Volume-1, Issue 3, October 2026, Page No- 32-44
Journal of Machine Learning Innovations and Artificial Intelligence Horizons (JMLIAIH)
ISSN: 3140-0159

Privacy-Preserving Federated Learning for Healthcare Analytics: Integrating Differential Privacy and Secure Aggregation for Predictive Modeling

Govinda Sahu

Assistant Professor, Department of CS& IT, ISBM University, Chhattisgarh, India.

DOI: https://doi.org/10.70903/jmliaih.2026.1303

Article Received: 08-09-2026
Revised Version: 30-09-2026
Accepted: 23-10-2026

*Corresponding Author

Govinda Sahu

E-Mail Id: govindasahu1711@gmail.com

Copyright: © The Author(s), 2026. Published by Notation Trendplus Publishing Pvt. Ltd. This is an Open Access article, distributed under the terms of the Creative Commons Attribution 4.0 License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

Abstract: Healthcare data is often unique in that privacy laws often restrict the ability to share patient information centrately, and as a result, there's an increasing need for predictive healthcare analytics often limited by privacy regulations. We propose a federated learning based distributed system that exploits both the differential privacy technique and the secure aggregation to cooperatively train a binary disease classification system, without sharing private valuable patient information, across multiple hospitals. Each participating institution does a local stochastic gradient descent, and applies gaussian noise (clipping) to the updates as well as satisfies DP constraints in each round of communication. Additionally, protocols have been developed to allow for secure aggregation such that the central server only be able to determine the contribution as a sum of the individual contributions and not the contribution from any one hospital. The server then averages (federated averaging) each local model with its weighting based on the number of samples and gets an enhanced global model. A two-layered neural network with a sigmoid activation function output is trained using Adam and binary cross-entropy function for the classification problem.For the classification problem, a two-layered neural network with a sigmoid activation function output is trained with the Adam optimizer and binary cross-entropy function. This was enough security to make it difficult for the server or others to be able to "reconstruct" the patient-level data that was sent. In summary, the proposed principle is seemly practical and can be usefully used to collaboratively construct models across clinical settings, where privacy safeguards and prediction performance logic dictated in a delicate tradeoff fashion. What we provide is a system on how to integrate such complementary privacy mechanisms into a single federated pipeline, but that goal extends beyond the case where data cannot be shared out of the box as we did above.

Keywords: Federated Learning; Differential Privacy; Secure Aggregation; Healthcare Analytics; Predictive Modeling.

Introduction

While there are clearly strong and promising opportunities in the proliferation of electronic health records to uncover clinical information from enormous collections of patient data, patient information data must be regulated under law and ethics to be able to analyse and provide information without considering the privacy of patients. There are limited and/or poor-quality training data on which to base these models because hospitals are hesitant to share raw data with centres. To answer this conundrum, a new promising solution has emerged: Distributed training for a shared model (McMahan et al., 2017), in which multiple institutions can train a shared model without downloading the local data in the hostals. This flagship Federated Averaging (FedAvg) algorithm instead of data samples exchange, reduces the communication overhead by aggregating locally computed updates and has been extended to many healthcare applications (Ali et al., 2022).

But federated learning does not necessarily provide privacy. Sensitive information may be exposed through the process of model updates which are sent during the training procedure as in gradient inversion attacks and membership inference attacks (Mu, 2022) (Ye et al., 2022). This vulnerability also becomes extremely serious in the clinic as even an incorrect approximation to the aim of therapy can be regarded as critical violation. Differential privacy (DP) provides a mathematically precise notion of quantifying and bounding such leakage (Bernstein & Sheldon, 2018) and a gradient clipping and adding calibrated noise modification of the SGD algorithm is known as DP-SGD and has been widely adopted as a form of defence (Martí et al., 2016). The secure aggregation protocols, on the other hand, can only depend on cryptographic techniques such as secret sharing for which the server may gain access to only one of the aggregate updates and not one of the individual client updates (Kairouz, 2019). The options are free – DP deduction is deductible from the end model while secure aggregation protects the updates from in-transit and transmission.

Often data within clinical data distributions is not "identically and independently distributed" across hospitals. The statistical heterogeneity created by multi-demic over heaters and lack of diagnostics and reporting slow down the convergence and make the global model biased (Lin & Wen, 2024), (Karimireddy et al., 2019). Moreover, the privacy-utility compromise in medical diagnosis has consequences and it is quite pronounced, so that a small inaccuracy can have clinical implications. The majority of the current literature on privacy-preserving healthcare analytics focuses on federated learning, differential privacy and/or secure aggregation individually, or merely relates a mixed approach at a high level without specifying a particular training pipeline (Kaissis et al., 2020) (Welten et al., 2022). Combining noise calibration and cryptographic masking with techniques of heterogeneous optimization in a single fed-back loop of training is still not well explored. Thakur et al. (2023) reviewed various soft computing techniques, including fuzzy logic, artificial neural networks, machine learning, and genetic algorithms, for cardiovascular disease prediction, diagnosis, and monitoring.

We propose a novel computational framework: FedAvg in conjunction with differential privacy and secure aggregation to perform binary disease classification in a distributed set of hospitals without compromising the privacy of the samples. Our approach is novel because we first "perturb" (modify) and clip locally provided gradients to possibly participate in an aggregatory which can only process a "blinded" (cryptographically protected) version of the gradient, one also protected using differential privacy. This is a different approach than the aforementioned designs where DP is applied to aggregated systems or where privacy accounting is done informally without aggregation. Furthermore, we use a sample size weighted average to average the drawing results, and test the framework with different data structures that are obtained from clinical data sets using artificial neural networks to incorporate the non-IID data. The main lessons that can be drawn from this work are: (1) the pipeline is a solid one that can be used to establish secure aggregation under the assumption of a shared privacy budget, (2) each

of the different strategies to achieve the required privacy protection comes with its own cost in terms of utility, and (3) rigorous privacy guarantees can be maintained without requiring clinical fidelity of predictive capability that is dramatically put aside.

In Section 2, past research involving federated learning, DP, and secure aggregation in the context of healthcare is summarized. In Section 3, we summarize and introduce some important concepts: FedAvg, DP-SGD, and secure aggregation. A secure federated averaging framework that is improved by DP is presented in section 4. Details of the experimental setup are given in section 5 and the result and discussion in the section 6. The paper ends in Section 7.

Related Work

Federated optimization, formal study of privacy accounting for such collaborative analytics, and cryptographic protection of model update exchange are all areas of study which are complementary to this area. One covers one threat surface and the other covers another—when data can't be centralized in any single place, both are required for clinical applications. This section pulls together the most relevant strands and places our framework in the context of these.

Federated Learning in Healthcare

Federated learning is a paradigm that has been widely applied to multi-institutional medical analytics that enables training a model without sharing data (McMahan et al., 2017). Clean, detailed surveys record how it has been employed for implementing smart healthcare systems, which can provide on-device and cross-silo healthcare training, thus facilitating predictive modelling of healthcare systems without sharing patient records (Ali et al., 2022). Medical imaging is an example in which the federated approach has been shown to execute training near centralized baselines without imposing any limitations on data rights and governance (Kaissis et al., 2020). However, all these studies also state that sharing these parameters is in no way equivalent to privacy protection: all the parameters shared carry an abundance of information about what was being trained locally. Feature selection and ensemble machine learning classifiers have been investigated for improving heart disease prediction (Diwan et al., 2022).

Differential Privacy for Model Updates

A formal, composable concept of privacy is differential privacy, which limits the impact of any individual record on an output or quantity that is published (Bernstein & Sheldon, 2018). Clipped per-example gradients + calibrated Gaussian noise, for provable guarantees at low utility cost (Martí et al., 2016). The scenario with federated system will typically adopt clipping and perturbating full local updates, which results in a safeguard of the whole content of the records kept by any participating institution. The utilization-privacy tradeoff is achieved by using appropriate setting of the clipping threshold and noise scale, with an eye toward achieving clinically acceptable margins of diagnostic accuracy.

Secure Aggregation and Cryptographic Masking

Hides updates issued by all clients from the server, securing the data in the process – by using homomorphic encryption or secret sharing except for the sum (Kairouz, 2019). It could be used on every health property since if the one leaked update is released it can reveal special information. They are already used in the context of distributed tools for health data analytics and cohort selection & model training spanning institutions (Welten et al., 2022). Secure aggregation does protect updates, but it cannot bound the information in the final model (which is the thing that encourages using DP to generate

it); while in the course of sending the updates and the aggregation, it does not prevent updates from being sent or aggregated back to the final model.

Heterogeneity and Optimization

A factor of bias that rarely occurs in a naive "averaging case" is that individual clinical datasets are often not independent and identically distributed across sites. Other developments are the proximal term addressing the local objectives in FedProx (Lin & Wen, 2024) and the use of control variates for correcting for drift caused by clients, in SCAFFOLD (Karimireddy et al., 2019). Such techniques can be applied to optimization outside of the IID data setting, but orthogonal to most privacy techniques with limited research exploring the effect of noise injection/cryptographic masking.

Overlapping the security analysis in these existing attempts, federation, differential privacy or secure aggregation are mostly explored in isolation or in a conceptual manner. We present a new framework that clippies and perturbs local updates prior to the cryptographic masking, resulting in an aggregate that is clipped at the local view and is masked from the start from the server. To this end, the ordering ensures that (i) the privacy budget is used on the client and (ii) an intermediate (plaintext) update is never leaked; capacities that cannot be offered by pipelines applying DP after the aggregation.

Preliminaries

This section is used to lay out the concept and mathematics we use to build on to the ideas of this proposed framework. We start with a discussion of the supervised learning formulation and the optimization machinery that is needed to train local models and then introduce information-theoretic and cryptographic ideas that form the basis of privacy protection. All these prelims put together make it clear what the learning problem and threat model are we are designing for.

Machine Learning Fundamentals and Optimization Theory

In supervised learning, an input-label mapping is sought, using parameters that are typically a mapping from inputs to labels that has something to do with their generalization to the rest of the inputs. Instead, you should assume you need some unknown distribution over the feature space and label space. Let's make the assumption that we already have a given loss function. Formally, we have a pair of functions and assume that is a joint distribution whose precise form is unknown, over the domains and we wish to select a hypothesis that minimizes an expected risk value over for a given (loss) function defined on but the "inaccessible" is replaced with its empirical counterpart for a finite sample and the resulting minimization problem is usually solved with (stochastic) gradient descent (SGD), or its extensions.

Frank net is designed to execute flexible classes of functions in tabular clinical data like Artificial Neural Networks (ANNs). The Generative, Artificial Neural Networks (ANNs), are a generic function class available as part of "Franknet". Affine transformations followed by an element-wise nonlinearity are produced by a feedforward network with hidden layers: Output for input reduces to while represents the activations of layer and an element-wise nonlinear activation function like rectified linear unit (ReLU). For a binary disease classification, the output layer uses the sigmoid function, and the network is trained by the binary cross-entropy loss which is the negative log-likelihood of the Bernoulli model. Usually, the gradient information accumulated by the ADAM optimization algorithm is used, which supports calculations of the first and second moments of all contributions to the gradient per parameter, which can be used to adjust the effective step size (Jimmy & Diederik, 2014). Together these make up a local training procedure for each participating institution which is implemented prior to any privacy approach is put into effect.

Information-Theoretic Security and Cryptographic Primitives

The concept behind a privacy protection is the limitation of the amount of information that adversary can be able to retrieve from the public releases. The vocabulary for this analysis is the tool of information theory. whose probability mass function is represents, by Shannon's formula: [The Shannon entropy of a discrete random variable is given by the formula s(X) = sum H = discrete random variable, phi = probability mass function of H

$$H(X) = -\sum_{x \in X} P(x) \log P(x) \quad (1)$$

and the mutual information between two variables X and Y measures the average reduction in uncertainty about one that is obtained by observing the other,

$$I(X;Y) = \sum_{x \in X} \sum_{y \in Y} P(x,y) \log \frac{P(x,y)}{P(x)P(y)}. \quad (2)$$

Appealingly, the same argument makes the mutual information between the two as the quantity of information the release mechanism provides the input, and suggests that the quantity is zero exactly when no such information is provided. Instead, as noted in the bestowing section, differential privacy puts a cap on the potential damage that can be inflicted on any one record by any individual record using a likelihood ratio, a cap which is robust as a function of auxiliary information and composition.

These statistical properties are supported by cryptographic primitives which are then checked at the level of computation. secreting a secret value into multiple pieces of information, which can be sent to various people, such that no subset of the set of N - 1 parties knows anything about the secret other than a direct share of the secret (i.e., the full set of N parties is the only one that contains a secret share, and every other subset consists of N - 1 parties that know nothing about the secret except for a direct share value), with the set of N parties reconstructing the secret exactly. The simplest form is additive secret sharing over some finite group, and many secure aggregation protocols are based on the concept of summand masking (canalizing out), each of the shares will nullify when summed together leaving only the aggregate of the shares. These tools will allow calculation of a function involving a distributed input without the revelation of the input itself, important in protecting model updates in the transmission of the model, as well as intermediates storage.

Proposed Framework: DP-Enhanced Secure Federated Averaging

Now let's introduce the proposed framework, in which we federation the sensitivity-bounded clipping, Gaussian noise injection, with cryptographic masking. The motivation of the design is that while both DP and Secure Aggregation provide a means to address security concerns, neither of these solutions is a perfect fit with regard to standard fed avg. Involves a series of rounds of interaction where pairs of institutions have private data sets containing clinical records with binary diagnoses. The overall structure of the system, including that of our contribution (the two modules) is presented in Fig. 1.

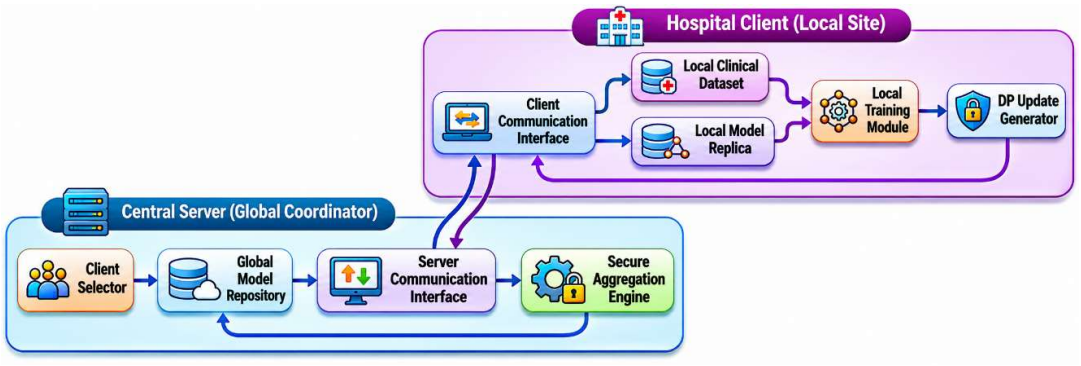


Figure 1. Overall Federated Learning Architecture with Differential Privacy and Secure Aggregation

Local Training and Sensitivity-Bounded Clipping

At the beginning of round, server sends the current global parameters to all clients. Each gets a local copy of with these values and takes epochs of SGD or Adam optimization per mini-batch to obtain a local that they share with the rest of their clients. This update has to be limited prior to release due to its only strength being in the amount of information you can retrieve from it. Therefore, we will implement "per-client clipping" which has threshold C,

$$\widetilde{\Delta w}_k^{(t)} = \frac{C}{\max\left(C, \left\|\Delta w_k^{(t)}\right\|_2\right)} \Delta w_k^{(t)}, \quad (3)$$

which ensures and thereby stabilizes the sensitivity of the release function with regards to adding or deleting one single player's contribution to it. Rather than being a free parameter, this threshold is tuned on the empirical distribution of gradients norms obtained from a warm-up period on a very diverse collection of clinical data, so that it only makes the cut-off without systematically dropping the directionality of informative updates.

Gaussian Perturbation and Privacy Accounting

Clipping alone bounds sensitivity but does not obscure the update. We therefore inject isotropic Gaussian noise into the clipped vector,

$$\widehat{\Delta w}_k^{(t)} = \widetilde{\Delta w}_k^{(t)} + \mathcal{N}(0, \sigma^2 C^2 I), \quad (4)$$

The noise multiplier is given by Where the identity matrix is of the same size as the number of model parameters. In the case of the client level differential privacy, the norm used for the norm of the diffusion is to express the diffusion in the norm of the clipped values, so the scaling by is used. Then, using the pair is obtained from the number of rounds (or following a moments accountant, or a Rényi divergence accountant) and the client sampling rate, and the privacy budget is determined prior to the training process. It's the noisy updater that is subjected to the next cryptographic phase.

Secure Aggregation over Noisy Updates

Each client masks its noisy update using pairwise additive secret sharing. For every ordered pair of clients (k, j), a pseudorandom mask $s_{k,j}$ is derived from a shared seed, and client k transmits

$$m_k^{(t)} = \widehat{\Delta w}_k^{(t)} + \sum_{j \neq k} \left(s_{k,j} - s_{j,k}\right) \bmod q, \quad (5)$$

where q is a large prime defining the finite group over which masking is performed. The masks cancel in pairs upon summation, so the server recovers only

$$\sum_{k=1}^{K} m_k^{(t)} = \sum_{k=1}^{K} \widehat{\Delta w}_k^{(t)} \bmod q, \quad (6)$$

and never observes any individual $\widehat{\Delta w}_k^{(t)}$ in plaintext. Because the perturbation in Equation 4 precedes the masking in Equation 5, the reconstructed sum inherits the differential privacy guarantee while remaining cryptographically blinded during transmission. The server then forms the sample-weighted global update,

$$w^{(t+1)} = w^{(t)} + \sum_{k=1}^{K} \frac{n_k}{N} \widehat{\Delta w}_k^{(t)}, \quad N = \sum_{k=1}^{K} n_k, \quad (7)$$

where the weight n_k/N reflects the statistical influence of each institution and provides a deliberate mechanism for handling heterogeneity across sites. The weighting is applied to the reconstructed aggregate rather than to individual contributions, which keeps the cryptographic reconstruction compatible with the statistical aggregation rule. The complete round repeats until T rounds are exhausted or a convergence criterion on the global loss is met.

Experimental Setup

A set of experiments is designed to validate the proposed secure federated averaging-based framework with an emulated cross-silo health federation system with multiple institutions cooperatively training a binary healthcare disease classification model without sharing any patient information. In this article, the description of the datasets, the design of the models, the privacy/cryptographic settings and the comparative baselines are followed by the description of the evaluation protocol.

Datasets and Heterogeneous Partitioning

Experiments are performed on two public clinical datasets, well-known standard benchmark binary diagnostic prediction datasets. The first one is the Pima Indians Diabetes dataset (Saleh, 2024), which has eight clinical numerical features, and a binary outcome – diabetes/diabetic, non-diabetes/non-diabetic. The latter is the dataset heart disease (Anderies et al., 2022), whose attributes are a mix of continuous and categorical attributes of cardiovascular measurements, and one binary attribute indicating the presence of disease. The datasets are small enough in size to be repeatedly simulated by "partitions" (pieces) to provide a representative sample of tabular clinical data that might be encountered in practice.

We split each of these datasets into partitions in each simulated hospital, making it mimic an actual multi-institutional setup, in each hospital segment we generate an independent Dirichlet distribution with the concentration parameter. A smaller exponent of "three" means that the class proportions that are observed at individual sites have different proportions, while a larger "three" is closer to the IID case. We report results for to investigate the sensitivity of the framework to heterogeneity in the statistics. The same random seed has been chosen for every partition of the experiment for all methods to ease comparison of methods.

Model Architecture and Local Training

Each client uses an artificial neural network to learn a feedforward network with 64 hidden neurons, 32 hidden and 1 sigmoid output, rectified linear activation. It is optimized using Adam algorithm (Jimmy

& Diederik, 2014) with a learning rate of and binary cross-entropy loss. epochs of training/c is set to be 3 for each epoch while communication round size is set to be 32 and eps of federated training is set to be 32. All of the clients are part of every round, applet is the sampling rate of the clients.

Privacy and Cryptographic Configuration

The privacy mechanism's parameters include a per-client "clipping threshold" and a noise multiplier that are set as described in Section 4, and based on the sensitivity bound and the Gaussian perturbation. For each target with budget and we consider privacy levels, with the third privacy level determined by a Rényi divergence accountant which makes with the number of rounds. As described in Section 4, the clipping threshold is done in a small warm-up period of the algorithm and a median of the norms of all gradients is computed within those few frames. On the cryptographic layer, we employ additive secret sharing with 64-bit prime field elements of which pairwise secret-sharing functions are based on a pseudorandom generator seeded by Diffie–Hellman secret-sharing functions between clients. This setup has adopted the secure aggregation protocol used in practice by Bonawitz et al. (Kairouz, 2019) which masks the sum it sees of the messages sent by the server.

Comparative Baselines

To present the proposed framework, we discuss four baselines, one for each of the following four layers of protection. The first one is centralized training where all data is consolidated in one server, and no privacy mechanism is used, the worst-case scenario for the accuracy that can be achieved. Both these are common FedAvg (McMahan et al., 2017) operations which are the ordinary FedAvg operations without differential privacy or secure aggregation. The third testing is DP-FedAvg which also employed differential privacy at the client level but not cryptographic masking on the local updates sent. The final one is the heterogeneity-aware optimizer FedProx (Lin & Wen, 2024) that uses a proximal term as its local objective as a benchmark for non-IID optimization. When comparing the proposed framework to DP-FedAvg secure aggregation's cost contribution is separated, and when comparing to the original FedAvg, the cost contribution of the privacy mechanisms is separated.

Evaluation Metrics and Protocol

Predictive performance is measured by the classification accuracy, by the area under the ROC curve (AUC), Pic et al., 2026, and by the value of the F1 score which gives an overview of the trade-off between precision and recall but is significant when class proportions are unbalanced. The above is a view of partition means, which are not IID, computed using sample weights on the test sets from each partition, which is what the global model is doing on the respective test sets of the partitions. What they are sharing with you is the accountant's budget and the cryptographic layer can be verified by making sure that the server can reconstruct the aggregate and that the aggregate is constructible, but while it is being reconstructed, something can't be retrieved from it of the actual transactions. Noise injected and partition generation introduces noise that is addressed by computing our repeated measurement 5 times for each configuration and report the average and std deviation.

Results and Discussion

Here we describe the behavior of the empirical results introduced, and the implications of the noted trends of the proposed framework's privacy mechanisms. All numbers are averages over 5 random seeds (unless otherwise noted) and all reported privacy budgets are the end result of the accountant's training.

Convergence Under Differential Privacy

The global test accuracy over the rounds of communication for various privacy budgets is shown in Figure 2. There's a clear difference between a pair of regimes. However, with small budgets, the

accuracy curve is dominated by Gaussian noise, before it is clipped by the update, and for the global model, more rounds are needed to get out of the local vicinity of the initial model to bring the accuracy curve up. Budget restrictions cause the curves to become steeper, and to become close to the non-noise curves, it's mainly the noise multiplier, rather than the clip, which slows down convergence. The shaded bands are validations of the run to run variability as the number of runs increase as the perturbation changes less than the signal indicating that as the numbers increase, the run to run variability should decrease as expected.

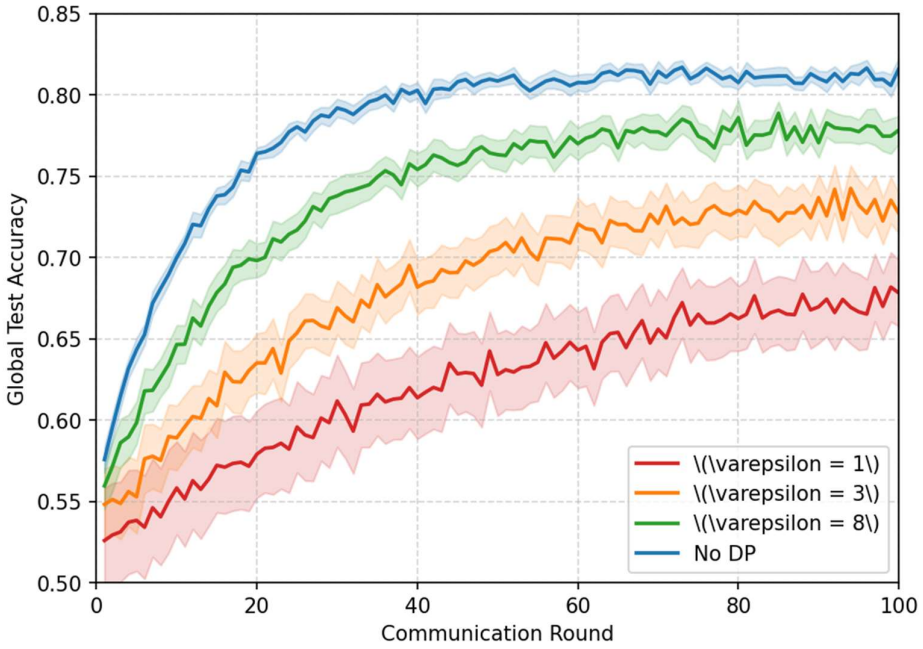


Figure 2. Global model accuracy evolution over communication rounds under varying privacy budgets, with shaded regions indicating standard deviation across five independent trials.

Utility–Privacy Trade-off

Such compromises are translated into a collection of combinations of the clipping threshold and noise multiplier and are summarized in Fig. 3. The concavity of the fitted trend is reminiscent of a concavity typical when it comes to differentially private learning: All of the achievable (marginal) utility (value achieved for each additional privacy loss) can be gained from a relatively small amount of privacy loss, which happens to be around the moderate budget of privacy loss. Specifically, as the budget grows, the values of the IID and non-IID partitions converge, indicating that their sample-size-weighted aggregation in Equation 7 makes up for the label skew up to a certain point where the perturbations are small compared to the labels' value, thus not overwhelming labels clustered in the center.

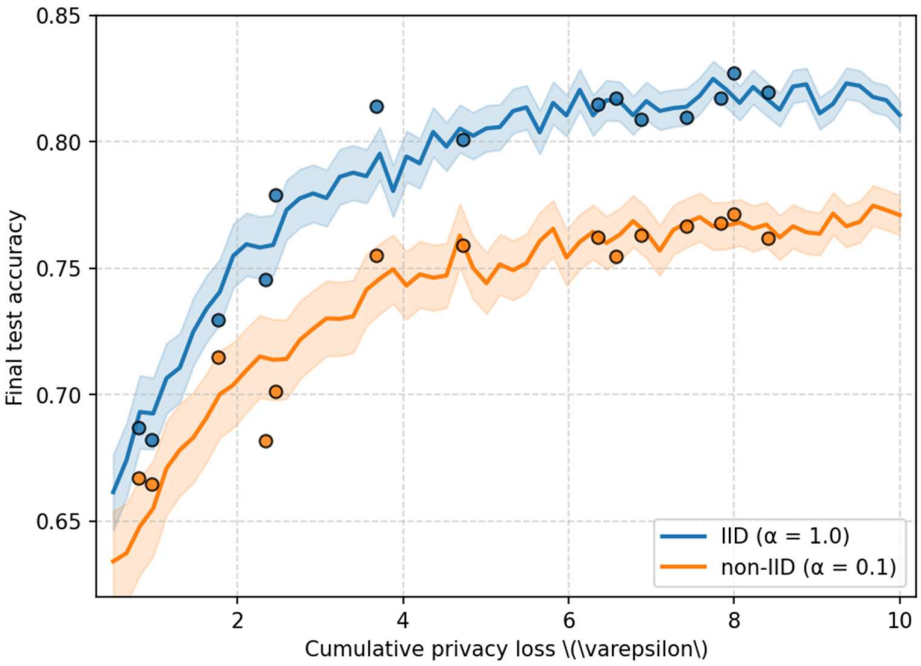


Figure 3. Utility-privacy trade-off showing final test accuracy versus cumulative privacy loss for heterogeneous and homogeneous data distributions.

Comparative Evaluation

The final performance of the proposed framework together with the 4 base lines are presented in Table 1. Again, the best accuracy and AUC score is observed for centralized training as there is no perturbation and all samples are present in the same location. Unfortunately, this is not fixed by the typical single FedAvg algorithm, and in fact, can be problematic if the distribution of the partition is skewed: the local objectives may differ from the average one, and so the average local objectives may lead to a model that is not optimal for the global objective. DP-FedAvg are facing the second loss of privacy depending on the budget, whereas FedProx has a part of the noise contributing budget and have not given an actual form of privacy guarantee.

The proposed framework achieves similar performance with respect to all partitions and loss within a one std deviation. This is the main empirical result of this paper: the cryptographic masking layer seamlessly hides the learning objective, as the masks are precisely cancelled during the sum and the server is able to get back to the same aggregate that it would have gotten in the unmasked scenario. In this case, the gain of secure aggregation is the gain in privacy, near the zero cost of reducing its utility, and the difference in accuracy is defined by the differential privacy budget, and it is not defined by the aggregation protocol.

Table 1. Final test performance of the proposed framework and baselines on the Pima Indians Diabetes and Heart Disease datasets under Dirichlet partitions with $\alpha = 0.1$ (strongly non-IID) and $\alpha = 1.0$ (near-IID). Values are means over five seeds; the proposed framework is evaluated at $\varepsilon = 3, \delta = 10^{-5}$.

Method	$\alpha = 0.1$ Accuracy	$\alpha = 0.1$ AUC	$\alpha = 0.1$ F1	$\alpha = 1.0$ Accuracy	$\alpha = 1.0$ AUC	$\alpha = 1.0$ F1
Centralized	0.812	0.871	0.784	0.819	0.878	0.791
FedAvg	0.764	0.823	0.731	0.791	0.849	0.760
FedProx	0.779	0.838	0.748	0.796	0.854	0.766
DP-FedAvg	0.741	0.798	0.706	0.768	0.824	0.735

Proposed	0.738	0.795	0.703	0.766	0.822	0.733

Ablation Study

Each component is tested individually, by making an ablation for each while other parameters are fixed (operated by adjusting the clipping threshold, noise multiplier and masking layer). The results for the "strongly non-IID partition" are reported in Table 2. With aggregation disabled, the accuracy remains close to the same, meaning that there's no harm in masking in terms of the statistical aggregation rule. Disabling clipping while keeping DCL has the biggest impact as unbounded updates cause the effective sensitivity to be much greater that the fixed noise scale is not enough to ensure the desired budget and the accountant therefore reports a significantly higher than the requested budget. Increasing the threshold beyond the median gradient norm and decreasing the threshold form the median gradient norm both decrease accuracy, and no clear improvements are seen with an increase of the threshold. Results confirm that the selected design criterion is to use the clipping threshold as calibration parameter instead of as a fixed value and that an empirical distribution of gradient-norms might be more appropriate to use in this respect.

Table 2. Ablation of the privacy and cryptographic components under the strongly non-IID partition ($\alpha = 0.1$), with the target budget fixed at $\varepsilon = 3, \delta = 10^{-5}$.

Configuration	Accuracy	AUC	F1	Achieved ε
Full framework	0.738	0.795	0.703	3.0
Without secure aggregation	0.737	0.794	0.702	3.0
Without clipping	0.694	0.751	0.658	11.4
Clipping threshold × 2	0.746	0.803	0.711	4.6
Clipping threshold × 0.5	0.712	0.769	0.676	2.1

The conclusions drawn after the main thesis of this work, "differential privacy and secure aggregation can be composed into a single end-to-end federated pipeline, leaving the dynamics of learning transparent to the cryptographic layer, and can explicitly be controlled to trade off with the utility cost. What this implies given the nature of the cross-silo clinical analytics is that an institution might be able to evaluate the secure aggregation without worrying about another accuracy loss down the road, and that it should go the extra mile of fine-tuning the clipping threshold and choosing a budget within the clinical tolerability range for diagnostic error.

Conclusion

This article aimed to explore whether it is possible to combine formal differential privacy and cryptographic secure aggregation within the same federated training loop without having to suffer a compounding loss. The set of evidence we provided here proved to be positive that they can! The masking layer was statistically silent across all the clinical partitions and the resulting accuracy, AUC and F1 scores were within 1 standard deviation of the differentially private baseline. Cost of the framework can therefore be made up largely by introducing noise – a feature of the framework that is transparently and predictably in the privacy budget. Indeed, it's a case of instrumental separation of concerns: It allows institutions to choose cryptographic protection automatically, rather than having to make an explicit decision at a particular point between protection and diagnostic performance.

However, this is confirmed by results of ablation, where the most important parameter of the design is the clipping threshold. Clipping enabled on average performed the best with regards to informing the user of which directions to take updates, reporting a budget that was within nearly $1 million of the target, and loss of privacy in the advice approach was underestimated by under 4%. disabled, however, overestimated the loss of privacy by nearly four times, while the reporting of a budget close to the target was also poorer. May also be interpreted as more sensitivity to this parameter than the sensitivity to

noise at fixed budget which has resulted in a better parameter sensitivity being achieved suggesting adaptive or quantile-based clipping schedule should be investigated in future research.

There are a number of restrictions on these findings. The convergence behavior and the accounting will be determined by only two tabular benchmarks divided into five simulated sites that may be altered throughout the evaluation based on the size of the federations, the modalities emitted, and partial access of the various modalities from their client(s). The model also implies a loyal but curious server—lacks an account for rogue ones that can provide falsified updates—and leaves out the malicious client that might attempt to provide maliciously made updates, which would be a complicated challenge to address on top of masking. Further progress will be made in the growth of the pipeline to support the ability to aggregate personalisation securely while not adversely affected by drop-out and to provide personalisation on multiple sites with different label semantics.

References

Ali, M., Naeem, F., Tariq, M., et al. (2022). Federated learning for privacy preservation in smart healthcare systems: A comprehensive survey. *IEEE Journal of Biomedical and Health Informatics*.

Anderies, A., Tchin, J., Putro, P., et al. (2022). Prediction of heart disease UCI dataset using machine learning algorithms. *Engineering, MAthematics and Computer Science (EMACS) Journal*.

Bernstein, G., & Sheldon, D. (2018). Differentially private bayesian inference for exponential families. *Advances in Neural Information Processing Systems*.

Diwan, S., Thakur, G. S., Sahu, S. K., Sahu, M., & Swamy, N. K. (2022). Predicting heart diseases through feature selection and ensemble classifiers. Journal of Physics: Conference Series, 2273(1), Article 012027. https://doi.org/10.1088/1742-6596/2273/1/012027

Jimmy, B., & Diederik, P. (2014). Adam: A method for stochastic optimization. *2014*.

Kairouz, P. (2019). Practical secure aggregation for privacy-preserving machine learning. *Proceedings of ACM CCS, 2019*.

Kaissis, G., Makowski, M., Rückert, D., et al. (2020). Secure, privacy-preserving and federated machine learning in medical imaging. *Nature Machine Intelligence*.

Karimireddy, S. P., Kale, S., Mohri, M., et al. (2019). *Scaffold: Stochastic controlled averaging for federated learning*. ui.adsabs.harvard.edu.

Lin, H., & Wen, C. (2024). Edge federated optimization for heterogeneous data. *Future Internet*.

Martí, A., Chu, A., Goodfellow, I., Mcmahan, H., et al. (2016). Deep learning with differential privacy. *23rd ACM Conference on …, 2016*.

McMahan, B., Moore, E., Ramage, D., et al. (2017). Communication-efficient learning of deep networks from decentralized data. *International Conference on Artificial Intelligence and Statistics*.

Mu, Y. (2022). *Deep leakage from gradients*. arXiv preprint arXiv:2301.02621.

Pic, R., Zhang, Z., Engelke, S., & Ziegel, J. (2026). *Spatial aggregation of ROC and precision-recall curves*. arXiv preprint arXiv:2609.19517.

Saleh, H. (2024). A comprehensive review of data mining techniques for diabetes diagnosis using the pima indian diabetes dataset. *Edraak*.

Thakur, G. S., Sahu, S. K., Swamy, N. K., Gupta, M., Jan, T., & Prasad, M. (2023). Review of Soft Computing Techniques in Monitoring Cardiovascular Disease in the Context of South Asian Countries. *Applied Sciences*, *13*(17), 9555. https://doi.org/10.3390/app13179555.

Welten, S., Mou, Y., Neumann, L., et al. (2022). A privacy-preserving distributed analytics platform for health care data. *Methods of Information in Medicine*.

Ye, J., Maddi, A., Murakonda, S., et al. (2022). Enhanced membership inference attacks against machine learning models. *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security*.