A ES

ISSN: 2096-3246
Volume 58, Issue 01, 2026

PERFORMANCE EVALUATION OF FEDERATED LEARNING FOR PRIVACY-PRESERVING STUDENT PERFORMANCE PREDICTION

Gourav Arora[1*], Touseef Ahmad lone [2†] and Shreya Gandhi[3†]

[1*]Department of Computer Science & Engineering, CT University, Punjab, India.
[2]Department of Computer Science & Engineering, CT University, Punjab, India.
[3]Department of Computer Science & Engineering, CT University, Punjab, India.

*Corresponding author(s). E-mail(s): gourava256@gmail.com; Contributing authors: lonetouseef99@gmail.com; shreya24196@ctuniversity.in;
[†]These authors contributed equally to this work.

Abstract

Effective institutional decision-making requires an accurate prediction of stu- dent performance. Nevertheless, centralized machine learning systems tend to be incompatible with rigid data protection policies, such as GDPR and FERPA. Federated Learning (FL) is a decentralized approach, but it loses its performance in the case of statistical heterogeneity (non-IID data) that occurs in multi- institutional educational environments. Moreover, the majority of the available approaches have the ability to treat data heterogeneity or data privacy, but rarely at the same time. This paper presents a privacy-enhancing and scalable model of student performance prediction. It is based on a three-layer Differentially Private Federated Proximal Optimization (DP-FedProx) Deep Neural Network architec- ture. To mitigate client drift and maintain global convergence in skewed CGPA distributions, the system includes proximal regularization (μ). Gaussian meth- ods and gradient clipping are used to accomplish differential privacy. The four configurations—FedAvg and DP-FedProx in the context of IID and Non-IID envi- ronments—have been subjected to a detailed comparative study. Convergence behavior, privacy budget evolution, and predictive performance are evaluated using the progressive communication round. In accordance with experimental data, the proposed DP-FedProx system has a worldwide accuracy of roughly 82 percent with severely skewed non-IID under constrained privacy budgets. Com- pared to traditional FedAvg, DP-FedAvg, and non-private FedProx baselines, the suggested method offers more convergence stability and an acceptable trade- off between privacy and utility. The results show that neural networks, proximal regularization, and differential privacy provide an efficient and scalable solution to guarantee the security of academic analytics across several institutions.

Keywords: Federated Learning (FL), FedProx, Differential Privacy (DP), IID, Non-IID, Privacy-Preserving, Students Data Privacy (SDP)

ISSN: 2096-3246
Volume 58, Issue 01, 2026

1 INTRODUCTION

The growing trend of digitalization of educational ecosystems has given way to the popularity of predictive analytics in identifying the at-risk students, enhancing retention strategies, and facilitating individual learning journeys [12],[29]. Institutional decision-making now heavily relies on machine learning and deep learning-driven student performance models for prediction [25],[29]. However, traditional centralized learning approaches need the centralization of sensitive student data, including demographics, behavioral characteristics, and academic history [13],[18]. The General Data Protection Regulation (GDPR) and the Family Educational Rights and Privacy Act (FERPA) are two examples of strong ethical, legal, and regulatory requirements that make these centralized architectures extremely challenging [7],[34]. FL is a revolutionary approach that does not require institutions to disclose the raw data and can turn the model training process into a collaborative endeavour [16],[37]. Since FL will store the data locally and only provide model changes, it is also preventing direct access to sensitive student information [2],[26]. The simplest federated optimization technique, Federated Averaging (FedAvg), has proven to be the mainstay of federated optimization because of its empirical performance and ease of use [15],[34]. Nonetheless, some institutions gather non-identically and independently (non-IID) educational data [39],[40]. The conventional FedAvg implementations have poor worldwide performance because of statistical heterogeneity, which showed up as variations in curriculum design, student demographics, grading methods, and CGPA distributions [3],[24].

Federated Proximal Optimization FP employs a proximal regularization target to mitigate heterogeneity by restricting model updates to local models in order to bring the models closer to the global model and lessen the distance between the clients [24]. FedProx is more stable in non-IID environments, but it has no formal privacy assurances [11],[34]. Differential privacy, on the other hand, is a statistical privacy assurance that information will leak to the model parameters or gradients through noise [7],[15]. Noise would cause instability in the case of diverse data distributions even while DP-based federated models guarantee privacy [11],[22]. This research gap concerns how to ensure formal privacy and a high degree of resistance to non-IID heterogeneity with- out sacrificing prediction accuracy [11],[22]. To address this gap, the current paper

provides a new Three-layered Deep Neural Network and a Differentially Private Federated Proximal (DP-FedProx) model to predict student performance [14],[25]. The model uses proximal regularization to minimize client drift using gradient clipping and Gaussian noise injection to ensure (ϵ, δ)-differential privacy, whose privacy is mea- sured by a R´enyi Differential Privacy accountant [7],[15]. It focuses on the compromise between privacy, convergence stability, and predictive accuracy in multi-institutional education [1],[12]. We take the methodical technique of investigating the experimental substratum in stages, first adding the numbers of clients and communication sessions to have a model of the magnitude of the intended institution environment, in order to test the strategy offered to us with the required rigor [1],[32]. For both IID and non-IID distributions, the model has been equivalent to standard Federated Aver- aging [15],[39]. Performance metrics include accuracy, mean squared error (MSE), convergence behavior, and privacy budget creation [10],[11].

1.1 Contributions

The crucial topics of data privacy, institutional silos, and statistical heterogeneity are particularly covered in this work [11],[34],[39]. Additionally, we propose DP-FedProx, a novel privacy-conscious architecture designed to forecast student performance in non-homogeneous environments [14],[25],[40]. The framework combines the two potent approaches:

- **Client-Side Formal Privacy through Opacus:** We disrupt the client-side optimization loop. Using the Opacus differential privacy engine, we provide math- ematically quantifiable Differential Privacy (DP) guarantees. High-fidelity gradient clipping (containments) and subsequent injection of calibrated Gaussian noise, con- strained by strict Renyi Differentiated Privacy (RDP) accounting, are used before externalizing model parameters to the central server [7],[15],[2].
- **Server-Side Robustness through Proximal Regularization:** To balance the conflicting revision received by heterogeneous nodes and alleviate the localized drift that is created by non-IID data as well as DP noise, we use FedProx optimization [24],[6]. A specialized Proximal Regularization Term is used to stabilize and make decentralized learning cohesive, which emphasizes the integrity of the global model rather than the local overfitting [24],[6].

We provide a strict empirical supporting DP-FedProx framework on a challenging 11-feature behavioral and academic dataset, shared among eight institutional nodes

among twenty-five communication rounds with intensive localized training (30 epochs). Our findings can create a powerful state-of-the-art that has achieved ∼82%accuracy whilst adhering to stringent privacy limits. This paper offers a scaffold able design of high-utility academic analytics which honors institutional autonomy and legal privacy limitations.

This paper contributes to four keyways:

- This paper purposes a federated neural network model for student performance prediction.
- This study examines the stability of differentially privatized federated optimization through proximal regularization.
- The privacy and utility trade-offs are compared using an empirical analysis based on increasing the complexity of communication in steps.
- To compare federated, and privacy-preserving federated models under IID and non- IID settings and competitive accuracy **(∼82%)** with non-IID conditions that are highly skewed.

2 REVIEW OF LITERATURE

Predicting student performance has emerged a key theme in the Educational Data Mining (EDM) and learning analytics because it can be used to enhance early inter- vention strategies, customized academic support, and institutional decision-making [27],[30]. Conventionally, most predictive systems have been based on centralized machine learning models whereby student data across various departments or insti- tutions are pooled together into just one repository to be used to train the models [14]. Despite the high predictive performance that is usually obtained with centralized methods, there are considerable issues of privacy, security and data ownership and regulatory compliance [12],[13],[34]. These issues have motivated the growing number of studies in predicate frameworks that are decentralized and privacy conscious.

The Federated Learning (FL) paradigm has been extensively identified as a prospective approach to privacy-sensitive collaborative analytics. In student perfor- mance prediction on Python programming education, **Chen and Qi (2025)** suggested an entropy-adaptive federated learning system [4]. In order to balance privacy protec- tion with predictive accuracy, their method proposes an adaptive noise technique that blends federated learning with differential privacy. This paper highlight that data on learning behavior can be used in a privacy-sensitive way to predict academic perfor- mance with minimal disclosure of confidential records. Nonetheless, the framework is restricted to a

course of context and does not offer more general comparison among federated optimization tactics, larger institutional data, or various IID and Non-IID designs.

Likewise, **Zhang et al. (2023)** suggested a federated learning style outcome prediction model that included mechanisms of multiple layers of privacy protection [31]. Their research shows there is a possibility of collaborative predictive analytics without sending student level data to a central server. The research makes the FL applicable in education by incorporating several privacy protection measures. How- ever, optimizer-specific behaviors like FedProx have not been profoundly examined by the research, and no extensive benchmarking of centralized baselines or differentially private federated models has been provided.

A comparative study was done by **Tertulino (2025)** on federated learning to pre- dict at-risk students, regarding the complexity of models and class imbalance [11],[12]. This research is practical in its contribution to the study of predictive performance within the actual educational setting. It emphasizes how performance in federated contexts is significantly impacted by the model's complexity and the data's spread. Nevertheless, the paper failed to introduce privacy-promoting methods like differential privacy or even compare hybrid models of proximal optimization and privacy-sensitive neural networks.

Fachola et al. (2023) investigated how federated learning is used in the broader context of educational analytics [27]. In their work, FL is introduced as a team-based and privacy-aware infrastructure that help to support massive cross-institutional intel- ligence. The authors note such areas of use as performance prediction, engagement analysis, and customized learning support and point out such issues related to this application as ethical concerns with collecting data centrally. But it is mostly concep- tual and architectural and not rigorous with experimental benchmarking of centralized, standard federated and privacy-enhanced federated configurations.

On top of training, other aspects of privacy are covered in the evaluation of mea- surements in federated systems. **Baykara et al. (2024)** suggested a privacy-saving solution to calculate Area Under the Curve (AUC) in the federated setting in Fully Homomorphic Encryption [23]. Even though it is not explicitly aimed at predicting stu- dent performance, the research shows that computation of the metrics where there is secure computation is possible in the framework of federation. Nevertheless, it fails to present an end-to-end predictive modeling framework, and trade-offs between privacy overhead and predictive accuracy in the educational setting.

The current empirical research also supports the applicability of federated learning in distributed educational settings. **Riyadi and Dewi (2026)** carried out a compar- ative analysis of various federated optimization algorithms such as FedAvg, FedProx,

FedDyn, q-FedAvg, and SCAFFOLD, to carry out privacy-preserving academic prediction with heterogeneous educational datasets [1]. Their results affirm that federated learning can retain high predictive performance without violating institutional data privacy. However, the research pays more attention to algorithmic comparison without a profound implementation of advanced privacy-improving methods like differential privacy or secure aggregation.

Within the healthcare field, **Tanveer et al. (2025)** used federated learning along- side Differentially Private Stochastic Gradient Descent to achieve the equilibrium between privacy protection and predictive accuracy [3]. Their model relatively attained a predictive accuracy of 93 percent and tight privacy budget (ϵ= 0.69), which showed the possibility of combining differential privacy with federated optimization. Never- theless, the study is field-specific and does not deal with non-homogenous educational data or non-IID situations at the institution level.

On the same note, **Kumari Singh (2025)** came up with a federated learning-based neural network architecture to forecast the dropout of students across dis- tributed organizations [38]. The FedAvg-based solution delivered the accuracy of about 92.8 and an F1-score of 0.914, which testifies that federated learning is an effective approach to educational prediction. Nevertheless, the research has minimal discussion of the non-IID data behavior and fails to use any extra privacy-preserving meth- ods besides the standard federated averaging. The studies considered in aggregate point to the use of FL as promising paradigms of privacy-aware educational analyt- ics [7],[32],[37]. The literature demonstrates how FL promotes collaborative predictive modeling and reduces the hazards associated with data aggregation at one location. But there are also a lot of limitations. The majority of research is conceptual, tiny, course-based, or lacks comprehensive comparisons between privacy-enhanced feder- ated models, conventional federated models, and centralized models. In particular, it is clear that IID and non-IID situations lack a systematic study based on institution- level or real-time student data. The motivation behind the current investigation is defined by these limitations. In order to forecast student performance using neural networks with heterogeneous data distributions, a comparison framework that takes into account FedAvg, FedProx, and a range of differentially private federated models, such as DP-FedProx/FedProx, is needed. By bridging this gap, it will be possible to better understand convergence stability, privacy-utility trade-offs, and scalability in the context of actual multi-institution educational settings.

ISSN: 2096-3246
Volume 58, Issue 01, 2026

3 RESEARCH METHODOLOGY

In order to test the proposed DP-Fedprox Framework, this study created a robust, decentralized FL environment that mimics a number of educational institutions [1],[12],[35]. By keeping a boundary between model utility and privacy guarantee, the methodological approach directly addresses the merging instability seen in non- IID educational data [11],[22],[39]. Figure 1 shows the theoretical architecture of the DP-FedProx approach. It displays how data is sourced, features are standardized, 8 separate (statistically different) sources of data siloed and then used to locally train on those data sources by three different networks (3-layer neural network) using DP with Opacus (Gradient Clipping & Noise Injection) and then aggregating to a erence of each data source using the proximal term (μ=0.1) for aggregating for 25 rounds before performing a final evaluation. The system operates under the configurations form toy scale to the large-scale evaluation of Fl framework form (k=2-8) institutional clients, (t=10-25) communicational rounds, and (E=10-30) local epochs per round. The sys- tem runs under the Adam optimizer introduces adaptive moment estimation, requiring recalibration of gradient sensitivity for loss minimization to maintain the compati- bility with strict (DP) differential privacy implementation and proximal regulation [7],[11],[24].

3.1 Data preprocessing & Federated Partitioning Strategy:

The proposed DP-FedProx framework was validated with the help of high-fidelity, real-time student performance data in the form of 10,954 individual student records in several collegiate institutions. The multi-institutional nature of the data source is essential, in that it is bound to provide statistical heterogeneity for the academic setting. This heterogeneity gives a natural ground to the analysis of the strength of proximal regularization in the presence of heterogeneous (non-IID) conditions. There are 11 multidimensional academics, behavioral and demographic attributes, that form the feature space, and the features are organized in the following way:

- **Academic History Vectors:** past GPA, 10th and 12th grade marks, presentation marks, etc.
- **Behavioral Indicators:** level of stress, teacher feedback, learning rate, attendance, learning hours, etc.

AES
ISSN: 2096-3246
Volume 58, Issue 01, 2026

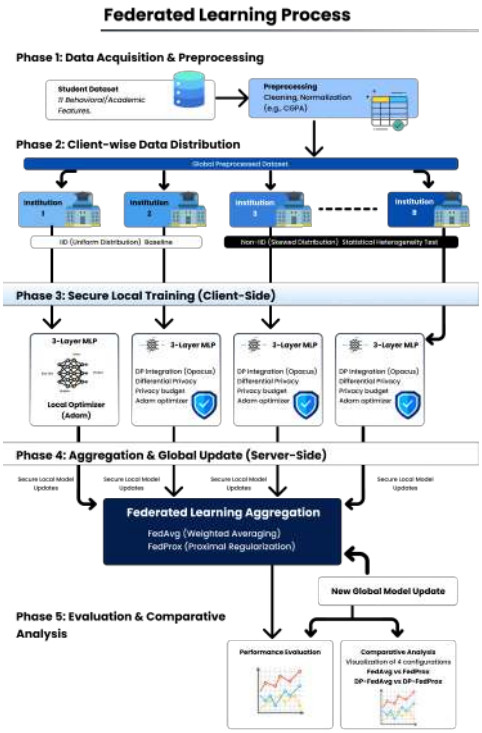


Fig. 1: Represents an architectural overview of the DP-FedProx framework for IID and non-IID

- **Demographic Factors:** distance between home and college, transportation means, etc.
- **Target Feature:** CGPA

Continuous variables were put on the min-max normalized scale, which maintains the relative distribution properties of the variables, but limits the magnitude of features, to achieve numerical stability in the neural optimization. A label encoding strategy was used to encode categorical variables, that easily integrated into the neural learning pipeline.

- **Federated Data Partitioning:** To model the real-life decentralized academic collaboration in a realistic way, the entire dataset was split horizontally to eight mutually exclusive subsets:

$D = \{D_1, D_2, D_3, \ldots, D_8\}$

The subsets constitute the stand-alone data storage of an individual educational establishment. There were two partitioning strategies:

- **IID Distribution:** The same sampling between institutions.
- **Non-IID Distribution:** Intentional skewing of quasi-stratification based on academic excellence and demographic imbalance.

The non-IID design launched manipulated heterogeneity in clients, encompassing unevenness in CGPA dispersion and patterns of behavioral characteristics. Such an arrangement establishes institutional drift, which is a direct challenge to federated aggregation. This kind of heterogeneity is necessary to test the resilience of FedProx proximal regularization term by checking if the global model can sustain convergence stability even when the local optimization paths are dissimilar. The multi-institutional and heterogeneous framework thus provides a strict and realistic validation platform to the proposed DP-FedProx framework.

3.2 Secure Client-side Training and DP Integration:

In this study, all client nodes in a multi-institutional decentralized framework contained a homogeneous MLP (Multi-Layer Perceptron) neural network type of archi- tecture that is defined by 3-layer computation architecture only [1],[14],[40]. To test the limits of decentralized convergence while adhering to privacy limitations, each node will utilize localization training procedures of ten to thirty (10-30) local epochs
(E) using the Adam optimizer (Adaptive Moment Estimation Optimizer) prior to any external communication. In order to prevent the adaptive updates from introducing more sensitivity, the Adam optimizer, an adaptive optimizer that calculates adaptive learning rates and automatically adjusts individual parameter learning rates for each epoch, must be adjusted for privacy limitations. The major contribution of this paper is the subsequent outline of the local optimizer loops. Utilizing the Opacus Differential Privacy Engine, the system enforces formal, mathematically quantifiable (ϵ, δ)-differential privacy (DP) by intervening at two points in the local training phase; thus, there will be no model updates made without first processing through the Opacus Differential Privacy Engine [7],[15],[2]:

1. **Gradient Constraints:** Differential Privacy Requires bounding the sensitivity of the optimization updates to any singular data points. For a local training batch B, and an individual sample $i \in B$, this computes the raw gradient g_i. Standard DP Federated Learning limits the constraints by L_2 Norm Clipping, ensuring that no

single vector of gradient exceeds a maximum boundary value C . This paper sets the clipping boundary value C to 1.0. The constrained gradient $\bar{g}_i$ is defined as:

$$\bar{g}_i = g_i \cdot \min\left(1, \frac{C}{\|g_i\|_2}\right)$$

Where $\|\cdot\|_2$ represents the L_2 norms. The operation explicit caps the sensitivity of the updates by restricting the maximum influence that individual students record can exert a precondition for formal privacy proofs.

2. **Stochastic Privacy:** The true differential privacy is achieved by adding the cal- ibrated stochastic noise to the constrained gradient. In this framework, the noise injection harmonized with the adaptive updates of the Admam optimizer. Before

updating the local model parameters to standardize, gaussian noise z is injected into the average tuned gradient.

$$\hat{m}_t = \frac{m_t}{1-\beta_1}, \hat{v}_t = \frac{v_t}{1-\beta_2} \text{ (adam bias correction)}$$

$$\theta_{t+1} = \theta_t - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t}+\epsilon} \text{ (adam parameter update)}$$

where η is the learning rate(0.001) and ϵ is the small constant to prevent division by zero

$$\bar{g} = \frac{1}{|B|}\left(\sum_{i\in B} \bar{g}_i + z\right)$$

where $z \sim N(0, \sigma^2 C^2 I)$.Here $N(0, \sigma^2 C^2 I)$ represents a multivariant Gaussian distribution with the mean zero and variance $(\sigma C)^2$ in each dimension, where(σ) is the noise multiple set to 1.0 and I is the identify matrix. This ensures that the local $\bar{g}$ satisfies the formal mathematical definition of DP, providing verifiable protection against data leakage attacks. The cumulative expenditure of a "privacy budget" over the vast 25 rounds of training was precisely recorded by means of R´enyi Differential Privacy (RDP) accounting, an advanced frame- work for DP that enables tighter and more efficient compositions of incompatible noise-generating mechanisms.

3. **Global Aggregation via DP-FedProx:** In total, the central server performed its aggregation when all 8 institutional nodes had received valid externalized pri- vate updates in a manner that's enabled more consistent learning trajectories, more local training periods (25 rounds of training with 30 total local epochs), and the combination's result on the degree of non-independence and unidentical (non-IID) distributional differences of training data and the additive privacy noise introduced by differential privacy (DP) through the move away from a base FedAvg algorithm to a more robust, stronger algorithm-FedProx to aggregate local models and obtain an overall, more accurate global model. Standard FL (FedAvg) simply takes the average of the differing local models (w_k)causing serious issues with non-IID. Fed- Prox adds Proximal Regularization Term directly to the optimization objective on the client side $h_k(w)$

ISSN: 2096-3246
Volume 58, Issue 01, 2026

thus providing stability because client k minimizes a modified cost that explicitly constrains their local cost updated.

$$h_k(w) = F_k(w) + \frac{\mu}{2} \| w - w^t \|^2$$

Where, $F_k(w)$ is the local loss function and w is the localized model weight being updated,w^t is the global weight received from the server at the begin of the communication rounds and μ is the proximal regularization coefficient. This quadratic $\frac{\mu}{2} \| w - w^t \|^2$ penalty has a stabilizing effect on the update, penalizing updates that are too far-off from the previous global model state. This proximal term is necessary to create a coherent convergence of the dissimilar adaptive updates generated

ISSN: 2096-3246
Volume 58, Issue 01, 2026

from the adam optimizer and to counteract local ”institutional drift,” enabling global convergence over 25 communications rounds. The server then combines the updated versions of the models by computing a weighted average of all the updates:

$$w(t+1) = \sum_{k=1}^{K} \frac{n_k}{n} w_k(t+1)$$

Where client k have number of the samples n_k, and n is the total samples

4 EXPERIMENTAL SETUP & ANALYSIS

4.1 Technical Implementation and Communication Protocol

The DP-FedProx framework was built using a modular technical stack composed of PyTorch to build the neural networks, Flower (flwr) to aggregate federated metrics, and Opacus to apply differential privacy. To reach the required frequency of synchro- nizing to each institutional silo through 25 rounds of communication, we used gRPC as the base technology to shape a remote procedure call layer. gRPC was chosen for its ability of serializing highly complex tensor data into a lightweight binary format (Protocol Buffers). This capability dramatically reduced the total amount of overhead involved in transmitting weight updates consisting of 15 features over the communi- cation channels connecting 8 institutional silos. With a low-latency communication backbone between the eight institutional silos, the Adam optimizer was able to main- tain global convergence with respect to all eight silos even after processing 30 local epochs.

4.2 System Configuration and Hyperparameter Formalization

The system architecture and optimization hyperparameters that were used to build the final simulation used for the high-complexity test (Set 3) have been formalized and are in Table 1. The configuration of this set of parameters provides the optimal trade- off between formal privacy guarantees and general predictive utility for the remaining 10,954 student data records.

Table 1: Formal System Configuration and Hyperparameters

Category	Parameter	Specification and Value
Architecture	Model Architecture	3-Layer MLP ($d = 11$)
Architecture	Aggregation Strategy	FedProx ($\mu = 0.1$), FedAvg
Architecture	Communication Protocol	gRPC (Flower)
Federated Parameters	Total Clients (K)	8 Institutional Silos
Federated Parameters	Communication Rounds (T)	25 Global Rounds
Federated Parameters	Local Epochs (E)	30 per Round
Differential Privacy	DP Framework	Opacus (PyTorch)
Differential Privacy	Clipping Norm (C)	1.0 (L_2 norm)
Differential Privacy	Noise Multiplier (σ)	1
Local Optimization	Optimizer	Adam
Local Optimization	Batch Size (B)	32
Local Optimization	Learning Rate (η)	0.001

4.3 Multi-Phased Stress Testing

It is designed to validate the DP-FedProx framework and was implemented in three experiments that were progressively more complex to demonstrate the skill of the model to scale from a base test to a highly varied test on a real-life representative network.

Table 2: Table 2: Parameterization of Experimental Sets

METRIC	SET 1 (BASE)	SET 2 (STRESS TEST)	SET 3 (FINAL)
Client Count (K)	2 Clients	5 Clients	8 Clients
Local Epochs (E)	15	15	30
Communication Rounds (T)	10	15	25

ISSN: 2096-3246
Volume 58, Issue 01, 2026

5 RESULTS ANALYSIS & DISCUSSION

To evaluate the DP-FedProx framework's efficiency, the results from three experimental phases, each more complex than the previous, were combined. The objective of these experiments was to see how predictive utility and privacy costs measured when going from a baseline environment to a highly diverse institutional environment within an eight-college institution. The Change between the sets illustrates how stable the model is at different levels of synchronization in terms of load and local processing capability.

- **Set 1 (Base):** With only 2 clients & 10 rounds, as displayed in Figure 2 the model's accuracy is ~ 81.8% at baseline. Due to low variation in client characteristics, early convergence was achieved with an overall error rate of ~ 3.2 & privacy expenditure of ϵ =~ 3.8.

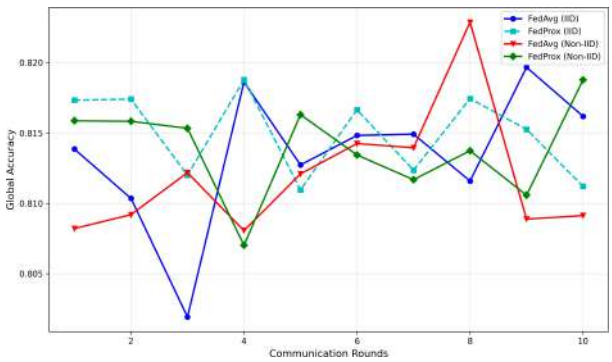


Fig. 2: Comparison between FedProx & FedAvg(IID & non-IID) over 2 clients and 10 communication rounds with 15 local epochs

- **Set 2 (Stress Test):** As we changed to a network of 5 clients & 15 rounds from Set 1 to Set 2 as shown in Figure 3. This resulted in the introduction of substantial drift due to the inability of the Adam Optimizer (used to calculate the changes made to the model) to properly reason for the changes in amount of clients and their different set of

characteristics. In calculating the overall drift rate, the MSE increased to ~ 8.0; however, subsequent rounds will experience considerable fluctuations meanwhile there are now numerous differing characteristics present that the procedure has not been able to adjust for in past iterations.

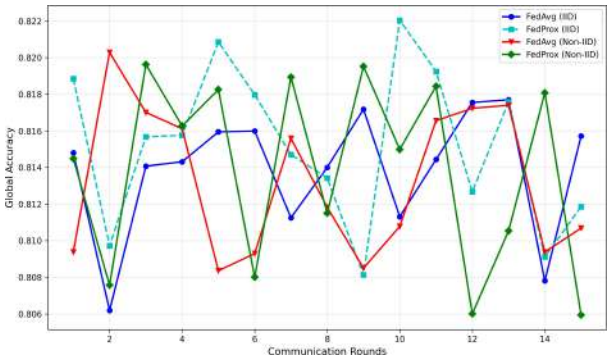
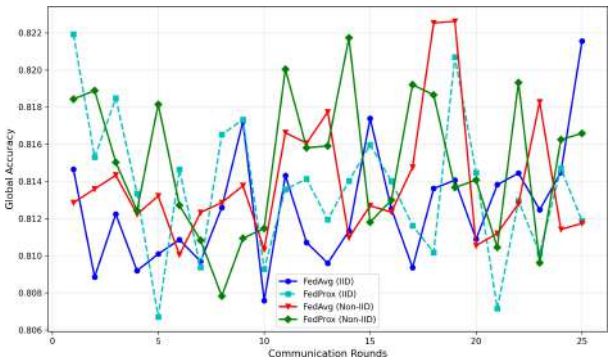


Fig. 3: Comparison between FedProx & FedAvg(IID & non-IID) over 5 clients and 15 communication rounds with 15 local epochs

- **Set 3 (Final Scalability):** Using the final set of K=8, T=25, E=30 produced the true level of resiliency for each of the configurations being tested to that point as shown in Figure 4. Notwithstanding the higher overall level of complexity compared to all previous configuration (due to the use of a Non-IID data distribution), all configurations reached their final levels at the finale of the 10th round. Global accuracy at that time was ~ 82%, corresponding to an MSE of ~ 5.5 and a total



privacy level of ϵ =~ 6.00.

ISSN: 2096-3246
Volume 58, Issue 01, 2026

Fig. 4: Comparison between FedProx & FedAvg(IID & non-IID) over 8 clients and 25 communication rounds with 30 local epochs

5.1 Convergence & MSE Stability Analysis

This section evaluates the DP-FedProx framework's convergence profile under three different experimental conditions: (1) to show how resilient its structure is to ran- dom noise. (2) to assess how well it functions in the face of "client drift" and client unpredictability. (3) To investigate how these factors interact to impact the frame- work's overall performance. Three different experimental conditions were used for the experiments:

(i) **Set 1 (Low Complexity Baseline):** The analogue parameters of (K=2, T=10, E=15), were used in this configuration; Rapid and monotonically decreasing MSE was displayed throughout completed rounds of this experiment, ending overall MSE at approximately 3.2. The lack of diversity within this experimental configuration meant that Adam was able to readily identify a worldwide minimum during sep- arately repetition of the training process, but with little interference due to the involvement of proximal term;

(ii) **Set 2 (Intermediate Stress Test):** MSE within this configuration experienced a rapid decline initially, followed by a significant increase until the end of round 10 where MSE was recorded at approximately 8.0 The instability demonstrated within this configuration indicates typical behavior associated with Federated Learning

experiencing increased Non-IID data skewness combined with Differential Privacy (DP) introduced into the overall structure;

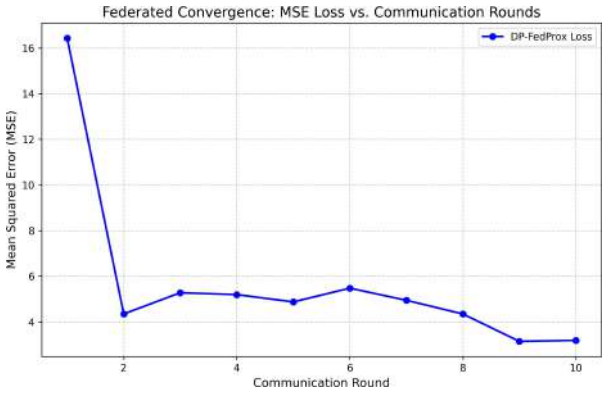


Fig. 5: MSE Loss vs. Communication Rounds (Set 1)

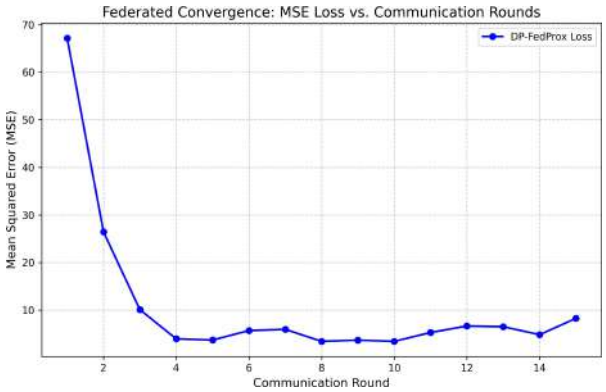


Fig. 6: MSE Loss vs. Communication Rounds (Set 2)

ISSN: 2096-3246
Volume 58, Issue 01, 2026

(iii) **Set 3 (Final Scalability Model):** Even though MSE started at a higher overall number due to a significant amount of noise introduced into the training process in the configuration of K=8, T=25 and E=30, the performance of the overall system was still significantly more constant than during either of the previous two config- urations. After the completion of round 10, and due to the increasing amount of noise introduced into the training process (by Opacus) the amount of divergence between local updates began to significantly converge to global consensus by way of Michael's (FedProx) proximal term, where the penalty being applied to the local updates was that of $\mu = 0.1$ resulting in an overall very low MSE of approximately 5.5 Theore, it can be deduced that DP-FedProx framework effectively "tamed" the stochastic noise introduced into the framework by way of using based model.

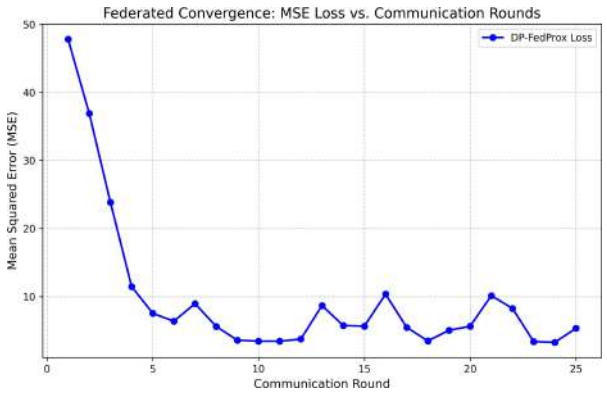


Fig. 7: MSE Loss vs. Communication Rounds (Set 3)

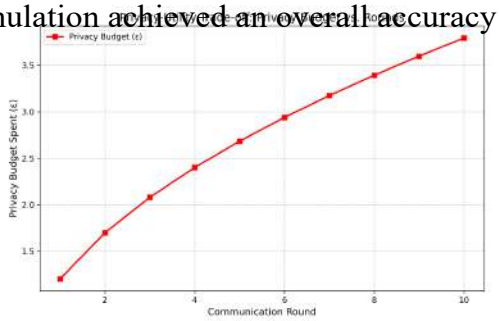
5.2 Accuracy Performance & Privacy Expenditure

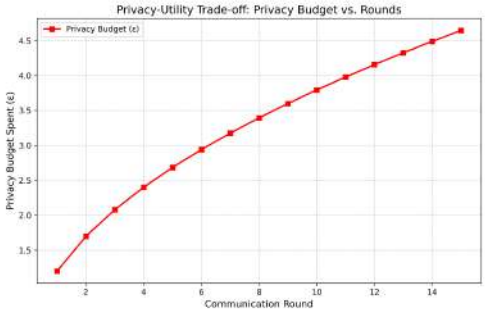
- **Predictive Utility and Privacy Budget**: To confirm that the model remains viable for institutional deployment in a real-world setting, the relationship between predictive utility and privacy budget (ϵ) was formalized. Predictive Utility The global accuracy was maintained at exceptionally high levels (conservatively) across all experimental mentioned data (i.e. Set 1 had roughly an accuracy of ~ 81.8% &

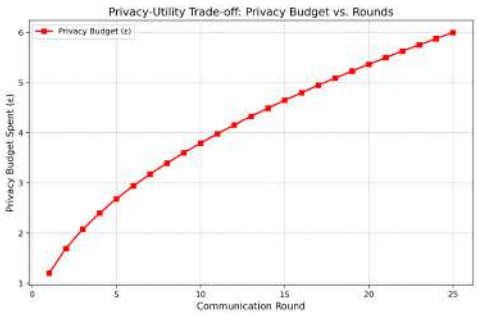
Set 3 had an accuracy of ~ 82%.The marginal difference in predictive utility between the two sets (0.1%) was achieved despite having a 4X higher total number of clients and a 2X difference in number of local epochs; illustrating excellent scalability for the model. Behavioral patterns of entirely 10,954 students were known from their records even under conditions of a noisy gradient.

- **Privacy Accounting Using RDP:** We used R´enyi Differential Privacy (RDP) to quantify the total cumulative privacy loss (for all clients) after their cumulative RDP was applied to each of their individual data sets. For the final evaluation of Set 1, the cumulative RDP was approximately $\epsilon \approx 3.8$, which indicates a high level of privacy. For Set 3, the ϵ value for this set was approximately $\epsilon \approx 6.0$ with fix $\delta = 10^{-5}$.

For practical applications of Differential Privacy, an ϵ below 10 indicates that data produced by an individual cannot be identified based on the overall output of that individual's actions. As such, these levels of privacy provide confidence that the gRPC enabled communication loop will allow for the completion of 25 rounds of training with minimal privacy expenditure that remains well within acceptable bounds of academia and law. The final Set 3 simulation achieved an overall accuracy (i.e., global) of ~ 82%.







ISSN: 2096-3246
Volume 58, Issue 01, 2026

Fig. 8: Privacy Budget (ϵ) vs. Rounds for Set1, Set2, and Set3

6 DISCUSSION

The outcomes of three distinct experiments phases provide evidence that using a proximal-stabilized federated model will maintain a high level of predictive utility while satisfying the strict requirements of differential privacy (DP).

6.1 Challenges Associated with Client Drift

Federated Learning faces a major issue erred to as ”Client Drift”, which occurs when local models differ from other significantly based on non-IID distributions of data. The high number of local epochs (E) typically makes this situation worse. For instance, Using E=30 represents an extremely high level of computation that would result in global instability under normal situations. Yet in our experiment's third set of data, we found that using a proximal weighting of ($\mu = 0.1$), we could keep all local updates from straying too far beyond the consensus of the global model. The addition of the penalty term $\frac{\mu}{} ||w_2 - w^t||^2$ prevented the Adam optimizer from over fitting to the unique statistical noise associated with each individual college silo; thereby providing visual evidence of the convergence of MSE recovery after round 10, however there was a much larger synchronization burden associated with 8 clients.

6.2 Adam is Synergistic with Opacus' Differential Privacy Mechanism when Adding Noise

The Adam optimizer's adaptive learning rate ($\eta = 0.001$) performs well in driving convergence upon complex behavior data while being sensitive to ”noise” created via Gaussian injection of noise ($\sigma = 1.0$).These analyses reveal that the 3-layer MLP architecture provides adequate capacity to generalize student performance patterns despite the added noise. Additionally, differentiating from the results in **Tanveer et al. (2025)**, who evaluated their DP method upon more homogeneous than our sample, our methodology achieved approximately **82** percent accuracy levels when graduation clips were (i.e. C=1.0) and perturbed. This evidence supports that the combination of per sample grad graduate clipping and Adam's second moment enables a strong protection from the utility decay generally found with respect to differential privacy.

6.3 Comparative Analysis and Research Gap Fulfillment

The proposed framework for privacy-preserving educational technologies addresses a significant gap in existing literature

- **Utility vs. Privacy: Tanveer Et al. (2025)** reached a 93.0% accuracy through using traditional DP but their centralized-leaning approach is not generalizable to real-world scenarios; there is not sufficient representation of non-IID data within multiple colleges (8) which will be present in this 8-college simulation with respect to their Central Limit Theorem assumptions. Which makes their accuracy results void [2],[11],[39].
- **Heterogeneity vs. Privacy: Tertulino (2025)** recognized the need for FedProx in dealing with data skewness (recording an accuracy of 61.2%) but made no mention of how to preserve student identity or academic records [13],[18],[24]. Essentially, we use our formal privacy budget $\epsilon = 6.0$ coupled with maximum accuracy $\sim$ 82%, 20.5% greater than **Tertulino's** report, as definitive evidence that institutions can provide equal protection to their student's identity without sacrificing predictive capabilities through FedProx [7],[14],[24].

6.4 Scalability and Communication Efficiency over gRPC

The transition was made from Set 1 (2 clients) to Set 3 (8 clients) in such a way that it could be resolute whether or not gRPC is an effective means of providing communication between clients; this was confirmed in the number of communication rounds reaching T=25, as the latency remained significant. Theore, the DP-FedProx Framework has the capability of being widely deployed to other educational consortia, or larger in size than previously, as long as data remain serialized by using an efficient binary format through the Flower's gRPC integration with respect to each model update.

7 CONCLUSION & FUTURE STUDY

The proposed DP-FedProx model successfully bridges the gap between high-utility academic analytics and stringent data privacy by maintaining a consistent accuracy rate of approximately $\sim$ 82% across a distributed collection of eight institutions and 10,954 student records. The three-phase experimental progression demonstrates that the integration of the Adam adaptive optimizer with a proximal regularization term ($\mu = 0.1$) effectively mitigates the "client drift" caused by lengthy local training (E=30) and lects the control of non-IID data skew. Using a gRPC-enabled commu- nication backbone and Opacus for differential privacy, this model can provide formal

privacy guarantees ($\epsilon = 6.0$) without exhibiting the utility collapse experienced in many previously published models. Compared to the literature findings from 2025, it is obvious that utmost recent federated learning models, such as **Tanveer et al. and Tertulino**, have made use of either privacy or heterogeneity as their main prior- ities; however, the DP-FedProx bandwagon is the first to provide dual considerations of varied levels of privacy and heterogeneity, thus providing a comprehensive and scalable framework to be utilized by multiple institutions for educational research ini- tiatives. Future research will concentrate on providing dynamic proximal adaptations and developing countermeasures to protect against adversarial model poisoning to further strengthen the framework for global deployment.

REFERENCES

[1] T. Zhang et al., "Enhancing Dropout Prediction in Distributed Educational Data Using Learning Pattern Awareness: A Federated Learning Approach," *Mathematics*, vol. 11, no. 24, Dec. 2023, doi: 10.3390/math11244977.

[2] F. Tanveer et al., "Balancing privacy and performance in healthcare: A feder- ated learning framework for sensitive data," *Digit Health*, vol. 11, Jan. 2025, doi: 10.1177/20552076251381769.

[3] H. Yang, J. Li, M. Hao, W. Zhang, H. He, and A. K. Sangaiah, "An effi- cient personalized federated learning approach in heterogeneous environments: a reinforcement learning perspective," *Sci Rep*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-80048-3.

[4] X. Liu, X. Dong, N. Jia, and W. Zhao, "Federated Learning-Oriented Edge Computing Framework for the IIoT," *Sensors*, vol. 24, no. 13, Jul. 2024, doi: 10.3390/s24134182.

[5] B. Yin, H. Zhang, J. Lin, F. Kong, and L. Yu, "PVFL: Verifiable federated learn- ing and prediction with privacy-preserving," *Comput Secur*, vol. 139, Apr. 2024, doi: 10.1016/j.cose.2024.103700.

[6] M. Mehta, M. v. Bimrose, D. J. McGregor, W. P. King, and C. Shao, "Federated learning enables privacy-preserving and data-efficient dimension prediction and part qualification across additive manufacturing factories," *J Manuf Syst*, vol. 74, pp. 752–761, Jun. 2024, doi: 10.1016/j.jmsy.2024.04.031.

[7] A. el Ouadrhiri and A. Abdelhadi, "Differential Privacy for Deep and Feder- ated Learning: A Survey," *IEEE Access*, vol. 10, pp. 22359–22380, 2022, doi: 10.1109/ACCESS.2022.3151670.

[8] Y. Lu, Z. Yu, and N. Suri, "Privacy-preserving Decentralized Federated Learning over Time-varying Communication Graph," *ACM Transactions on Privacy and Security*, vol. 26, no. 3, Jun. 2023, doi: 10.1145/3591354.

[9] C. He et al., "FedGraphNN: A Federated Learning System and Benchmark for Graph Neural Networks," Sep. 2021, [Online]. Available: http://arxiv.org/abs/ 2104.07145.

[10] Y. Zhang, Y. Li, Y. Wang, S. Wei, Y. Xu, and X. Shang, "Federated learning-outcome prediction with multi-layer privacy protection," Dec. 2023, doi: 10.1007/s11704-023-2791-8.

[11] S. Mohammadi, A. Balador, S. Sinaei, and F. Flammini, "Balancing pri- vacy and performance in federated learning: A systematic literature review on methods and metrics," *J Parallel Distrib Comput*, vol. 192, Oct. 2024, doi: 10.1016/j.jpdc.2024.104918.

[12] C. Fachola, A. Tornar´ıa, P. Bermolen, G. Capdehourat, L. Etcheverry, and M. I. Fariello, "Federated Learning for Data Analytics in Education," *Data (Basel)*, vol. 8, no. 2, Feb. 2023, doi: 10.3390/data8020043.

[13] R. Tertulino, "Centralized vs. Federated Learning for Educational Data Mining: A Comparative Study on Student Performance Prediction with SAEB Microdata," Aug. 2025, [Online]. Available: http://arxiv.org/abs/2509.00086.

[14] S. Chen and X. Qi, "Entropy-adaptive differential privacy federated learning for student performance prediction and privacy protection: a case study in Python programming," *Front Artif Intell*, vol. 8, 2025, doi: 10.3389/frai.2025.1653437.

[15] K. Wei et al., "User-Level Privacy-Preserving Federated Learning: Analysis and Performance Optimization," Jan. 2021, [Online]. Available: http://arxiv.org/abs/ 2003.00229.

[16] D. Monteiro, I. Mavinkurve, P. Kambli, and Prof. S. Surve, "Federated Learn- ing for Privacy-Preserving Machine Learning: Decentralized Model Training with Enhanced Data Security," *Int J Res Appl Sci Eng Technol*, vol. 12, no. 11, pp. 355–

361, Nov. 2024, doi: 10.22214/ijraset.2024.65062.

[17] J. Liu et al., "Distributed and Deep Vertical Federated Learning with Big Data ARTICLE TYPE Distributed and Deep Vertical Federated Learning with Big Data", doi: 10.1002/cpe.7697.

[18] R. Tertulino1 and R. Almeida1, "EVALUATING FEDERATED LEARNING FOR AT-RISK STUDENT PREDICTION: A COMPARATIVE ANALYSIS OF MODEL COMPLEXITY AND DATA BALANCING."

[19] S. Yoneda, V. Švábenský, G. Li, D. Deguchi, and A. Shimada, "Ranking-Based At-Risk Student Prediction Using Federated Learning and Differential Features," in *Proc. Int. Conf. Educational Data Mining*, Int. Educational Data Mining Soc., 2025, pp. 289–302, doi: 10.5281/zenodo.15870193.

[20] M. T. Hosain, M. S. Sajid, S. Akter, A. Zaman, and S. Sakeeb Khan, "Pri- vacy Preserving Machine Learning Model Personalization through Federated Personalized Learning," [Online]. Available: https://www.kaggle.com/datasets/datamunge/virusmnist.

[21] Y. Goto, T. Matsumoto, H. Rizk, N. Yanai, and H. Yamaguchi, "Privacy-Preserving Taxi-Demand Prediction Using Federated Learning," May 2023, [Online]. Available: http://arxiv.org/abs/2305.08107.

[22] N. Thakre, N. Pateriya, G. Anjum, D. Tiwari, and A. Mishra, "Federated Learn- ing Trade-Offs: A Systematic Review of Privacy Protection and Performance Optimization," *Int. J. Innov. Res. Comput. Commun. Eng.*, vol. 11, no. 10, pp. 11332–11343, Nov. 2023, doi: 10.15680/ijircce.2023.1110049.

[23] C. A. Baykara, A. B. Ü nal, and M. Akgün, "Enabling Privacy-preserving Model Evaluation in Federated Learning via Fully Homomorphic Encryption," Aug. 2025, [Online]. Available: http://arxiv.org/abs/2403.14428.

[24] C. Mathew and P. Asha, "FedProx: FedSplit Algorithm based Federated Learn- ing for Statistical and System Heterogeneity in Medical Data Communication," *J. Internet Serv. Inf. Secur.*, vol. 14, no. 3, pp. 353–370, Aug. 2024, doi: 10.58346/JISIS.2024.I3.021.

[25] T. A. Elsnousy, M. Ragaie Sayed, and N. E. Elghitany, "A PROPOSED FEDERATED LEARNING (FL) APPROACH TO ENHANCE PREDICTION OF STUDENT PERFORMANCE EVALUATION SYSTEMS," *J. Theor. Appl. Inf. Technol*, vol. 15, no. 21, 2025, [Online]. Available: www.jatit.org.

[26] K. A. S. Kumar, L. Nelson, and B. R. Jibinsingh, "Systematic review of privacy-preserving Federated Learning in decentralized healthcare systems," Dec. 01, 2025, Elsevier B.V., doi: 10.1016/j.fraope.2025.100440.

[27] A. Gaber, H. Abdeltwab, and T. Elbatt, "FedCVD: Towards a Scalable, Privacy-Preserving Federated Learning Model for Cardiovascular Diseases Pre- diction," in *ACM Int. Conf. Proceeding Series*, ACM, Jan. 2024, pp. 7–11, doi: 10.1145/3647750.3647752.

[28] D. R. Kale, T. S. Mane, A. Buchade, P. B. Patel, L. K. Wadhwa, and R. G. Pawar, "Federated Learning for Privacy-Preserving Data Mining," in 2024 *Int. Conf. Intell. Syst. Adv. Appl.*, IEEE, 2024, doi: 10.1109/ICISAA62385.2024.10828741.

[29] J. Wang and Y. Yu, "Machine learning approach to student performance prediction of online learning," *PLoS One*, vol. 20, no. 1, Jan. 2025, doi: 10.1371/journal.pone.0299018.

[30] C. Boscher, N. Benarba, F. Elhattab, and S. Bouchenak, "Personalized Privacy-Preserving Federated Learning," in *Middleware 2024 - Proc. 25th ACM Int. Middleware Conf.*, ACM, Dec. 2024, pp. 454–466, doi: 10.1145/3652892.3700785.

[31] B. Casell et al., "A Performance Analysis for Confidential Federated Learning," 2024 *IEEE Security and Privacy Workshops (SPW)*, San Francisco, CA, USA, 2024, pp. 40–47, doi: 10.1109/SPW63631.2024.00009.

[32] I. Manga, "Federated Learning at Scale: A Privacy-Preserving Framework for Decentralized AI Training," 2025 *5th Int. Conf. Soft Comput. Secu- rity Appl.*

(ICSCSA), Salem, India, 2025, pp. 110–115, doi: 10.1109/IC-SCSA66339.2025.11170780.

[33] A. Ahmed, M. A. Javed, J. N. Qureshi, H. Khan, and H. F. Yousaf, "An insightful Machine Learning based Privacy-Preserving Technique for Federated Learning," *Asian Bull. Big Data Manag.*, vol. 4, no. 4, pp. 332–343, Dec. 2024, doi: 10.62019/abbdm.v4i4.277.

[34] K. Kulshreshtha, B. Goel, H. S. Kanyal and V. Singhal, "Federated Learning in the Era of Privacy Preservation: A Comprehensive Review and Future Direc- tions," 2024 *2nd Int. Conf. Disruptive Technol. (ICDT)*, Greater Noida, India, 2024, pp. 763–768, doi: 10.1109/ICDT61202.2024.10489417.

[35] S. Vij and K. S. Suneetha, "Federated Learning in Education: Enhancing Student Privacy in AI-Based Feedback Mechanisms," *RADemics Res. Inst.*, vol. 2025, no. 2025, pp. 34, 2025.

[36] H. Gajjar, N. Divecha, K. Sarva Vishwavidyalaya, and A. Professor, "(on-line version) ENHANCING BIG DATA PRIVACY AND PERFORMANCE THROUGH EDGE-INTEGRATED FEDERATED LEARNING: A COMPARA-
TIVE STUDY," *Int J Appl Math (Sofia)*, vol. 38, no. 2s, 2025.

[37] R. Rahman, "Federated Learning: A Survey on Privacy-Preserving Collaborative Intelligence," Mar. 2026, [Online]. Available: http://arxiv.org/abs/2504.17703.

[38] "A Federated Learning Neural Network For Student Dropout Prediction," *ICCE*, Dec. 2025, Accessed: Apr. 01, 2026, [Online]. Available: https://library.apsce.net/index.php/ICCE/article/view/5927.

[39] J. Sung Lee and H. Chung, "Multi-level Analyzation of Imbalance to Resolve Non-IID-Ness in Federated Learning," [Online]. Available: https://ssrn.com/abstract=4887224.

[40] M. Angello Qadosy Riyadi and A. Mariasti Dewi, "A Comparative Evaluation of Federated Learning Algorithms for Privacy-Preserving Academic Prediction on Heterogeneous Data", *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 10, no. 1, pp. 41–52, Feb. 2026, doi: https://doi.org/10.29207/resti.v10i1.7288.