*(7706), S55-S55.

[18] Geyer, R. C., Klein, T., & Nabi, M. (2017). Differentially private federated learning: A client level perspective. *arXiv preprint arXiv:1712.07557*. https://doi.org/10.48550/arXiv.1712.07557

[19] Haidar, M., & Kumar, S. (2021). Smart healthcare system for biomedical and health care applications using aadhaar and blockchain. *In IEEE 5th International Conference on Information Systems and Computer Networks (ISCON)*, 1-5.

[20] Hill, S., Zhou, Z., Saul, L., & Shacham, H. (2016). On the (in) effectiveness of mosaicing and blurring as tools for document redaction. *Proceedings on Privacy Enhancing Technologies*, 4 (2016) 403–417.

[21] Jarin, I., & Eshete, B. (2022). Dp-util: Comprehensive utility analysis of differential privacy in machine learning. *In Proceedings of the Twelfth ACM Conference on Data and Application Security and Privacy*, 41-52. https://doi.org/10.1145/3508398.3511513

[22] Ji, Z., Lipton, Z. C., & Elkan, C. (2014). Differential privacy and machine learning: a survey and review. *arXiv preprint arXiv:1412.7584*.

[23] Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255-260.

[24] Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., & Zhao, S. (2021). Advances and open problems in federated learning. *Foundations and Trends® in Machine Learning*, *14*(1–2), 1-210.

[25] Kifer, D., Smith, A., & Thakurta, A. (2012). Private convex empirical risk minimization and high-dimensional regression. *In Conference on Learning Theory*, JMLR Workshop and Conference Proceedings.

[26] Lee, J., & Kifer, D. (2021). Scaling up differentially private deep learning with fast per-example gradient clipping. *Proceedings on Privacy Enhancing Technologies*, 128–144. https://doi.org/10.2478/popets-2021-0008

[27] Liu, B., Ding, M., Shaham, S., Rahayu, W., Farokhi, F., & Lin, Z. (2021). When machine learning meets privacy: A survey and outlook. *ACM Computing Surveys (CSUR)*, *54*(2), 1-36.

[28] Liu, H., Peng, C., Tian, Y., Long, S., Tian, F., & Wu, Z. (2022). GDP vs. LDP: A survey from the perspective of information-theoretic channel. *Entropy, 24*(3), 430. https://doi.org/10.3390/e24030430

[29] Malekzadeh, M., Hasircioglu, B., Mital, N., Katarya, K., Ozfatura, M. E., & Gündüz, D. (2021). Dopamine: Differentially private federated learning on medical data. *arXiv preprint arXiv:2101.11693*. https://doi.org/10.48550/arXiv.2101.11693

[30] McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. *In Artificial intelligence and statistics*, 1273-1282.

[31] McMahan, H. B., Moore, E., Ramage, D., & y Arcas, B. A. (2016). Federated learning of deep networks using model averaging. *arXiv preprint arXiv:1602.05629*

[32] McMahan, H. B., Ramage, D., Talwar, K., & Zhang, L. (2017). Learning differentially private recurrent language models. *arXiv preprint arXiv:1710.06963*. https://doi.org/10.48550/arXiv.1710.06963.

[33] Mohandas, R., Veena, S., Kirubasri, G., Thusnavis Bella Mary, I., & Udayakumar, R. (2024). Federated Learning with Homomorphic Encryption for Ensuring Privacy in Medical Data. *Indian Journal of Information Sources and Services, 14*(2), 17–23. https://doi.org/10.51983/ijiss-2024.14.2.03

[34] Odeh, A., & Taleb, A. A. (2023). A Multi-Faceted Encryption Strategy for Securing Patient Information in Medical Imaging. *Journal of Wireless Mobile Network Ubiquitous Computing and Dependable Application (JoWUA), 14*(4), 164-176.

[35] Papernot, N., Abadi, M., Erlingsson, U., Goodfellow, I., & Talwar, K. (2016). Semi-supervised knowledge transfer for deep learning from private training data. *arXiv preprint arXiv:1610.05755*. https://doi.org/10.48550/arXiv.1610.05755

[36] Pati, S., Baid, U., Zenk, M., Edwards, B., Sheller, M., Reina, G. A., & Bakas, S. (2021). The federated tumor segmentation (fets) challenge. arXiv preprint arXiv:2105.05874. https://doi.org/10.48550/arXiv.2105.05874

[37] Peter, K., Sewoong, O., & Pramod, V. (2014). Extremal mechanisms for local differential privacy. *In Advances in neural information processing systems*, 2879–2887.

[38] Pfohl, S.R., Dai, A.M., & Heller, K. (2019). Federated and differentially private learning for electronic health records. *arXiv:10.48550/arXiv.1911.05861*

[39] Sadilek, A., Liu, L., Nguyen, D., Kamruzzaman, M., Serghiou, S., Rader, B., & Hernandez, J. (2021). Privacy-first health research with federated learning. *NPJ Digital Medicine*, *4*(1), 132-138, https://doi.org/10.1038/s41746-021-00489-2

[40] Shen, Z., & Zhong, T. (2021). Analysis of application examples of differential privacy in deep learning. *Computational Intelligence and Neuroscience*, *2021*(1), 4244040. https://doi.org/10.1155/2021/4244040

[41] Sindhusaranya, B., Yamini, R., Manimekalai Dr, M. A. P., & Geetha Dr, K. (2023). Federated Learning and Blockchain-Enabled Privacy-Preserving Healthcare 5.0 System: A Comprehensive Approach to Fraud Prevention and Security in IoMT. *Journal of Internet Services and Information Security, 13*(4), 199-209.

[42] Tayebi Arasteh, S., Ziller, A., Kuhl, C., Makowski, M., Nebelung, S., Braren, R., & Kaissis, G. (2023). Private, fair and accurate: Training large-scale, privacy-preserving AI models in medical imaging. *arXiv e-prints*, *arXiv-2302*. https://doi.org/10.48550/arXiv.2302.01622

[43] Udayakumar, R., Anuradha, M., Gajmal, Y. M., & Elankavi, R. (2023). Anomaly detection for internet of things security attacks based on recent optimal federated deep learning model. *Journal of Internet Services and Information Security, 13*(3), 104-121.

[44] Udayakumar, R., Suvarna, Y.P., Yogesh, M.G., Vimal, V.R., & Sugumar, R. (2023). User Activity Analysis Via Network Traffic Using DNN and Optimized Federated Learning based Privacy Preserving Method in Mobile Wireless Networks. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 14*(2), 66-81.

[45] Voigt, P., & Von Dem Bussche, A. (2017). The eu general data protection regulation (gdpr). *A Practical Guide, 1st Ed., Cham: Springer International Publishing*, *10*(3152676).

[46] Wang, D., Smith, A., & Xu, J. (2019). Noninteractive locally private learning of linear models via polynomial approximations. *In Algorithmic Learning Theory*, 898-903.

[47] Zhang, J., Zheng, K., Mou, W., & Wang, L. (2017). Efficient private ERM for smooth objectives. *arXiv preprint arXiv:1703.09947*.

[48] Zhao, J., Chen, Y., & Zhang, W. (2019). Differential privacy preservation in deep learning: Challenges, opportunities and solutions. *IEEE Access*, *7*, 48901-48911. https://doi.org/10.1109/ACCESS.2019.2909559

[49] Zhou, Z. H. (2016). Learnware: on the future of machine learning. *Frontiers of Computer Science, 10*(4), 589-590.

[50] Ziller, A., Usynin, D., Remerscheid, N., Knolle, M., Makowski, M., Braren, R., ... & Kaissis, G. (2021). Differentially private federated deep learning for multi-site medical image segmentation. *arXiv preprint arXiv:2107.02586*. https://doi.org/10.48550/arXiv.2107.02586

## Authors Biography

**Cina Mathew,** "Experienced educator and researcher with 14 years of teaching experience. Holding a Master's degree in Computer Science and Engineering, she is currently pursuing a PhD in Machine Learning. Her passion lies in exploring innovative teaching methods and staying updated on the latest advancements in AI and ML. With a strong academic background and extensive teaching experience, she is dedicated to shaping the next generation of tech enthusiasts. She is excited to collaborate and contribute to the field of machine learning, driving progress and excellence in education and research."

**Dr.P. Asha,** is an Professor of Computer Science and Engineering Department, Sathyabama Institute of Science and Technology, Chennai. She has 10+ years of working experience in Teaching and Research. Her area of Specialization includes Machine Learning, Internet of Things, Artificial Intelligence, Parallel Computing, Computer Graphics, Visualization and Data Mining. She has 100+ research publications and 60+ in Scopus/ Web of Science Journals and Conferences along with 6 Granted Patents and Copyrights. She serves as a reviewer for many Scopus and Web of Science Journals with High Impact factor. She has also published books on Machine Learning, Web of Things, Data Analytics with Python and Industry 4.0 support by Machine Learning.