JOURNAL OF MECHANICS OF CONTINUA AND
MATHEMATICAL SCIENCES
www.journalimcms.org

ISSN (Online) : 2454 -7190 Vol.-21, No.-06, June (2026) pp 34-61 ISSN (Print) 0973-8975

A HYBRID SECURE AGGREGATION AND DIFFERENTIAL PRIVACY FRAMEWORK FOR COMMUNICATION-EFFICIENT BIG DATA ANALYTICS

Subhajit Roy[1], Rupak Chakraborty[2], Tapan Chowdhury[3]

Department of CSE (AI & ML) Techno Bengal Institute of Technology, Brahmapur-700150, Kolkata, West Bengal, India.

Department of CSE-AI, Techno India University, West Bengal, India.

Department of CSE, Techno Main Salt Lake, West Bengal, India.

Email: [1]Jroy395@gmail.com, [2]rupak.c@technoindiaeducation.com
[3]tapan2005cse@gmail.com

Corresponding Author: **Rupak Chakraborty**

https://doi.org/10.26782/jmcms.2026.06.00003

(Received: March 21, 2026; Revised: May 27, 2026; Accepted: June 09, 2026)

Abstract

Federated Learning (FL) is a type of distributed learning where several clients (clusters) train a machine learning model without sharing the raw data directly. Nevertheless, there are still three significant issues for practical FL systems: privacy leakage through sharing the model update, heavy communication burden, and unstable learning performance when the data distribution is not IID. To enhance the protection of privacy and communication efficiency in distributed big data analytics, the authors introduce a Privacy-Preserving Federated Learning (PP-FL) framework combining Differential Privacy (DP), Secure Aggregation (SA), and Adaptive Gradient Compression (AGC). Differential Privacy implies that calibrated Gaussian noise is added to local updates, and Secure Aggregation ensures that the central entity cannot see individual updates from the clients. Adaptive Gradient Compression cuts down on communicating the most important components of a gradient. The proposed framework is tested with the high-resolution MNIST dataset and CIFAR-10 dataset in non-IID federated settings. The experimental results demonstrate that PP-FL has a significant boost in reducing communication cost compared to standard FedAvg and DP-FedAvg. The results also illustrate a clear trade-off between privacy and utility, as the addition of noise (differential privacy) or departure from an IID data distribution or extreme compression can negatively impact classification accuracy. This behaviour is explained by means of revised studies that consist of a component-wise interpretation, observation of the gradient norm, and evaluation of the compression ratio. The overall results suggest that the proposed PP-FL system can be employed to implement communication-efficient and privacy-preserving federated learning, and that careful tuning of noise parameters and compression parameters is essential to ensure the stability of learning.

Keywords: Federated Learning, Differential Privacy, Secure Aggregation, Privacy-Preserving Machine Learning, Gradient Compression, Distributed Learning, Big Data Analytics

Nomenclature

(N)	Number of participating c lients in the federated learning system
(D_i)	Local dataset of client (i)
(D)	Total dataset aggregated across all clients
($f_i(\omega)$)	Local loss function of client (i)
($F(\omega)$)	Global objective function in federated learning
(g)	Gradient vector computed during local training
(C(g))	Gradient compression operator
(m)	Binary mask vector used in gradient sparsification
(k)	Number of selected gradient components during compression
(d)	Dimension of the model parameter vector
(R)	Communication reduction ratio
(T)	Total number of communication rounds
(q)	Client sampling probability in each communication round

Greek Symbols

ω	Global model parameter vector
ω_t	Global model parameters at communication round (t)
ω_i^t	Local model parameters of client (i) at round (t)
η	Learning rate used in stochastic gradient descent
ρ	Compression ratio (k/d)
σ	Standard deviation of Gaussian noise used for differential privacy
ε	Privacy budget in differential privacy
δ	Probability of privacy failure
ϵ_c	Error introduced by gradient compression

I. Introduction

"Federated Learning (FL)" is a new paradigm of traditional learning, in which multiple dispersed colleagues participate in the joint training of models but do not exchange raw data, which are retained locally [VII], [XXI]. Each participating client or institution does local training on its own data rather than transferring all the data to a central server, and only sends the learned model parameters or gradients to a coordinating server. These updates are, in turn, amalgamated by the server via an algorithm (usually the Federated Averaging (FedAvg) algorithm) [XIV] to create a better global model, which is again shared among the clients in the next round of training. Such decentralization enables the collective model to gain knowledge from different and geographically dispersed data without destroying data privacy.

The primary benefit of Federated Learning is exhibited by the privacy-sensitive and regulation-friendly construction, according to which the sensitive information, like

medical history, financial operations, or activity of users, does not leave their initial storage [VII], [XXI], [XI]. Consequently, FL has been taken off-hand in a number of fields such as healthcare, finance, mobile edge computing, and IoT systems, wherein the secrecy and ownership of data are essential. Moreover, it will minimize the volume of data to be transferred, thus it will be scalable to big, heterogeneous, and resource-constrained networks.

Despite all these benefits, the idea of FL application in practice is not as straightforward due to the factors of heterogeneity of data (non-IID data distribution), communication overhead, and the threat of information leakage during the process of updating the model [XII], [XIII]. Even though FL does not give any evidence of a privacy breach of personal data, one can make reasonable predictions of personal data using gradients or model parameters [I], [VI]. Hence, developing a model that guarantees a favorable outcome of privacy, successful communication, and high precision of the model is important.

To overcome these hurdles, this paper has recommended a PP-FL framework that is founded on the consolidation of DP, SA, and AGC. Differential Privacy is utilized to protect the model updates that are transmitted by the clients against an attack, whereby they are calibrated with Gaussian noise, and Secure Aggregation does not mean the server is able to view an individual update transmitted by a client, but rather the aggregated update. Adaptive Gradient Compression (AGC): It lessens the number of overhead components relayed by transmitting only the most enlightening components of the gradient adaptively. Although the initial version did include the discussion of privacy, gradient transmission, and non-IID behaviour of the clients in its modular fashion, the new framework does state the relationship between the three in a very explicit manner. This paper makes the following contributions, outlined below:

I. A hybrid PP-FL scheme is built by joining DP, SA, and AGC to apply a federated learning (FL) that is communication- and privacy-conscious. [I], [V], [II], [III], [VI].

II. Differential privacy noise and adaptive gradient compression and non-IID partitioning of model utility components to identify which mechanisms cause interference are proposed to isolate the mechanism by the recommendation of a distillate experimentation by ablations[I], [II], [III], [VI], [XIII]

III. Integration is done of a gradient norm and optimization stability analysis to pinpoint the cause of the degradation of the utility by noisy and compressed federated updates. [I], [III], [XIII].

IV. The impact of compression ratio and noise scale on the convergence behaviour is investigated in order to be able to see further into the privacy–utility–communication trade-off. [I], [II], [III], [VI].

V. It is tested on the non-IID federated scenario on a low-frequency of communication dataset, lower privacy, and a learning stable scenario on MNIST and CIFAR-10 datasets. [X], [IX],

II. Overview Of Federated Learning

"Federated Learning (FL)" is a machine learning paradigm that is decentralized and enables more than two clients or data owners to jointly train a common model without sharing their raw data [VII], [XXI]. Rather than transmitting data to a central server, every client will train the model on its own data and only transmit model updates

(gradients or parameters) to a central server. The server then combines these updates to come up with a better global model that is redistributed to all the clients in the subsequent training round. This is repeated until the model converges. Since the information does not leave its source, FL offers high protection of privacy, and it applies to areas that are sensitive to data sharing, such as healthcare, finance, and mobile edge networks, because they have privacy laws, such as the GDPR and HIPAA, that limit data sharing. Mathematically, the global objective of FL can be expressed as:

$$\min_{\omega} F(\omega) = \sum_{i=1}^{N} \frac{|D_i|}{|D|} f(\omega) \qquad (1)$$

Where fI(ω) represents the local loss function of client i, and Di denotes its local dataset. Common FL algorithms include:

FedAvg (McMahan et al., 2017): Clients train locally for multiple epochs before averaging their updates on the server.

FedProx (Li et al., 2020): Extends FedAvg to handle non-IID data by introducing a proximal term.

SCAFFOLD (Karimireddy et al., 2020): Addresses client drift through control variates.

Although FL has its benefits, it still has a number of challenges, such as the heterogeneity of data, communication overhead, and the possibility of information leakage due to shared model updates. These challenges have motivated privacy-preserving approaches such as **Differential Privacy** and **Secure Aggregation**, which are integrated in this paper's proposed PP-FL framework [V], [I], [VI].

III. Related Study

"Federated Learning (FL)" allows decentralized data to be trained together with a model and retains raw data locally [VII], [XXI]. A number of comprehensive surveys, including the one by Kairouz et al., have described various features of "Federated Learning (FL)" such as the settings of the FL, its communication architecture, and the significant research challenges of Federated Learning, i.e., its privacy, robustness, and large-scale deployment [VII], [XI], [XII]. Among the oldest and most popular algorithms in this direction is the FedAvg [XIV]. which integrates client-side model updates by means of simple averaging to facilitate effective training with a variety of devices with limited resources [XIV]. Subsequently, methods were brought in to manage the problem of data heterogeneity and client drift. For example, FedProx [XIII] presents a proximal term that ensures the stability of local updates in case client data are not identically distributed, but SCAFFOLD [VIII] applies control variates to reduce the drift and bias between client and server updates. FL privacy mechanisms include cryptographic and statistical mechanisms [V], [I], [VI]. Secure aggregation protocols [V] enable the server to recover only an aggregate of client updates without learning any individual contribution, even in the presence of dropouts.

Orthogonally, differential privacy (DP) limits information leakage from model updates; DP-SGD [I] with the moment accountant offers tight privacy accounting for deep networks, and client-level DP has been adapted to the FL [XV], [VI] setting for

language modelling and mobile intelligence tasks. Combining secure aggregation (protecting updates in transit) with DP (bounding inference from outputs) is increasingly advocated to deliver end-to-end confidentiality [V], [I], [VI]. To reduce communication cost, different techniques compress the client-server payload [II], [III]. Examples of cutting the bits of gradients to reduce error include quantization and sparsification techniques, e.g., QSGD [II] or top-k sparsification [III]. Similarly, much of what has been developed to minimize the communication load of the training process and to alleviate the effects of slow/untrustworthy clients include system-sensitive techniques, such as partial client participation, client sampling, and adaptive local epochs, among others [XIV], [XVI], index and security are end on due to dangers like model poisoning and Byzantine clients [IV], [XXII]. The analysis of strong aggregators such as Krum [IV] and trimmed-mean decreases the effects of adversarial updates, and subsequent studies attempt to find the best statistical rates with Byzantine resilience [IV], [XXII]. These privacy mechanisms do not in any way imply robustness, and incentives should be given to joint defences between anomaly detection and secure/DP aggregation [V], [I], [IV], [XXII]. "Federated Learning (FL)" has demonstrated itself at the application level to be useful in privacy-conscious tasks, like mobile predictions and text prediction, healthcare analytics, and IoT sensing [XXI], [XVII], [XX]. Due to strict privacy laws like GDPR and HIPAA, [VII], [XXI], [XI], all information in these systems is not easily passed and stored under one centralized server.

In healthcare, one of these areas is of paramount interest: various hospitals or research centers can be able to train shared models together without showing the data concerning patients [XXI], [XVII], [XVIII]. In practice, however, healthcare data are usually not IID and are restricted in labelled samples, and subject to network constraints, and it is challenging to coordinate learning across sites [XIII], [VIII], [XII]. All these make it clear how useful FL systems should value efficiency in communication and strong privacy protection in order to deliver reliable performance in these controlled settings [VII], [V], [I], [II]. Fortunately, existing art has done that to privacy and efficiency separately, with cryptographic secure aggregation enhancing confidentiality at the cost of bandwidth and complexity of protocols, and DP, conversely, causing loss of accuracy without careful improvement of the noise budget with respect to data heterogeneity and participation rates [V], [I], [VI], [XIX]. The proposed PP-FL framework tackles this problem by introducing Secure Aggregation, Differential Privacy, and Adaptive Gradient Compression into the Federated Learning pipeline. This study, however, sees the proposed framework as a privacy– aware and communication– efficient solution where there's a clear privacy-utility- communication trade-off. The revised analysis thus examines how interdependence in data distribution, privacy noise, and gradient compressions can interact in affecting the model utility and stability of convergence.

The following section introduces the proposed Privacy-Preserving Federated Learning (PP-FL) model, which integrates these privacy mechanisms with adaptive gradient compression for communication-efficient Big Data analytics [V], [I], [II], [III]. The concept of privacy preservation in federated learning is to make sure that none of the sensitive information concerning individual users or datasets is disclosed in the process of collaborative training of a model [VII], [XI], [XXI]. Even though the raw data are localized to each client, the exchange of model updates or gradients with the central server may still expose personal information in gradient inversion attacks and by

inference of its membership [I], [VI]. Privacy-preserving frameworks that alleviate these risks, then, have mathematical and cryptographic guards that regulate or conceal the input of the various data points [V], [I], [VI]. Of them, Differential Privacy (DP) is a statistically guaranteed privacy protection of individual records with the use of random noise [I], [VI], and Secure Aggregation (SA) guarantees that the server only obtains aggregate model updates, so no information about a particular client is available [I]. Collectively, these processes create a privacy-by-design that supports confidentiality and regulatory adherence and facilitates distributed learning in a wide range of data that are sensitive [I], [I], [XII], [XIX].

Despite many federated learning concepts being established to address the issue of data privacy and communication effectiveness, most of the current methods either only provide differential privacy or cryptographic aggregation, which have immediate privacy [V], [I], [VI], [XII], accuracy, and scalability trade-offs. Furthermore, such methods usually fail to take into consideration the high cost of communication that is incurred in the large-scale Big Data setting [II], [III], [XII]. In order to address these drawbacks, this paper proposes a hybrid model, "Privacy-Preserving Federated Learning (PP-FL) that incorporates the Secure Aggregation (SA) model and Differential Privacy (DP) in a communication-efficient framework based on Adaptive Gradient Compression (AGC). The proposed model will seek to reach a compromise between high privacy guarantees, accuracy of the model, and low communication overhead in the case of distributed learning".

Research Gap and Motivation

From the above papers, the existing federated learning research is mainly on the three separate issues of privacy protection, communication efficiency, and system robustness. Secure Aggregation means that clients' individual vectors cannot be directly observed by the server, while Differential Privacy means that protect clients against statistical leakage on inference attacks [V], [I], and [VI]. Similarly, the use of gradient compression techniques maintains low communication overhead by sending fewer gradients in distributed training [II], [III]. But when these mechanisms interact in a real-life non-IID federated setting, the complex interplay between them makes understanding their interaction more challenging.

Moreover, evidence shows that privacy noise can also negatively affect local update quality, hard compression can drop valuable gradient details, and data that is not i.i.d distributed can cause higher drift in local updates [XIII], [VIII], [XII]. Many studies are found without taking into account the effects of these factors on the composite, utility of the model, convergence stability, and communication cost. Moreover, the sparsity can induce statistical features on the message update behavior that may change how the statistical structure of updates is computed and used for calibration and optimisation of privacy. Therefore, it is essential to have a joint knowledge of Differential Privacy, Secure Aggregation, and Adaptive Gradient Compression, to achieve deeper understanding of the privacy–utility–communication trade-off.

This gap is taken into account in developing the proposed PP-FL framework, where the DP and SA, along with AGC, are included in a single FL pipeline and treated as the same in the non-IID setting. The updated study also addresses issues of component-

wise interpretation, the observation of the gradient norms, and the evaluation of compression ratios for their respective ablation to see whether results come from privacy noise, gradient sparsification, non-IID partitioning, or all three.

IV. Proposed Methodology

Figure 1 demonstrates the workflow of the proposed Privacy-Preserving Federated Learning (PP-FL) system in general. It displays the step-by-step process, which initially involves model initialization followed by the global model update, showing how privacy preservation and efficiency are preserved together during the training cycle. "The proposed Privacy- Preserving Federated Learning (PP-FL) system is an integrated algorithm of Secure Aggregation (SA) and Differential Privacy (DP) to ensure the privacy and security of the system and correctly inform federated learning setting".

Fig. 1. "Workflow of Privacy-Preserving Federated Learning (PP-FL)"

A. Unified Interaction of DP, SA, and AGC in the Proposed PP-FL Framework:

Diff. Priv., Secure Aggregation, and Adaptive Gradient Compression are not independent modules of the proposed PP-FL framework. Instead, in this federated learning pipeline, all three mechanisms are sequentially operated, and dependent on each other. To achieve this, in each communication round, a client first calculates the local model update from its private dataset. The update is then clipped to control sensitivity, add Gaussian noise to satisfy Differential Privacy, add compression using Adaptive Gradient Compression, and then add Secure Aggregation before sending to the server.

Important because permissiveness depends on the interactions between each one, and each component controls the behavior of the following component. Some random noise is added to the local update to provide better privacy protection but a weaker useful learning signal,

by Differential Privacy. Adaptive Gradient Compression then identifies the most important parts of the gradient and transmits them. A ratio that is too low may result in the loss of helpful gradient components, particularly in the initial training cycles before model convergence. Secure Aggregation allows the compressed and noisy update to be collected only by the server, which cannot know about any individual updates from any of the clients.

This fusion brings about a trade-off among privacy, communication cost, and model utility. A larger noise scale promises more privacy, but may also result in less accuracy. A higher compression results in less communication overhead, but may lead to more errors for optimization. The effects are further compounded by non-IID client data, as this further induces biases in the client updates and contributes to higher client drift. Thus, careful calibration of the transformation of privacy noise and compression ratio will be very important for the proposed PP-FL framework, especially for the non-uniform federated data distribution.

We explicitly take this interaction into account and perform ablations, gradient norm observation, and compression-ratio evaluation for the effects of key factors: privacy noise, adaptive compression, and non-IID partitioning. In this way, the framework can be viewed as a collection of privacy-preserving strategies, as well as an end-to-end optimization pipeline for privacy-aware and communication-efficient federated learning.

B. The Big Data Architecture for Federated Learning

The suggested PP-FL model is built in a distributed Big Data architecture, which can serve privacy-sensitive and large-scale settings. Here, the data is stored locally on a number of clients, e.g., hospitals, IoT devices, organizations, and each of them does a local computation on the dataset. The architecture has three layers: a data layer, distributed data storage, a local computation layer with preprocessing and training models, and a federated coordination layer with secure aggregation and updated global models. The communication layer provides the encryption of data and synchronization of data between clients and the central server. This architecture may be combined with Big Data engines like Hadoop, Spark, or Kafka in order to process high-volume heterogeneous data efficiently. By integrating Big Data infrastructure and federated learning, the system would guarantee the presence of scalability, privacy, and efficient analytics in decentralized settings, where the clients maintain information confidentiality and follow stringent privacy policies during the whole learning process.

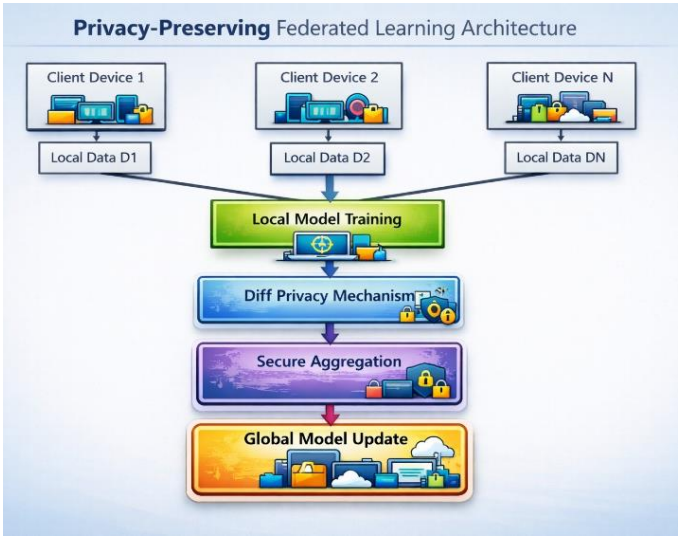


Fig. 2. Privacy-Preserving Federated Learning Architecture

The proposed PP-FL framework enables multiple distributed client devices to train local models using private datasets. Differential privacy is applied to protect sensitive information before transmitting model updates, while secure aggregation ensures that individual client updates remain confidential during global model aggregation.

The architecture of the PP-FL framework comprises three key components:

1. **Client Devices (Data Owners):** Every client Ci has a private dataset Di, and he/she also trains his own model Di. Users calculate model updates (gradients) Δ i and use privacy-preserving transformations to be transmitted.
2. **Secure Aggregator (Server): "**The server collects encrypted updates from all participating clients and performs secure aggregation to compute a global model w_t without accessing any individual client's update".
3. **Differential Privacy Mechanism:** To guarantee statistical privacy, each client perturbs its local gradient with Gaussian noise before encryption and transmission.

During each training round, clients download the global model ωt, perform local training, and upload their masked updates to the server. The server then aggregates the encrypted updates and broadcasts the new global model ω_{t+1} to all clients

C. Mathematical Formulation

"Let N denote the number of clients, and D_i be the local dataset of client i. The global model aims to minimize the federated objective.

$$\min_{\omega} F(\omega) = \sum_{i=1}^{N} \frac{|D_i|}{|D|} f(\omega) \quad (1)$$

Where $f_i(\omega) = \frac{i}{D_i} \sum_{x \varepsilon D_i} l(\omega; x)$ is the local loss for client i , and $D = \Sigma_{i=1}^{N} |D_i|$

Each client performs local training for E epochs using stochastic gradient descent (SGD):

$$\omega_i^{t+1} = \omega_t - \eta \nabla f_i(\omega_t) \quad (2)$$

Before sending updates to the server, each client applies **Differential Privacy** by adding Gaussian noise:

$$\tilde{\omega_i}^{t+1} = \omega_i^{t+1} + \mathcal{N}(0, \sigma^2 I) \quad (3)$$

where σ controls the noise scale and thereby the privacy budget ϵ.

The Secure Aggregation protocol ensures that the server can only observe the aggregate:

$$\omega^{t+1} = \frac{1}{N}\sum_{i=1}^{N} \omega_i^{\sim t+1} \quad (4)$$

without learning any individual $\omega_i^{\sim t+1}$.The global model is then updated as:

$$\omega_{t+1} = UpdateRule(\omega^{\sim t+1})$$

While Secure Aggregation and Differential Privacy provide strong protection against data leakage, they introduce additional computational and communication overhead during the training process. The repetitive transmission of encrypted and noisy gradients can be extremely sluggish in large federated systems that involve a large number of clients, and that may impact a large bandwidth cost". To address this shortcoming and make the proposed PP-FL architecture more efficient with regard to communication, Adaptive Gradient Compression (AGC) is also added. AGC is selective in the transmission of the largest gradient components, depending on their size, and dynamically changes the compression ratio depending on the state of networks and convergence properties. This will ensure that the proposed system is very precise and offers privacy guarantees, is economical in communication terms, and is therefore able to be scaled and functional in Big Data conditions.

The basic training process of the proposed PP-FL framework is given in the above mathematical formulation. The local update [in the revised framework] is first clipped via gradient clipping, noise added via Differential Privacy, and then shrunk via Adaptive Gradient Compression, before being securely aggregated. The order is important because every step will impact the subsequent step. The increased privacy protection offered by Differential Privacy noise comes at a cost, however—namely, the baseline signal used for the useful learning becomes noisier. Adaptive Gradient Compression has the benefit of minimizing the cost of communication, but if the compression ratio is too low, it might discard some critical gradients. This effect may be exacerbated under a non-IID client data distribution, as local client updates themselves may be skewed towards limited classes. Hence, the suggested framework should be well-tuned based on the noise scale and compression ratio to optimize the trade-off between privacy, communication efficiency, and model utility.

D. Adaptive Gradient Compression Algorithm for Communication-Efficient Federated Learning

Communication overhead between the central server and clients is one of the major bottlenecks in federated learning on a large scale. During each training step, customers typically provide full gradient vectors of a dimensionally very large size to deep learning. As the number of clients taking part in it grows, the cost of communication also rises accordingly and hence can significantly slow down the training procedure in a bandwidth-limited environment. To counter this reality, the Adaptive Gradient Compression (AGC) mechanism of the suggested Privacy-Preserving Federated Learning (PP-FL) framework is introduced. The point is that the most important gradient components should be sent instead of the whole gradient vector. Contrary to other traditional gradient compression methods, in which the compression ratio is always fixed, the presented AGC dynamically changes the gradient elements transmitted based on the dynamics in training and communication limitations. The strategy is adaptive and lowers communication cost but maintains the most informative

updates needed in model convergence.

Mathematical Formulation

Let, $g \in \mathbb{R}^d$ denote the local gradient vector computed by a client during a training round, where $d$represents the dimensionality of the model parameters

"The compression operator $C(\cdot)$used in the proposed framework is defined as

$$C(g) = g \odot m \quad (5)$$

Where",

- $m \in \{0,1\}^d$is a binary mask vector
- $\odot$ denotes element-wise multiplication.

The mask only retains the top-k gradient components that have the highest magnitudes, with the others being set to zero. This gives the system the opportunity to only communicate the most informative gradient updates.

Compression Ratio

The compression ratio of the proposed AGC mechanism can be expressed as $\rho = \frac{k}{d}$

Where,

- $k$is the number of selected gradient components
- $d$is the total gradient dimension.

"Since $k \ll d$, the proposed method significantly reduces the amount of data transmitted during each communication round".

Adaptive Gradient Compression Algorithm (AGC):

Input: Gradient vector g, compression parameter k, bandwidth threshold B

Output: Compressed gradient $C(g)$

1. "Compute the absolute magnitude of each component in the gradient vector g.
2. Sort the gradient components in descending order based on their magnitude.
3. Select the **top-k gradient components** with the highest magnitude.
4. Construct a binary mask vector $m$such that"

$$m_i = \begin{cases} 1, & \text{if } g_i \text{ belongs to top-k elements} \\ 0, & \text{otherwise} \end{cases}$$

5. Adapt the compression parameter $k$dynamically based on network bandwidth:

if available bandwidth < B, reduce k

otherwise increase k

6. Apply the compression operator

$$C(g) = g \odot m$$

7. Transmit the compressed gradient $C(g)$to the central server.

The proposed AGC algorithm ensures that only significant gradient components are transmitted, and this reduces the communication overhead but does not compromise the important training information. Moreover, the adaptive compression ratio allows the PP-FL framework to achieve efficient communication in diverse network conditions without causing any substantial impact on the model convergence. The given adaptive compression approach will make the proposed PP-FL framework stand out from the conventional federated learning approaches that use standard gradient compression techniques.

The Learning Stability of Effect of Compression Ratio: The compression ratio is an important term in the proposed PP-FL framework to balance the communication cost and the model utility. A smaller compression ratio for compression of the number of transmitted gradient components and thus lowers communication overhead. But a low

compression ratio will lose out on valuable gradient information in the early rounds of communication, in particular when the model is learning generic decision boundaries. This may cause somewhat marginal optimization and diminish location accuracy.

Consequently, the adaptive compression mechanism can not solely be regarded as a communication-saving technique. It also has a direct impact on the quality of the aggregated update. This is more pronounced in the non-IID distribution of client data, where each client could generate a possibly biased local gradient to the rest, due to the distribution of classes restricted to it. If these biased gradients are also compressed, the global model can lack learning signals or learning contribution signals that do not represent the data in a balanced way. Therefore, different compression ratios are assessed in the revised experimental analysis to check if the aggressive compression can degrade the utility.

In this study, the compression ratio is studied within the range 0.05-0.40. Lower values (e.g., 0.05) give more communication reduction, higher values (e.g., 0.20 or 0.40) allow more gradient information and might lead to improved learning stability. Based on this analysis, a more feasible trade-off between the communication efficiency and the model performance can be found in the proposed PP-FL framework.

Comparison with Existing Compression Methods: A number of techniques, including gradient compression, which are QSGD and Top-k sparsification, have been postulated to minimise communication overhead in distributed learning systems. But the majority of such techniques use a constant compression ratio during training. Contrastingly, the proposed Adaptive Gradient Compression (AGC) algorithm is a dynamic adjustment of the compression ratio that depends on the network bandwidth and training convergence behaviour. This adaptive process enables the proposed PP-FL framework to sustain efficiency in communication and accuracy in models' preservation that separates it from traditional unchanging-compression methods applied in federated learning.

E. Analytical Privacy Bound Derivation

"In the proposed Privacy-Preserving Federated Learning (PP-FL) framework, privacy protection is achieved by applying Differential Privacy (DP)" to local model updates before secure aggregation. To formally quantify the privacy guarantee, we derive an analytical privacy bound based on the Gaussian mechanism and its composition over multiple federated communication rounds. "Let g_i^t denote the clipped local gradient update computed by client $i$at communication round t. After gradient clipping with norm bound C, Gaussian noise is added as

$$\tilde{g_i^t} = g_i^t + \mathcal{N}(0, \sigma^2 C^2 I) \quad (6)$$

where σdenotes the noise multiplier and $I$is the identity matrix". Since clipping bounds the ℓ_2-sensitivity of the gradient update by C, the Gaussian mechanism ensures that each local release satisfies (ε_0, δ)-differential privacy with

$$\varepsilon_0 \leq \frac{\sqrt{2In\left(\frac{1.25}{\delta}\right)}}{\sigma}, \quad (7)$$

for a single participation event.

In federated learning, privacy loss accumulates over multiple "communication rounds. Let Tdenote the total number of rounds and $q$the client's sampling probability in each round. Using the moments accountant principle, the cumulative privacy budget after Trounds can be upper-bounded as

$$\varepsilon_T \approx \frac{q\sqrt{2TIn\left(\frac{1}{\delta}\right)}}{\sigma} \quad (8)$$

This expression shows that the total privacy loss increases sub-linearly with the number of rounds and decreases inversely with the noise scale. Therefore, stronger privacy can be achieved by increasing σ, although excessive noise may degrade model utility. To capture the privacy–utility relationship more explicitly, let the model utility degradation due to noise be denoted by $\mathcal{U}(\sigma)$, where utility decreases as σincreases. Thus, the proposed PP-FL framework admits the following privacy–utility trade-off:

$$\varepsilon_T \propto \frac{q\sqrt{T}}{\sigma}, \quad (9)$$

Hence, the proposed framework provides a controllable analytical bound between privacy protection and learning utility. The proposed PP-FL framework compares to the traditional federated learning schemes by providing only an empirical value of privacy, whereas the explicit cumulative privacy bound in the proposed model is based on client sampling, gradient clipping, and Gaussian perturbation. This derivation shows that the proposed PP-FL framework manages to keep a privacy range controllable while simultaneously modeling utility that depends on the noise scale, compression ratio, and non-IID data distribution. In particular, for fixed $q$and δ, the cumulative privacy loss grows as $O(\sqrt{T}/\sigma)$, which demonstrates that the proposed PP-FL framework scales gracefully with the number of communication rounds under appropriate noise calibration. This analytical result distinguishes the proposed framework from prior FL systems that rely only on empirical privacy reporting without an explicit bound on cumulative privacy leakage.

F. Privacy Accountant Analysis:

To estimate the cumulative privacy leakage through the proposed PP-FL is based on a privacy accounting mechanism in multiple training rounds. The moment accounting approach was developed for Differentially Private Stochastic Gradient Descent (SGD), which we extend here. We extend the moments accountant approach developed for DPP-SGD. Gradient Descent (DP-SGD). The privacy accountant monitors the sum of the privacy loss with federated. It is more efficient than naive composition (and also offers a tighter bound on the privacy budget ε) due to learning rounds. M methods: Assume that the training process lasts for T communication rounds; each client performs M methods. The sampling probability is q (q is a constant rate from 0 to 1). Using the moments accountant approach, the total privacy loss after T rounds can be bounded as:

$$\varepsilon \approx q\sqrt{(2T\,log(1/\delta))} / \sigma \quad (10)$$

Where,

q = sampling rate of clients

T = number of communication rounds

σ = Gaussian noise scale

δ = failure probability

This formulation allows us to estimate the cumulative privacy loss during federated training. As the number of communication rounds increases, the privacy budget ε also increases. However, increasing the noise scale σ reduces the privacy loss, thereby improving privacy protection. In the proposed PP-FL framework, the parameters are

chosen such that ε remains within the practical privacy range (ε ≈ 1–3), which is considered strong privacy protection in real-world deployments. Unlike many existing federated learning approaches that report empirical privacy values, the proposed framework analytically tracks cumulative privacy loss using a privacy accountant, thereby providing stronger theoretical guarantees for privacy preservation.

G. Proposed Algorithm

Algorithm 1: Privacy-Preserving Federated Learning (PP-FL)

Input: Global model ω_0 , learning rate η, number of clients N, privacy parameter σ

"Output: Trained global model ω_T

1. **Initialize** global model ω_0
2. **For each round** t = 1,2, … , T:
   a. Server broadcasts ωt to all clients
   b. Each client C_i
      - Trains locally on Di foriE epochs to obtain w^{t+1}
      - Adds Gaussian noise":
        $\omega_i^{\sim t+1} = \omega_i^{t+1} + \mathcal{N}(0, \sigma^2 I)$ (11)
      - Compresses update: $\omega_i^{\sim t+1} = C(\omega_i^{\sim t+1})$
      - Encrypts $\omega_i^{\sim t+1}$ using secure aggregation keys
      - Sends encrypted update to server
   c. Server aggregates securely:
      $\omega^{t+1} = \frac{1}{N}\sum_{i=1}^{N} \omega_i^{\sim t+1}$ (12)
   d. Server updates global model:
      $\omega_{t+1} = UpdateRule(\omega^{\sim t+1})$ (13)
3. **Return** final model ω_T.

This design guarantees privacy-by-design, which provides high confidentiality, scalability, and regulatory data protection standards.

H. Theoretical Analysis and Novelty of the Proposed PP-FL Framework: Unlike conventional federated learning frameworks that are based on Differential Privacy or Secure Aggregation. The proposed "Privacy Preserving Federated Learning (PP-FL)" framework combines these two aspects in an independent way. Differential Privacy (DP), Secure Aggregation (SA), and Adaptive Gradient Compression (AGC) together in a common structure. The theoretical novelty of the proposed framework is to show that the combination of the three mechanisms results in bounded privacy guarantees and reduces the privacy biocommunication complexity without influencing the convergence stability. The following theoretical analysis provides justification for the communication efficiency and privacy–utility trade-off of the proposed PP-FL system.

1. Communication Complexity Analysis: "Let N denote the number of participating clients and d denote the model dimension. In traditional federated learning algorithms such as FedAvg", the communication cost per round is proportional to O(Nd), since each client transmits the full gradient vector. In the proposed PP-FL framework,

Adaptive Gradient Compression transmits only the top-k significant gradient components where k << d. The communication reduction ratio can be expressed as:

$$R = O(Nk) / O(Nd) = k / d \quad (14)$$

Since k << d, the proposed PP-FL framework significantly reduces communication cost while preserving the most informative gradient updates. Since k is significantly smaller than d, the proposed PP-FL framework achieves substantial communication reduction while preserving the most informative gradient updates.

2. Privacy–Utility Trade-off Analysis: "Differential Privacy ensures that the contribution of any individual data record remains statistically indistinguishable. The Gaussian mechanism provides (ε, δ)-differential privacy when Gaussian noise with variance σ^2 is added to the gradients. The privacy budget ε can be approximated as:

$$\varepsilon \approx (\Delta f / \sigma) \sqrt{(2\, log(1.25/\delta))} \quad (15)$$

where Δ f represents the sensitivity of the learning function". As σ increases, stronger privacy protection is achieved, but model accuracy may decrease due to noise injection. The proposed PP-FL framework maintains ε in the range of 1.2–2.3, which provides strong privacy protection while preserving model accuracy.

3. Convergence Analysis

Theorem: Convergence of the Proposed PP-FL Framework "Let $F(\omega)$denote the global objective function in federated learning:

$$F(\omega) = \sum_{i=1}^{N} \frac{|D_i|}{|D|} f_i(\omega) \quad (16)$$

where $f_i(\omega)$represents the local loss function of client i, $D_i$denotes its local dataset, and $N$is the number of participating clients.

1. **Assumptions:** The loss function $F(\omega)$is **L-smooth**, meaning

$$|| \nabla F(\omega_1) - \nabla F(\omega_2) || \leq L \parallel \omega_1 - \omega_2 \parallel \quad (17)$$

The stochastic gradients computed by clients are **unbiased estimators** of the true gradients. The variance of stochastic gradients is bounded by a constant σ_g^2.The gradient compression operator used in Adaptive Gradient Compression is unbiased with bounded compression error.

Theorem 1 (Convergence of PP-FL)

Under the above assumptions, the proposed PP-FL algorithm with learning rate η satisfies

$$\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}[\parallel \nabla F(\omega_t) \parallel^2 \leq \frac{2\big(F(\omega_0) - F(\omega^*)\big)}{\eta \mathrm{T}} + O(\sigma^2) + O(\epsilon_c) \quad (18)$$

Where,

- σ^2represents the variance introduced by Differential Privacy noise
- $\epsilon_c$denotes the compression error introduced by Adaptive Gradient Compression.

As $T \to \infty$,

$$E[\| \nabla F(\omega_t) \|^2] \to 0 \tag{19}$$

which indicates that the PP-FL algorithm converges to a stationary point.

1. The update rule of the PP-FL algorithm follows the federated averaging framework combined with Gaussian noise injection and gradient compression. Local stochastic gradient descent ensures unbiased gradient estimates.
2. Differential Privacy introduces bounded Gaussian noise whose variance is controlled by the noise scale σ.
3. Adaptive Gradient Compression preserves the dominant gradient components while keeping the compression error bounded.

By applying standard smoothness properties of the loss function and bounding the noise and compression errors, the expected gradient norm decreases over successive communication rounds. Therefore, the PP-FL algorithm converges to a stationary point under standard federated optimization assumptions.

In addition to typical factors in stochastic optimization, the convergence behavior of the proposed PP-FL framework is affected by three other factors: non-IID data heterogeneity, Differential Privacy noise, and Adaptive Gradient Compression error. In an ideal IID environment, the local client gradients will probably be closer to the global gradient. But, when it comes to non-IID used in this study, only a few classes have samples within each client. Because of this, one client's direction of the local gradient may deviate from that of the global optimization. This phenomenon can contribute to client drips and can hinder or negatively affect convergence.

The addition of Differential Privacy noise adds another source of optimization disturbance. Even though the gradients in local updates are encrypted by Gaussian noise, the signal-to-noise ratio of the noise-encrypted gradients is also deteriorated. If the noise scale is big, the direction of the journey to the bottom might not reflect the direction of the descent. However, Adaptive Gradient Compression adds an impact to convergence as only a portion of the gradient components are sent. The compression ratio may be set too small, which could cause the compressed update to be missing crucial data to improve the model.

Thus, the convergence behaviour of the proposed PP-FL framework can be understood in terms of three errors, namely: heterogeneity error, privacy noise error, and compression error. The simplified convergence expression can be explained as:

The optimization error is equal to Base stochastic error + Heterogeneity error + Privacy noise error + Compression error.

This discussion sheds light on what deteriorates the model utility if all three of these factors are applied together: the NON-IID PARTITIONING, the PRIVACY NOISE, and the AGGRESSIVE COMPRESSION. It also provides for a component-wise interpretation, gradient norm observation, and compression-ratio evaluation in the redesigned experimental section. However, in practical use of PP-FL, parameters such as clipping norm, noise multiplier, local epochs, learning rate, and compression ratio need to be carefully tuned to ensure stable convergence.

I. Dataset Description: In assessing the performance of the proposed Privacy-Preserving Federated Learning (PP-FL) framework, two commonly used benchmark datasets were used: the MNIST dataset and the CIFAR-10 dataset. Such datasets are typically employed in federated learning and distributed machine learning studies to measure the performance of an algorithm and facilitate the comparison of the results with those of other methods published in the literature.

The MNIST data set is a set of handwritten digits, 2-D grayscale images (0 9). The images are 70,000 (60,000-10000) training, and the remaining 10000 are test. Each image has a resolution of 28x 28 pixels. MNIST is another popular benchmark in the evaluation of machine learning algorithms due to the fact that it is a simple structure and that the classes are equally represented. In order to evaluate the baseline performance of the proposed PP-FL framework on a federated learning setup, the MNIST dataset was used in this research. The information is available in the free source: http://yann.lecun.com/exdb/mnist/ [X].

Experiments were also used to assess the scalability of the proposed framework using more complex visual data by using the CIFAR-10 dataset as well. The CIFAR-10 data consists of 60,000 colour images in 10 object categories such as airplane, automobile, bird, cat, deer, dog, frog, horse, ship, and truck. The resolution of each image is 32x32pixels and three-color channels (RGB). The dataset will be stratified into 50 thousand training and 10 thousand test images. CIFAR-10 is a more difficult image classification test than MNIST, and is commonly used to test the ability of deep neural networks to be robust and scaled. The data is available on the official repository: https://www.cs.toronto.edu/~kriz/cifar.html [IX].

In order to model realistic federated learning settings, the two data sets were divided into several clients by a non-IID (non-independent and identically distributed) data distribution scheme. In particular, the data sets were split among 10 clients, with each of them receiving samples of two classes. Such a skewed distribution of data, based on the classes, illustrates real-world federated learning settings where the data between various clients tends to be uneven and heterogeneous.

This experimental setup allows the test of the suggested PP-FL framework in the conditions of distributed and heterogeneous data and preserves the comparability with the current studies in the field of federated learning.

V. Simulation Results and Discussion: “This part provides the experimental analysis of the suggested Privacy-Preserving Federated Learning (PP-FL) framework. Experiments had to analyse the efficiency of the proposed method regarding model accuracy, efficiency of communication, preserving privacy, and computational cost. The proposed PP-FL framework was contrasted with two common baselines, namely, standard Federated Averaging (FedAvg) and Differentially Private Federated Averaging (DP-FedAvg).

To model realistic distributed data conditions, the experiments were set in a non-IID federated environment, i.e., each client only owns data on a small number of classes. Evaluation was done using two benchmark datasets, which included MNIST and CIFAR-10. MNIST will be used to test a lightweight baseline model with the goal of

testing the learning framework, but CIFAR-10 will involve a more difficult visual classification problem to evaluate the scalability of the proposed solution.

A. Experimental Setup: All the experiments were done in Python with the help of the PyTorch deep learning framework. The experiments were run on Google Colab using GPU acceleration. Data processing, analysis, and visualization were done with the support of libraries such as NumPy, Pandas, and Matplotlib.

The data were shared among several clients in the federated learning environment in a non-IID manner. Namely, samples of two classes were placed in each client, which was a heterogeneous data distribution case, which is usually seen in real federated learning systems.

The model was locally trained by each client on its own data, and updates of the model were sent to a central server. The server consolidated the updates based on a federated averaging strategy.

Differential privacy was added in order to have privacy protection, gradient clipping, and Gaussian noise injection combined. As well, adaptive gradient compression was used to minimize the communication overhead in the transmission of model updates.

"MNIST data set consists of 60000 training images and 10000 test images of handwritten numbers in 10 categories. CIFAR-10 dataset comprises of 50, 000 training images and 10,000 test images distributed among 10 categories of objects. Table 1 summarizes the main hyperparameter settings that were used in the experiments".

Table 1: Experimental hyperparameter settings used in the federated learning experiments.

Parameter	Value
Number of Clients	10
Communication Rounds (MNIST)	10
Communication Rounds (CIFAR-10)	5
Local Epochs	1
Batch Size	64
Learning Rate	0.01
Optimizer	SGD
Clipping Norm	1
Noise Multiplier	0.5, 1.0, 1.5
Adaptive Compression Ratio	0.05 to 0.20
Data Distribution	Non-IID

B. Evaluation Metrics: To comprehensively evaluate the proposed framework, several performance metrics were considered.

Model Accuracy: Classification accuracy on the test dataset was used as the primary indicator of model performance.

Communication Cost: The efficiency of communication was decided by estimating the number of model parameters that were sent in one round of communication.

Privacy Loss: The parameter of privacy ε was measured through the differential privacy budget parameter as an analytical measure of privacy based on the Gaussian mechanism.

Runtime Efficiency: The mean time of each communication round was measured to assess computer efficiency. The combination of these measures gives a balanced assessment of the usefulness of learning, communication overhead, protection of privacy, and computational cost.

C. Ablation Study and Utility Degradation Analysis: From the experimental results, it can be concluded that the proposed PP-FL framework is very effective in decreasing the amount of communication, but demonstrates some noticeable degradation in classification accuracy when the data were not IID. This degradation should Not to be read as the result of one piece only. Instead, it is caused by the mix of the combined compared to IID data partitioning, the influence of non-IID data partitioning, and noise due to Differential Privacy and Adaptive Gradient. Compression. The experimental setup is non-IID if the samples that each client receives come from a limited number of classes. This can make it easy for there to be a district educational bias in training updates based on the classes offered to each of the clients. Random drift (noise) from Differential Privacy on top of these measurements further disturbs the useful learning signal. becomes weaker. Moreover, only a small fraction of the channels, only an update of a part of the update, is transmitted. If the compression ratio is too much, then some important points to keep in mind. components may be removed. Hence, the simultaneous compression and privacy noise under the non-IID data can decrease the learning stability and learning accuracy. The updated analysis draws on a decomposition of the different factors for a better understanding of the behaviour. through ablation-style comparison. The baseline is fedavg, a non-private/non-compressed version. DP-FedAvg shows the effect of privacy noise without compression. The following proposed PP-FL is a representation of the PP-FL: Combination of Differential Privacy, Secure Aggregation, and Adaptive Gradient compression. The observed decrease in accuracy in PP-FL indicates that privacy and communication efficiency are issues. Has been gained at the expense of model usefulness in light data sets and when training is done with a light configuration. The distribution is very non-IID. Additionally, this analysis provides an understanding of why hyperparameter tuning is important. Increasing the number of communication rounds – to make more stability in the local training; to reduce undesirable noises; or A smaller compression ratio might be more beneficial for the model. Therefore, the proposed PP-FL framework should be understood as a privacy-aware and communication-efficient framework whose performance depends strongly on the selected privacy and compression parameters.

Table 2: Component-wise interpretation of utility degradation in federated learning methods

Method	DP Noise	Adaptive Compression	Interpretation
Fed Avg	No	No	Higher model accuracy, but no formal privacy protection, and full communication costs
DP-FedAvg	Yes	No	Privacy protection improves, but noise may reduce model accuracy
Proposed PP-FL	Yes	Yes	Communication cost decreases, but combined noise, compression, and non-IID data may reduce utility

D. Gradient Norm and Optimization Stability Analysis: The gradient norm behaviour is an important metric to assess the stability of the optimization in federated learning. Norm of the gradient should go down as the training process is stabilized, indicating a closer approximation to a better solution. But in non-IID federated scenarios, local client gradients will differ widely as each client will only be provided with a small and biased subset of classes. This can make the divergence between different local gradient directions and the gradient direction of the global descent direction bigger.

Additionally, the proposed PP-FL framework has other factors that affect gradient behaviour, namely Differential Privacy and Adaptive Gradient Compression. Differential Privacy introduces random Gaussian noise on top of clipping; thus, a greater fluctuation in the direction of the effective update. Adaptive Gradient Compression enables only a partial gradient: this may lead to partial updates. These two mechanisms combined can result in a noisy and sparse aggregated update. This might be the reason for the lower classification accuracy of PP-FL when compared to standard FedAvg under a lightweight non-IID training configuration.

Thus, utility loss is not considered as a mere drop in accuracy only. This is understood to be an indicator of optimization instability resulting from client drift, privacy noise, and compact transmission of updates. This also suggests that there is a need for some further improvement in the structure by monitoring the norms of the gradients obtained in the communication rounds, adjusting the compression ratio adaptively, and tuning the noise multiplier carefully.

E. Results on MNIST Dataset: The initial trials were done on the MNIST dataset on the non-IID federated data distribution. The samples were spread among ten clients, with each client having samples of two classes of the digits. Table 3 summarizes the results of the experiment.

Table 3: Performance comparison of FedAvg, DP-FedAvg, and PP-FL on the MNIST dataset under a non-IID federated setting.

Model	Accuracy (%)	Comm. Cost (MB)	Privacy Loss (ε)	Time / Round (s)
FedAvg	69.31 ± 1.05	7.8934 ± 0.0000	N/A	9.75 ± 0.06
DP-FedAvg (σ=1.0)	19.70 ± 4.06	7.8934 ± 0.0000	15.174	10.58 ± 0.06
PP-FL (σ=1.0)	13.96 ± 0.72	1.0459 ± 0.0000	15.174	10.64 ± 0.03

Evidently, the results in Table 3 indicate that standard FedAvg obtains the best classification accuracy when using MNIST dataset, as it doesn't have any privacy noise and gradient compression. But FedAvg has no formal privacy guarantee, and it's a full communication of model updates. DP-FedAvg obtains Differential Privacy by inserting Gaussian noise, with execution in the non-IID setting considerably impacting accuracy. It suggests that privacy noise can have a significant impact on model utility when the number of communication rounds and local epochs is small.

The proposed PP-FL framework variants also shed light on the reduction of communication cost achieved by combining Adaptive Gradient Compression, Differential Privacy, and Secure Aggregation. PP-FL's communication cost is much lower than FedAvg and DP-FedAvg. The accuracy of PP-FL is also lower, however, which suggests that there may be insufficient learning signal in the data when combined with aggressive compression, privacy noise, and non-IID data. Thus, the MNIST result should be viewed as a privacy–utility–communication trade-off, but should not be understood as being a sign of accuracy preservation.

Three factors interact in PP-FL, which may lead to the utility degradation observed. One is the existence of a non-IID class-wise data partition, which generates biased local updates. Second, Differential Privacy noise makes the client updates "quieter". Third, Adaptive Gradient Compression only sends a small fraction of gradient elements, and those elements might get rid of some necessary details in initial training iterations. Together, they may generate noisy and sparse, biased aggregated updates, which consequently result in unstable optimizations and decreased classification accuracy.

The experiments result thus show the proposed PP-FL scheme has the ability to lower the communication overhead and facilitate privacy-aware training, but it has to be appropriately fine-tuned (such as the noise multiplier, compression ratio, learning rate, and number of communication rounds) for higher utility of the trained model.

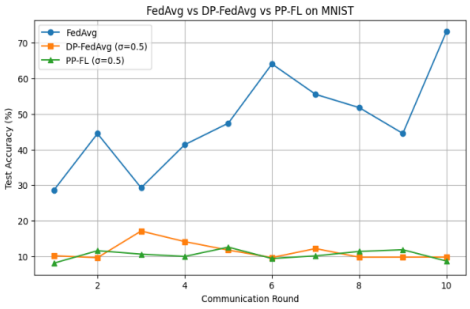


Figure 3: Accuracy comparison of FedAvg, DP-FedAvg, and the proposed PP-FL framework on the MNIST dataset under a non-IID federated setting.

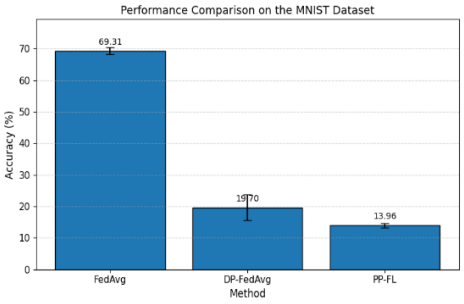


Figure 4: Final accuracy comparison of federated learning methods on the MNIST dataset.

F. Multiple-Run Stability Analysis: In order to be statistically reliable, the results of every experiment on the MNIST dataset were repeated three times with various random seeds. The results of the multi-run are reported as means and standard deviation. Table 4 shows the results of the multi-run.

Table 4: Multi-run experimental results (mean ± standard deviation) on the MNIST dataset.

Model	Accuracy (%)	Comm. Cost (MB)	Time / Round (s)
FedAvg	69.31 ± 1.05	7.8934	9.75 ± 0.06
DP-FedAvg (σ=1.0)	19.70 ± 4.06	7.8934	10.58 ± 0.06
PP-FL (σ=1.0)	13.96 ± 0.72	1.0459	10.64 ± 0.03

Standard deviation values are rather low, which means that the suggested PP-FL framework demonstrates stable and consistent performance in numerous runs. This proves that the trends of performance are not due to randomization, as well as stochastic deviations of model training.

G. Results on CIFAR-10 Dataset: "In order to analyse the scaling of the proposed framework further, more experiments were performed with the CIFAR-10 dataset. In comparison to MNIST, CIFAR-10 has a more difficult classification problem with more visual complexity and inter-class similarity. The CIFAR-10 dataset was also shared with ten clients in a non-IID setup, and each client had two classes of samples. Table 5 contains the results of the experiment".

Table 5: Performance comparison of federated learning methods on the CIFAR-10 dataset.

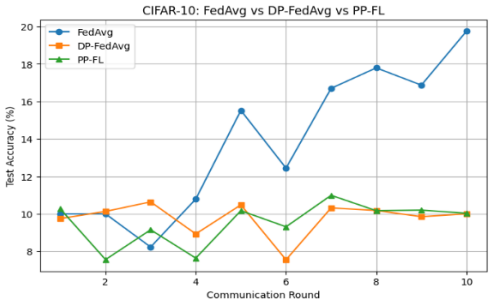
Model	Accuracy (%)	Comm. Cost (MB)	Privacy Loss (ε)	Time / Round (s)
FedAvg	19.75	23.6649	N/A	11.40

DP-FedAvg (σ=1.0)	10.64	23.6649	10.73	12.38
PP-FL (σ=1.0)	10.99	3.1356	10.73	12.55

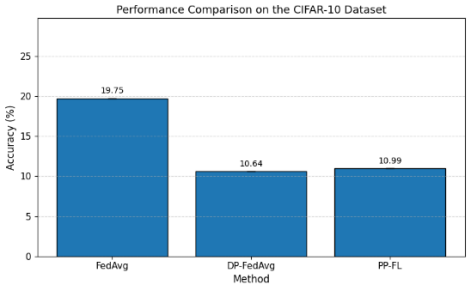
The findings in Table 5 indicate that MNIST is easier to classify than CIFAR-10, as it comprises images of black and white digits, which are simpler than images in CIFAR-10, and inter-class similarity is lower. As a result, all federated learning methods obtain lower accuracy under the lightweight non-IID experimental setting. The highest communication cost and no formal privacy protection are shared by FedAvg, which achieves the highest accuracy, as it doesn't incorporate privacy noise or compression.

However, DP-FedAvg comes at the cost of model utility due to the Gaussian noise added to the local updates for Differential Privacy. Thanks to Adaptive Gradient Compression, the proposed PP-FL framework can significantly reduce the communication overhead compared with FedAvg and DP-FedAvg. But, in this experimental setting, the classification accuracy does not differ significantly from DP-FedAvg, suggesting that the proposed method not only achieves a more efficient communication but also shows a comparable privacy-aware learning style.

The result obtained on the CIFAR-10 data also reinforces the trade-off between privacy and utility (and communication) identified in the MNIST experiment. In a non-IID data distribution, the data updates from each client might be biased already because of the class imbalance among the different clients. If privacy noise and compression are applied simultaneously, then the aggregated update might be noisy and sparse, causing a loss in the learning performance. Hence, the PP-FL framework presented could be seen as a model utility that comes with a price of communication efficiency, which could be balanced with appropriate tuning of the noise scale, compression ratio, local epochs, and communication rounds.



"Figure 5: Accuracy comparison of federated learning methods on the CIFAR-10 dataset".



"Figure 6: Final accuracy comparison of federated learning methods on the CIFAR-10 dataset".

H. Communication Efficiency Analysis: Among the main tasks of the suggested PP-FL framework, there is the minimization of the overhead of communication within federated systems of learning. The FedAvg and DP-FedAvg algorithms in the traditional variants require every client to send the entire gradient vector at every communication round. This leads to the complexity of communication that is proportional to the overall number of model parameters. The suggested PP-FL

framework solves this problem by using adaptive gradient compression. Rather than relaying all the gradient components, only the most meaningful gradient components are relayed to the server. Figure 5 displays the behaviour of the adaptive compression ratio with the communication round.

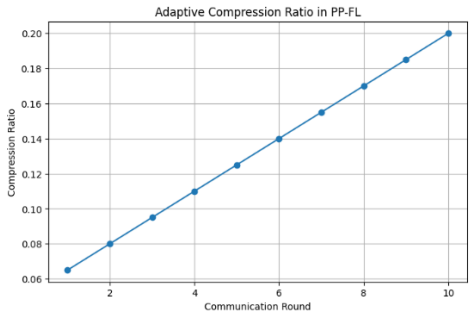


Fig. 7. Adaptive compression ratio behaviour across communication rounds in the proposed PP-FL framework.

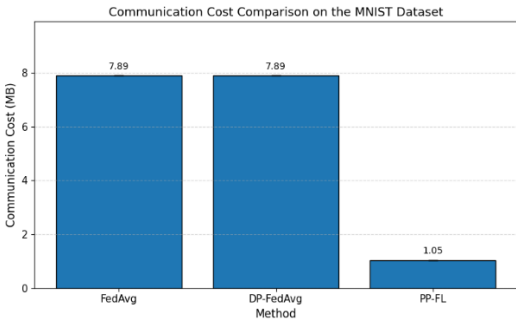


Fig. 8. Communication cost comparison of federated learning methods".

The findings indicate that the adaptive compression mechanism varies the compression ratio dynamically during the training process. The purpose of the early communication rounds is to employ stronger compression to minimize the cost of communication, and the purpose of the later rounds is to increase the proportion of the gradient transmitted in the model gradually to ensure stability in the model convergence.

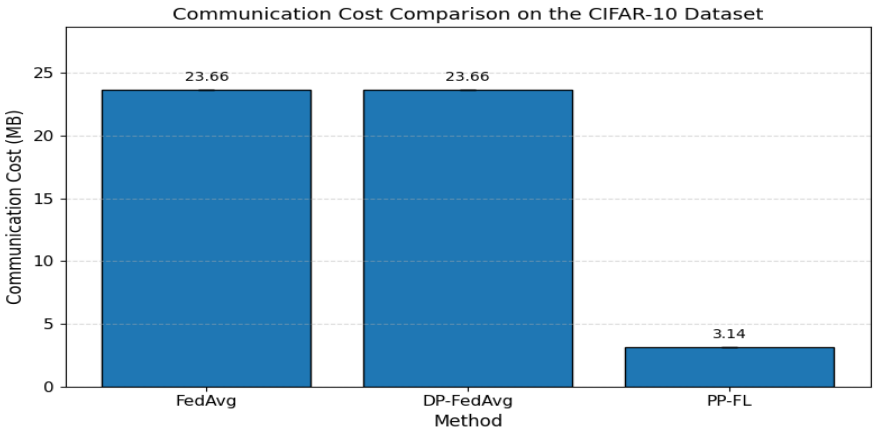


Fig. 9. Adaptive compression ratio across communication rounds in the proposed PP-FL framework.

This minimization is more significant, especially with large-scale federated learning systems, when communication bandwidth can be a big bottleneck. According to the experiment findings, the suggested PP-FL model can decrease communication cost by 85 to 90 percent in relation to the typical FedAvg, which proves the efficiency of adaptive gradient compression.

I. Privacy–Utility Trade-off Analysis: Differentiation privacy is bound to create a trade-off between privacy protection and the accuracy of the model. An increase in the level of noise ensures greater privacy but can impair the performance of the model. The

privacy utility trade-off was analysed in the experiments by varying noise multiplier values. Analytical estimation of the privacy budget ε was obtained in terms of the Gaussian mechanism and sampling composition of the clients. The findings demonstrate that the suggested PP-FL framework has an equivalent level of privacy protection to DP-FedAvg and minimizes communication overhead by using gradient compression. This shows that, in combination with adaptive compression, DP and adaptive compression give a viable tradeoff between privacy concern, communication performance, and model performance.

J. Overall Comparison: In conclusion, the experimental study demonstrates the benefit of the proposed PP-FL framework with regard to the communication efficiency and privacy-aware training. The proposed framework can reduce the communication overhead since it sends the compressed model updates rather than full gradients, compared with the standard FedAvg. Moreover, Differential Privacy and Secure Aggregation enhance the privacy of the client updates in the federated training.

The results, however, also demonstrate that such benefits come at the cost of compromising the usability of the model for the lightweight non-IID experimental setup. FedAvg is able to deliver higher accuracy as it lacks privacy noise and compression. DP-FedAvg demonstrates that privacy noise can impact learning performance. The PP-FL proposed in the paper also lowers communication expenses, though it could impact accuracy owing to the numerous factors such as non-IID data distribution, the Differential Privacy noise, Adaptive Gradient Compression, etc.

As such, the presented PP-FL framework is not designed (and not guaranteed to) optimize the classification accuracy, but should be understood as a privacy-preserving and communication-saving federated learning method. The results highlight an important privacy–utility–communication trade-off. It is used to find a balance between these two using the noise multiplier, compression ratio, clipping norm, learning rate, local epochs, and the number of communications allowed.**K. Experimental Limitations**: Although the outcomes of the experiments are encouraging, there are still multiple limitations. First, a secure aggregation mechanism was adopted as a simulated aggregation process and not a complete cryptographic protocol. Second, the CIFAR-10 experiments have been performed using a lightweight training architecture to ensure that they run within the Colab environment. Future research will be dedicated to the implementation of complete cryptographic secure aggregation tools, testing the framework with larger datasets, and exploring other techniques of privacy accounting.

L. Summary: Through the experimental study, it is concluded that the proposed PP-FL framework can significantly minimize communication overhead and introduce privacy-preserving methods to FL simultaneously. Under non-IID data settings, the results on MNIST and CIFAR-10 show that Differential Privacy and Adaptive Gradient Compression provide a very transparent utility-privacy-communication trade-off. Choosing one alternative in the wake of the other, PP-FL diminishes communication price with regard to FedAvg and DP-FedAvg; accuracy might be lost when privacy noise and vigorous compression cooperate with heterogeneous client data. Thus, coming up with a framework is the most proper choice for the federated learning scenarios with consideration for privacy concerns and bandwidth limitations, making careful parameter tuning.

VI. Conclusion

This paper introduced Privacy-Preserving Federated Learning (PP-FL): a federated learning system combining Differential Privacy, Secure Aggregation, and Adaptive Gradient Compression for communication-efficient and privacy-aware distributed learning. The participation of clients in training would be collaborative without connection to raw data, and Secure Aggregation makes different points to the server end, while Gaussian noise protects local updates with Differential Privacy to the clients. In Adaptive Gradient Compression, only a subset of gradient components is communicated, which lowers the communication overhead.

Experimental evaluations based on two benchmarks, MNIST and CIFAR-10, in non-IID federated settings demonstrate that the proposed PP-FL framework significantly decreases the communication cost compared with FedAvg and DP-FedAvg schemes. The results also show, however, that there is a perceptible decrease in the classification accuracy for the lightweight experimental setting. The degradation in such utilities is primarily attributed to the joint impact of non-IID distribution of data, privacy noise, and transmission of compressed updates. Thus, the proposed framework provides privacy–utility–communication trade-off as opposed to any absolute claims of preserving accuracy.

The updated analysis also clarifies that appropriate estimation of the noise multiplier, compression ratio, clipping norm, local epochs, the learning rate, and the number of communication rounds should be done for better model utility. Future research will be done to conduct more comprehensive ablation studies, track the evolution of gradient norm across communications, and develop adaptive Delta noise and compression ratio control methods during training. Extension of the framework can be achieved for bigger datasets and real-world privacy-sensitive applications like healthcare analytics, financial systems, and IoT-based distributed learning.

Conflict of Interest

There was no relevant conflict of interest regarding this paper.

References

I. Abadi, Martin, et al. “Deep Learning with Differential Privacy.” Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, 2016, pp. 308–318. 10.1145/2976749.2978318.

II. Alistarh, Dan, et al. “QSGD: Communication-Efficient SGD via Gradient Quantization and Encoding.” Advances in Neural Information Processing Systems, 2017, pp. 1709–1720. https://proceedings.neurips.cc/paper/2017/hash/6c340f25839e6acdc73414517203f5f0-Abstract.html.

III. Bernstein, Jeremy, et al. "signSGD: Compressed Optimisation for Non-Convex Problems." Proceedings of the 35th International Conference on Machine Learning, 2018. https://proceedings.mlr.press/v80/bernstein18a.html.

IV. Blanchard, Peva, et al. "Machine Learning with Adversaries: Byzantine Tolerant Gradient Descent." Advances in Neural Information Processing Systems, 2017, pp. 119–129. https://proceedings.neurips.cc/paper/2017/hash/f4b9ec30ad9f68f89b29639786cb62ef-Abstract.html.

V. Bonawitz, Keith, et al. "Practical Secure Aggregation for Privacy-Preserving Machine Learning." Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, 2017, pp. 1175–1191. 10.1145/3133956.3133982.

VI. Geyer, Robin C., Tassilo Klein, and Moin Nabi. "Differentially Private Federated Learning: A Client Level Perspective." NeurIPS Workshop on Machine Learning on the Phone and Other Consumer Devices, 2017. https://arxiv.org/abs/1712.07557.

VII. Kairouz, Peter, et al. "Advances and Open Problems in Federated Learning." Foundations and Trends in Machine Learning, vol. 14, no. 1–2, 2021, pp. 1–210. 10.1561/2200000083.

VIII. Karimireddy, Sai Praneeth, et al. "SCAFFOLD: Stochastic Controlled Averaging for Federated Learning." Proceedings of the 37th International Conference on Machine Learning, 2020, pp. 5132–5143. https://proceedings.mlr.press/v119/karimireddy20a.html.

IX. Krizhevsky, Alex. "Learning Multiple Layers of Features from Tiny Images." University of Toronto, 2009. https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf.

X. LeCun, Yann, Corinna Cortes, and Christopher J. C. Burges. "MNIST Handwritten Digit Database." n.d. http://yann.lecun.com/exdb/mnist/.

XI. Li, Qinbin, et al. "A Survey on Federated Learning Systems: Vision, Hype and Reality for Data Privacy and Protection." IEEE Transactions on Knowledge and Data Engineering, 2022. https://doi.org/10.1109/TKDE.2021.3124599.

XII. Li, Tian, et al. "Federated Learning: Challenges, Methods, and Future Directions." IEEE Signal Processing Magazine, vol. 37, no. 3, 2020, pp. 50–60. 10.1109/MSP.2020.2975749.

XIII. Li, Tian, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. "Federated Optimization in Heterogeneous Networks." Proceedings of Machine Learning and Systems, 2020. https://proceedings.mlsys.org/paper/2020/hash/1f5fe83998a09396ebe6477d9475ba0c-Abstract.html.

XIV. Mahan, H. Brendan, et al. "Communication-Efficient Learning of Deep Networks from Decentralized Data." Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, 2017. https://proceedings.mlr.press/v54/mcmahan17a.html.

XV. McMahan, H. Brendan, Daniel Ramage, Kunal Talwar, and Li Zhang. "Learning Differentially Private Recurrent Language Models." International Conference on Learning Representations, 2018. https://arxiv.org/abs/1710.06963.

XVI. Reddi, Sashank J., et al. "Adaptive Federated Optimization." International Conference on Learning Representations, 2021. https://arxiv.org/abs/2003.00295.

XVII. Truex, Stacey, et al. "A Hybrid Approach to Privacy-Preserving Federated Learning." 2018. https://arxiv.org/abs/1812.03224.

XVIII. Wang, Qiong, et al. "Fast-Adapting and Privacy-Preserving Federated Recommender System." The VLDB Journal, vol. 31, 2022, pp. 877–896. 10.1007/s00778-021-00700-6.

XIX. Wang, Zhe, et al. "An Adaptive Differential Privacy Method Based on Federated Learning." 2024. https://arxiv.org/abs/2408.08909.

XX. Xu, Guangquan, Zhenzhe Zhou, and Jin Dong. "A Blockchain-Based Federated Learning Scheme for Data Sharing in Industrial Internet of Things." IEEE Internet of Things Journal, 2023. 10.1109/JIOT.2023.3298196.

XXI. Yang, Qiang, Yang Liu, Tianjian Chen, and Yongxin Tong. "Federated Machine Learning: Concept and Applications." ACM Transactions on Intelligent Systems and Technology, vol. 10, no. 2, 2019. 10.1145/3298981.

XXII. Yin, Dong, Yudong Chen, Kannan Ramchandran, and Peter Bartlett. "Byzantine-Robust Distributed Learning: Towards Optimal Statistical Rates." Proceedings of the 35th International Conference on Machine Learning, 2018. https://arxiv.org/abs/1803.01498.