



## A Comparative Evaluation of Federated Learning Algorithms for Privacy-Preserving Academic Prediction on Heterogeneous Data

Michael Angello Qadosy Riyadi<sup>1\*</sup>, Adinda Mariasti Dewi<sup>2</sup>

<sup>1,2</sup>Department of Information Technology (Surabaya City Campus), Telkom University, Surabaya, Indonesia

<sup>1\*</sup>[michaelangelo@student.telkomuniversity.ac.id](mailto:michaelangelo@student.telkomuniversity.ac.id), <sup>2</sup>[adindamariasti@student.telkomuniversity.ac.id](mailto:adindamariasti@student.telkomuniversity.ac.id)

### Abstract

*The rapid growth of educational data enables predictive analytics for academic performance, yet privacy regulations like GDPR and FERPA severely restrict centralized data sharing. Although Federated Learning (FL) has succeeded in privacy-sensitive fields such as healthcare, its application in education remains underexplored, lacking systematic comparative studies of multiple FL algorithms across diverse educational datasets—especially emphasizing recall and ROC-AUC as critical metrics for early identification of students at academic risk. This study fills this gap by evaluating five FL algorithms—Federated Averaging (FedAvg), Federated Proximal (FedProx), Federated Dynamics (FedDyn), Fair Federated Averaging (q-FedAvg), and SCAFFOLD—for privacy-preserving prediction of academic outcomes. Three public datasets were purposefully selected for their representativeness and heterogeneity: Predict Students Dropout and Academic Success (binary dropout prediction with socioeconomic factors), Student Performance (multi-class grade prediction in secondary education), and xAPI-Edu-Data (multi-class performance based on online learning activities). Local neural networks employed Stratified 5-Fold Cross-Validation, while FL algorithms ran for 50 communication rounds. Global models, particularly q-FedAvg and FedProx, consistently surpassed local models, with q-FedAvg achieving 0.7668 accuracy, 0.6813 recall, and 0.8810 ROC-AUC on Predict Students Dropout; 0.8580 accuracy and 0.9871 recall on Student Performance; and 0.7396 accuracy and 0.8815 ROC-AUC on xAPI-Edu-Data. Paired T-tests confirmed significant recall gains for most global models ( $p < 0.05$ ). These results highlight FL's ability to handle data heterogeneity and privacy constraints while improving predictive performance, thereby supporting timely educational interventions and enhanced student retention policies.*

**Keywords:** academic; educational; federated learning; machine learning; privacy

*How to Cite:* 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.  
*Permalink/DOI:* <https://doi.org/10.29207/resti.v10i1.7288>

*Received:* October 22, 2025

*Accepted:* February 3, 2026

*Available Online:* February 16, 2026

*This is an open-access article under the [CC BY 4.0 License](#)*

*Published by [Ikatan Ahli Informatika Indonesia](#)*

### 1. Introduction

The surge in educational data, including grades, attendance, and digital interactions, has transformed predictive analytics, facilitating tailored learning and dropout prevention [1-4]. However, stringent privacy laws like GDPR and FERPA restrict centralized data collection, posing significant challenges [5-7]. Conventional machine learning (ML) methods, which depend on centralized data, raise privacy concerns and struggle with non-IID data stemming from diverse demographics, curricula, and evaluation systems [8], [9-12]. Balancing predictive accuracy with privacy protection remains a critical barrier to ML adoption in education. Federated Learning (FL) offers a promising approach by enabling distributed model training without sharing sensitive data, preserving privacy while maintaining performance [13-15]. This study is crucial

for developing secure, efficient, and privacy-compliant predictive systems in distributed educational settings.

Although FL has been widely adopted in healthcare for sensitive data, its application in education is still in its infancy [16-18]. Previous educational studies primarily used traditional ML techniques like random forests and logistic regression, achieving high accuracy but compromising privacy due to centralized data needs [19-22]. Limited FL research with Federated Averaging (FedAvg) shows promise, but advanced algorithms like Federated Proximal (FedProx), Federated Dynamics (FedDyn), and Fair Federated Averaging (q-FedAvg) are underexplored in education [23], [24]. Research gaps include a focus on binary classification, limited evaluation of recall and ROC-AUC for class imbalance, and insufficient analysis of data heterogeneity and dataset size effects on FL performance [25], [26]. These

gaps highlight the need for FL to address privacy and heterogeneity in education. Several studies have explored ML for academic predictions but highlight privacy limitations. J. Wang *et al.* [27] used logistic regression with Taylor expansion on a MOOC dataset, achieving 0.9330 accuracy, yet relied on centralized data. Similarly, W. Zou *et al.* [28] applied multi-model fusion (Random Forest, SVM, GradientBoosting, KNN, AdaBoost) for Computer Science majors' performance prediction, with GradientBoosting yielding 0.5100 accuracy, but lacked privacy measures. S. N. Thakur *et al.* [29] achieved perfect accuracy (1.00) using logistic regression with recursive feature elimination on the University of Minho dataset, ignoring privacy concerns. G. L. E. M. Niño *et al.* [30] classified high school math performance with KNN (0.9100 accuracy) on a dataset of 711 students, again without privacy considerations. G. Al-Tameemi *et al.* [31] combined PCA and K-Means with supervised methods (FDN, RF, DT) on the OULAD dataset, achieving 0.9370 accuracy with decision trees, but used centralized data. Lastly, E. F. Agyemang *et al.* [32] reviewed ML algorithms (RF, KNN, SVM, DT, Logistic Regression) on the xAPI-Edu-Data dataset, achieving 0.9243 accuracy with Random Forest, yet failed to address privacy. These studies demonstrate high predictive accuracy but are unsuitable for sensitive educational contexts due to centralized data requirements, reinforcing the need for FL-based approaches.

This study evaluates five FL algorithms—FedAvg, FedProx, FedDyn, q-FedAvg, and SCAFFOLD—for predicting academic outcomes on heterogeneous datasets, comparing global and local models using accuracy, precision, recall, F1-score, and ROC-AUC. By leveraging FL to address privacy concerns, this research provides a secure approach for non-IID data. Key contributions include a thorough evaluation of five FL algorithms, including the understudied FedDyn and q-FedAvg, analysis of data heterogeneity and dataset size impacts, and practical insights for secure academic prediction systems, advancing FL in education.

The primary objective of this study is to systematically evaluate and compare the performance of five advanced Federated Learning algorithms—Federated Averaging (FedAvg), Federated Proximal (FedProx), Federated Dynamics (FedDyn), Fair Federated Averaging (q-FedAvg), and SCAFFOLD—in predicting academic outcomes across heterogeneous educational datasets while strictly preserving data privacy. Specifically, this research aims to assess how these algorithms handle non-IID data distributions and class imbalances common in educational contexts, with particular emphasis on recall and ROC-AUC as key metrics for identifying students at risk. By benchmarking global FL models against local centralized models, the study seeks to demonstrate the effectiveness of FL in overcoming privacy barriers imposed by regulations such as GDPR and FERPA, thereby enabling secure collaborative analytics across distributed educational institutions.

## 2. Methods

The research methodology, depicted in Figure 1, starts with the use of three educational datasets: Predict Students Dropout and Academic Success, Student Performance, and xAPI-Edu-Data. In the preprocessing phase, categorical variables were transformed into numerical formats using dummy variable encoding. The datasets were split into three subsets to emulate a federated learning scenario with three clients. Each client trained a neural network model employing a Stratified 5-Fold Cross-Validation approach, applied independently to their respective data. Data normalization was conducted using StandardScaler, fitted solely on the training data to avoid data leakage, and then applied to the test data. The local models from each client were combined using five federated learning algorithms—Federated Averaging (FedAvg), Federated Proximal (FedProx), Federated Dynamics (FedDyn), Fair Federated Averaging (q-FedAvg), and SCAFFOLD—to create distinct global models for each algorithm.

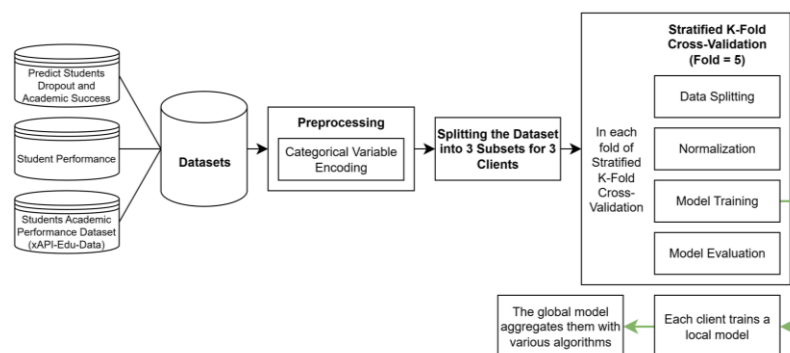


Figure 1. Research Flow

### 2.1 Dataset

This research employs three widely recognized public educational datasets for classification tasks: Predict Students Dropout and Academic Success, Student

Performance, and xAPI-Edu-Data. Table 1 summarizes the characteristics of these datasets. Their selection was driven by their frequent use in existing literature, enabling effective comparisons with previous studies and enhancing the contribution of this work to

advancing federated learning applications in education [20], [32].

Table 1. Characteristics of the Datasets Used in This Study

| Dataset                                               | Instances | Features | Number of Classes | Class Proportion (%)                                    |
|-------------------------------------------------------|-----------|----------|-------------------|---------------------------------------------------------|
| Predict Students Dropout and Academic Success         | 4,424     | 36       | 3                 | Class 0: 32.12%,<br>Class 1: 17.95%,<br>Class 2: 49.93% |
| Student Performance                                   | 649       | 30       | 2                 | Class 0: 15.41%,<br>Class 1: 84.59%                     |
| Students Academic Performance Dataset (xAPI-Edu-Data) | 480       | 16       | 3                 | Class 0: 26.46%,<br>Class 1: 43.96%,<br>Class 2: 29.58% |

To enable the federated learning simulation, each dataset was split into three subsets, each representing a client [33], [34]. The partitioning was achieved by dividing the total number of instances ( $N$ ) into three parts, as described by Equation 1.

$$\text{Client Size} = \left\lfloor \frac{N}{C} \right\rfloor \quad (1)$$

$N$  represents the total instances,  $C$  is the number of clients ( $C = 3$ ), and  $\lfloor \cdot \rfloor$  indicates the floor function (rounding down). The data was distributed sequentially: Client 1 was assigned instances from index 0 to  $\left(\left\lfloor \frac{N}{3} \right\rfloor - 1\right)$ , Client 2 from  $\left(\left\lfloor \frac{N}{3} \right\rfloor\right)$  to  $\left(2 \left\lfloor \frac{N}{3} \right\rfloor - 1\right)$ , and Client 3 from  $\left(2 \left\lfloor \frac{N}{3} \right\rfloor\right)$  to  $(N - 1)$ . Any remaining instances from uneven division were assigned to Client 3.

The selection of three clients was deliberate, aiming to emulate a realistic small-scale federated educational setting involving a limited number of collaborating institutions (e.g., three distinct schools or departments) while introducing substantial statistical heterogeneity and class distribution shifts across clients (as clearly demonstrated in Table 2).

Table 2. Class Proportion Distribution Across Clients

| Dataset                                               | Client   | Class Proportion                                     |
|-------------------------------------------------------|----------|------------------------------------------------------|
| Predict Students Dropout and Academic Success         | Client 1 | Class 0: 28.56%, Class 1: 17.03%,<br>Class 2: 54.41% |
|                                                       | Client 2 | Class 0: 33.58%, Class 1: 19.34%,<br>Class 2: 47.08% |
|                                                       | Client 3 | Class 0: 56.87%, Class 1: 43.13%,<br>Class 2: 12.12% |
| Student Performance                                   | Client 1 | Class 0: 8.33%, Class 1: 91.67%                      |
|                                                       | Client 2 | Class 0: 6.94%, Class 1: 93.06%                      |
|                                                       | Client 3 | Class 0: 31.02%, Class 1: 68.98%                     |
| Students Academic Performance Dataset (xAPI-Edu-Data) | Client 1 | Class 0: 39.38%, Class 1: 40.62%,<br>Class 2: 20.00% |
|                                                       | Client 2 | Class 0: 17.50%, Class 1: 46.25%,<br>Class 2: 36.25% |
|                                                       | Client 3 | Class 0: 22.50%, Class 1: 45.00%,<br>Class 2: 32.50% |

This configuration provides sufficient non-IID challenges to rigorously test the robustness of the evaluated FL algorithms, while remaining computationally feasible for reproducible experiments on standard hardware, in line with common practices in federated learning studies employing modestly sized public datasets.

## 2.2 Local Model Architecture: Neural Network

In the federated learning framework, the local model is designed as an artificial neural network [21], [35]. The forward pass processes input features  $x \in \mathbb{R}^d$  through two hidden layers equipped with ReLU activation functions, followed by an output layer. For binary classification, a sigmoid activation is used to generate the positive class probability ( $\hat{y} \in [0,1]$ ). For multiclass classification, a softmax activation produces class probabilities ( $\hat{y} \in \mathbb{R}^K$ ) where  $(\sum_{i=1}^K \hat{y}_i = 1)$ . This process is formally expressed in Equations 2–4.

$$h_1 = \text{ReLU}(W_1 x + b_1) \quad (2)$$

$$h_2 = \text{ReLU}(W_2 h_1 + b_2) \quad (3)$$

$$\hat{y} = \sigma(W_3 h_2 + b_3) \text{ if binary, else } \text{Softmax}(W_3 h_2 + b_3) \quad (4)$$

Here,  $W_1, W_2, W_3$  denote weight matrices, and  $b_1, b_2, b_3$  represent bias vectors. To improve generalization and mitigate overfitting, a dropout mechanism is implemented after each hidden layer, randomly disabling a fraction of units during training.

The hidden layers consist of 64 and 32 neurons, respectively. Table 3 provides a complete summary of the neural network architecture for enhanced reproducibility.

Table 3. Experimental Configuration

| Layer          | Neurons/Units  | Activation Function                      | Dropout Rate | Notes                                                                         |
|----------------|----------------|------------------------------------------|--------------|-------------------------------------------------------------------------------|
| Input          | $d$ (variable) | –                                        | –            | $d$ = number of input features                                                |
| Hidden Layer 1 | 64             | ReLU                                     | 0.2          | Applied after ReLU activation                                                 |
| Hidden Layer 2 | 32             | ReLU                                     | 0.2          | Applied after ReLU activation                                                 |
| Output         | 1 or 3         | Sigmoid (binary)<br>Softmax (multiclass) | –            | 1 neuron for binary classification<br>3 neurons for multiclass classification |

## 2.3. Federated Learning (FL)

Federated Learning (FL) is a distributed machine learning approach that facilitates collaborative model training without transferring sensitive client data to a central server, ensuring data privacy and security, as depicted in Figure 2 [36–38].

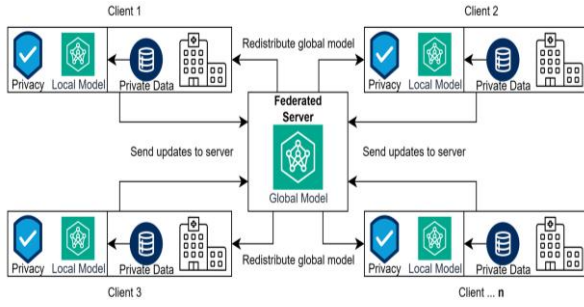


Figure 2. Overview of the Federated Learning Process

In FL, local models are trained on each client's device, with model weights sent to the server for aggregation to update the global model iteratively. FL enhances privacy by keeping model updates confidential during transmission, reducing risks of data reconstruction attacks. This makes FL highly suitable for sensitive fields like education, where compliance with regulations such as GDPR and FERPA is critical, and it supports robust model generalization across diverse, non-IID data distributions while preserving client data sovereignty.

In this study, FL is implemented on three educational datasets with  $C$  clients, where each client  $c \in \{1, 2, \dots, C\}$  possesses a local dataset  $D_c = (x_i, y_i)_{i=1}^{n_c}$ . The number of local instances  $n_c \approx N/C$ , where  $N$  is the total number of instances in the dataset, and  $x_i \in \mathbb{R}^d$  represents the feature vector. The objective of FL is to minimize the global loss function, as shown in Equation 5.

$$L(w) = \frac{1}{N} \sum_{c=1}^C \sum_{(x_i, y_i) \in D_c} l(f(x_i; w), y_i) \quad (5)$$

$w \in \mathbb{R}^p$  denotes the global model parameters,  $f(x_i; w)$  is the neural network's prediction function, and the loss function—binary cross-entropy for binary classification or categorical cross-entropy for multiclass classification—is defined in Equation 6.

$$l(\hat{y}, y) = -[y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})] \text{ if binary,} \\ \text{else } -\sum_{i=1}^K y_i \log(\hat{y}_i) \quad (6)$$

$(\hat{y} = f(x_i; w) \in [0, 1])$  is the predicted probability for the positive class in binary classification, and  $(\hat{y} = f(x_i; w) \in \mathbb{R}^K \text{ with } (\sum_{j=1}^K \hat{y}_j = 1))$  represents the predicted probability distribution over  $K$  classes for multiclass classification.

#### 2.4 Federated Averaging (FedAvg)

Federated Averaging (FedAvg) is an aggregation method that integrates local model weights from clients to create a global model [39]. Each client optimizes its local model to minimize the local loss, as shown in Equation 7.

$$L_c(w_c) = \frac{1}{n_c} \sum_{(x_i, y_i) \in D_c} l(f(x_i; w_c), y_i) \quad (7)$$

Local training employs the Adam optimizer, updating parameters for client  $c$  in round  $t$  according to Equations 8–10.

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) \nabla L_c(w_c^t) \quad (8)$$

$$V_t = \beta_2 V_{t-1} + (1 - \beta_2) (\nabla L_c(w_c^t))^2 \quad (9)$$

$$w_c^{t+1} = w_c^t - \eta \frac{m_t}{\sqrt{V_t + \epsilon}} \quad (10)$$

$m_t$  is the first moment estimate,  $v_t$  is the second moment estimate,  $\eta$  is the learning rate,  $\beta_1, \beta_2 \in [0, 1]$  are momentum parameters, and  $\epsilon$  is a small constant. Post-training, local weights  $\{w_c^{t+1}\}_{c=1}^C$  are aggregated at the server using Equation 11.

$$w^{t+1} = \frac{1}{C} \sum_{c=1}^C w_c^{t+1} \quad (11)$$

This process iterates over multiple rounds to refine the global model.

#### 2.5 Federated Proximal (FedProx)

Federated Proximal (FedProx) extends FedAvg to address data heterogeneity by incorporating a proximal penalty into the local loss function [40]. The modified loss function is defined in Equation 12.

$$L_c(w_c) = \frac{1}{n_c} \sum_{(x_i, y_i) \in D_c} l(f(x_i; w_c), y_i) + \frac{\mu}{2} |w_c - w^t|^2 \quad (12)$$

$\mu$  is the proximal penalty parameter that ensures local weights  $w_c$  stay close to global weights  $w^t$ , reducing model drift in non-IID data. Local training follows the Adam optimizer as in Equations 8–10 with the adjusted loss. Aggregation uses Equation 11, consistent with FedAvg.

#### 2.6 Federated Dynamics (FedDyn)

Federated Dynamics (FedDyn) tackles data heterogeneity by introducing a local control variable  $h_c$  for each client, adding a dynamic penalty to the local loss [41]. The loss function is defined in Equation 13.

$$L_c(w_c) = \frac{1}{n_c} \sum_{(x_i, y_i) \in D_c} l(f(x_i; w_c), y_i) + \frac{\alpha}{2} |w_c - w^t - h_c|^2 \quad (13)$$

$\alpha$  is the penalty parameter. After local training using equations (10)–(12), the control variable is updated, and local weights  $\{w_c^{t+1}\}_{c=1}^C$  are aggregated using equation (13), with adjustments based on the global control variable to maintain model consistency.

#### 2.7 Fair Federated Averaging (q-FedAvg)

Fair Federated Averaging (q-FedAvg) enhances fairness by weighting clients with higher local losses more heavily [25]. The aggregation is performed using  $q$ -norm normalization, as shown in Equation 14.

$$w^{t+1} = \frac{\sum_{c=1}^C n_c (L_c(w_c^t))^q w_c^{t+1}}{\sum_{c=1}^C n_c (L_c(w_c^t))^q} \quad (14)$$

$q$  is the fairness parameter, and  $L_c(w_c^t)$  is the local loss for client  $c$ . Local training uses the Adam optimizer as in Equations 8–10, with the standard local loss from Equation 7.

## 2.8 Stochastic Controlled Averaging for Federated Learning (SCAFFOLD)

Stochastic Controlled Averaging for Federated Learning (SCAFFOLD) uses global and local control variables to mitigate model drift in heterogeneous data [42]. Each client maintains a local control variable  $c_c$ , and the server holds a global control variable  $c$ . Local weight updates account for gradient differences between local and global models. After training with the Adam optimizer as in Equations 8–10, weights are aggregated using Equation 11, and the global control variable is updated as in Equation 15.

$$c^{t+1} = c^t + \frac{1}{C} \sum_{c=1}^C (w_c^{t+1} - w^t) \quad (15)$$

Local control variables are also updated to ensure stable convergence on non-IID data.

## 2.9 Experimental Setup

The experimental setup for this study is summarized in Table 4. Model performance was assessed using a comprehensive set of metrics: accuracy, precision, recall, F1-score, and ROC-AUC. Accuracy reflects the overall proportion of correct predictions, precision measures the accuracy of positive class predictions, recall gauges the detection rate of the positive class, F1-score balances precision and recall, and ROC-AUC evaluates the model's ability to distinguish classes, particularly for imbalanced educational datasets with varying class distributions [43], [44]. To ensure robust and equitable evaluation, a Stratified 5-Fold Cross-Validation approach was applied independently to each client's data [45].

Table 4. Experimental Configuration

| Aspect                     | Detail                                                                                                        |
|----------------------------|---------------------------------------------------------------------------------------------------------------|
| Datasets                   | Predict Students Academic Success, Student Performance, Students Academic Performance Dataset (xAPI-Edu-Data) |
| Client Count (C)           | 3                                                                                                             |
| Training Round (T)         | 50                                                                                                            |
| Cross-Validation Folds (K) | 5                                                                                                             |
| Optimizer                  | Adam                                                                                                          |
| Adam Hyperparameters       | $\beta_1, \beta_2, \epsilon$                                                                                  |
| Epochs per Training Round  | 1                                                                                                             |
| Dropout Probability        | 0.2                                                                                                           |

The federated learning workflow was designed to support multiple aggregation strategies, including FedAvg, FedProx, FedDyn, q-FedAvg, and SCAFFOLD, enabling a comprehensive comparative evaluation under a unified experimental setting. For each aggregation method, the global model parameters were initialized randomly to ensure unbiased starting conditions. In methods that require auxiliary variables, such as SCAFFOLD and FedDyn, global and local control or correction variables were initialized to zero to facilitate variance reduction and drift compensation during training.

Model evaluation was conducted using a cross-validation strategy to improve the robustness of performance assessment across heterogeneous client data distributions. Within each fold, every client's dataset was partitioned into mutually exclusive training and testing subsets. This setup ensured that all aggregation methods were evaluated under identical data splits, allowing for fair and consistent comparisons. The training process was organized into multiple communication rounds, during which decentralized optimization and centralized aggregation were performed iteratively.

At the beginning of each communication round, participating clients synchronized their local models with the latest global parameters distributed by the server. Local training was then performed independently at each client using the Adam optimizer for a fixed number of epochs. The optimizer's first- and second-order moment estimates were reinitialized at the start of each training cycle to maintain consistency across communication rounds. To reduce the risk of overfitting and improve generalization, dropout regularization was applied during local updates.

Several aggregation methods incorporated modifications to the standard local optimization objective to address data heterogeneity and client drift. FedProx constrained local updates by introducing a proximal term that penalized deviations from the global model parameters, while FedDyn applied correction terms to dynamically adjust local objectives based on historical updates. In the SCAFFOLD method, local control variates were updated following each local training phase to reduce gradient variance caused by non-identically distributed client data. After completing local optimization, each client transmitted its updated model parameters to the central server for aggregation.

The server aggregated the received client updates according to the selected aggregation strategy. FedAvg computed a simple arithmetic mean of the local model parameters, serving as a baseline aggregation approach. q-FedAvg applied a weighted aggregation scheme, where client contributions were scaled according to their local loss values raised to a predefined exponent, thereby prioritizing underperforming clients. FedDyn integrated correction-based adjustments during the aggregation stage to mitigate client drift, while SCAFFOLD updated both the global model parameters and the associated global control variables to ensure convergence stability. This aggregation process was repeated for all communication rounds and cross-validation folds, resulting in a final global model for each federated learning method.

All experiments were conducted on a Windows 11 Home 64-bit system equipped with an AMD Ryzen 5 6600H processor operating at 3.3 GHz with 12 logical cores, 16 GB of RAM, and integrated AMD Radeon™ Graphics. The experimental framework was implemented using Python version 3.10.9, with NumPy 1.24.3, TensorFlow 2.13.0, and Keras 2.13.1 as the

primary software libraries. To enhance reproducibility and ensure accessibility for researchers without specialized hardware, all training and evaluation processes were executed exclusively on the CPU, without reliance on dedicated GPU acceleration.

### 3. Results and Discussions

The efficacy of federated learning (FL) models was assessed across three educational datasets—Predict Students Dropout and Academic Success, Student Performance, and xAPI-Edu-Data—using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. These were evaluated through 5-fold cross-validation over 50 communication rounds. Local models were benchmarked against global FL algorithms (FedAvg, FedProx, FedDyn, q-FedAvg, SCAFFOLD) to examine their effectiveness in managing data heterogeneity while ensuring privacy. Performance trends, convergence behaviors, and heterogeneity impacts are depicted in Table 5 and Figure 3, with average metrics summarized in Table 6.

For the Predict Students Dropout and Academic Success dataset, global algorithms, particularly q-FedAvg and FedAvg, demonstrated strong convergence due to the dataset's larger size and simpler features, surpassing local models in recall and ROC-AUC. On the Student Performance dataset, constrained by smaller sample size and complex features, local models underperformed, but global algorithms like q-FedAvg and FedProx achieved notable improvements in key metrics, despite challenges from data heterogeneity affecting convergence. For xAPI-Edu-Data, global algorithms, especially q-FedAvg, exhibited greater stability and superior recall and ROC-AUC compared to variable local models.

Convergence analysis, as illustrated in Figure 3 and supported by the progression of metrics in Table 5 across communication rounds, reveals distinct behaviors among the algorithms. On the larger Predict Students Dropout and Academic Success dataset, most global algorithms achieved rapid convergence within 25–30 rounds, with q-FedAvg and FedAvg exhibiting the smoothest trajectories and minimal oscillations. In contrast, smaller datasets (Student Performance and xAPI-Edu-Data) displayed slower initial convergence and occasional fluctuations, particularly for FedProx and SCAFFOLD, attributable to pronounced client drift in highly non-IID settings. q-FedAvg consistently demonstrated superior stability, reaching near-optimal recall and ROC-AUC earlier than other variants, reflecting its fairness-aware aggregation mechanism that mitigates the influence of poorly performing clients.

Figure 3 presents the convergence trajectories of the local model and the five global FL algorithms across 50 communication rounds, averaged over 5-fold cross-validation, with accuracy plotted on the primary y-axis and recall on the secondary y-axis for each dataset. Subplot (a), corresponding to the Predict Students

Dropout and Academic Success dataset, illustrates rapid and stable convergence of global models, particularly q-FedAvg and FedAvg, which achieve near-maximal accuracy and recall by approximately round 25 with minimal oscillations. Subplot (b) for the Student Performance dataset reveals slower initial progress and greater variability in local models, while global algorithms—led by q-FedAvg and FedProx—gradually surpass local performance, reaching superior recall despite limited sample size. Subplot (c) on the xAPI-Edu-Data dataset highlights pronounced fluctuations in local model performance and the stabilizing effect of federated aggregation, with q-FedAvg exhibiting the most consistent upward trend in both metrics throughout the rounds. Overall, the figure visually confirms the robustness of advanced FL algorithms in handling non-IID data and class imbalance, especially on smaller and more heterogeneous datasets.

Regarding computational complexity, Table 6 reports average training time per round and peak memory usage across all rounds and folds. Global FL algorithms required approximately 3.0–3.8 seconds per communication round (roughly  $2\times$  the local model time of  $\sim 1.6$  seconds), primarily due to the additional aggregation step and, in some cases, penalty term computations (FedProx, FedDyn). Memory consumption for global methods ranged from 19–25 MB compared to 9–11 MB for local training, reflecting the maintenance of multiple client states during simulation. Among the variants, q-FedAvg and FedAvg incurred the lowest overhead, while FedProx exhibited the highest due to its proximal term optimization. These costs remain modest and feasible on standard hardware, confirming FL's practicality for educational deployments without requiring specialized infrastructure.

Global FL algorithms consistently addressed data heterogeneity through aggregation, achieving better convergence than local models, particularly on larger datasets. Even with smaller datasets like Student Performance, FL provided advantages, though limited by sample constraints. These results highlight FL's potential for privacy-preserving predictions in diverse educational settings.

Paired T-tests revealed significant differences only in recall between local and global models, with FedProx ( $T = -4.5169$ ,  $p = 0.0457$ ), FedDyn ( $T = -4.7894$ ,  $p = 0.0409$ ), q-FedAvg ( $T = -5.0382$ ,  $p = 0.0372$ ), and SCAFFOLD ( $T = -4.8430$ ,  $p = 0.0401$ ) showing  $p$ -values  $< 0.05$ , except for FedAvg ( $T = -4.1735$ ,  $p = 0.0529$ ;  $p \geq 0.05$ ). No significant differences were observed in accuracy, precision, F1-score, or ROC-AUC across the datasets for any global algorithms ( $p \geq 0.05$ ).

The superior performance of q-FedAvg in recall and ROC-AUC across all three datasets can be attributed to its adaptive client weighting scheme, which assigns higher influence to clients with greater local loss during aggregation.

Table 5. Performance Metrics for Local and Global Methods Across Academic Datasets

| Dataset                                               | Round | Method                   | Accuracy | Precision | Recall | F1-Score | ROC-AUC | Training Time (s) | Memory Usage (MB) |
|-------------------------------------------------------|-------|--------------------------|----------|-----------|--------|----------|---------|-------------------|-------------------|
| Predict Students Dropout and Academic Success         | 1     | Local                    | 0.6784   | 0.5830    | 0.5350 | 0.4973   | 0.7699  | 1.5653            | 12.6008           |
|                                                       |       | Federated Averaging      | 0.5635   | 0.4278    | 0.4245 | 0.3872   | 0.6098  | 3.0734            | 23.0745           |
|                                                       |       | Federated Proximal       | 0.5628   | 0.4990    | 0.4239 | 0.3863   | 0.6301  | 3.2839            | 27.7448           |
|                                                       |       | Federated Dynamics       | 0.5895   | 0.4773    | 0.4495 | 0.4168   | 0.6609  | 3.1391            | 24.6099           |
|                                                       |       | Fair Federated Averaging | 0.6027   | 0.4591    | 0.4755 | 0.4382   | 0.6852  | 3.1278            | 23.2773           |
|                                                       |       | SCAFFOLD                 | 0.5936   | 0.4729    | 0.4605 | 0.4195   | 0.6512  | 3.1229            | 24.6216           |
|                                                       | 25    | Local                    | 0.6736   | 0.5503    | 0.5312 | 0.4964   | 0.7674  | 1.8944            | 9.6529            |
|                                                       |       | Federated Averaging      | 0.7659   | 0.7104    | 0.6767 | 0.6849   | 0.8809  | 3.5022            | 18.6315           |
|                                                       |       | Federated Proximal       | 0.7612   | 0.7008    | 0.6707 | 0.6779   | 0.8800  | 4.6055            | 19.5958           |
|                                                       |       | Federated Dynamics       | 0.7623   | 0.7032    | 0.6743 | 0.6818   | 0.8801  | 3.5178            | 18.3628           |
|                                                       |       | Fair Federated Averaging | 0.7695   | 0.7147    | 0.6804 | 0.6887   | 0.8805  | 3.4891            | 17.2687           |
|                                                       |       | SCAFFOLD                 | 0.7655   | 0.7093    | 0.6743 | 0.6826   | 0.8796  | 3.5021            | 16.7341           |
|                                                       | 50    | Local                    | 0.6738   | 0.5621    | 0.5319 | 0.4982   | 0.7557  | 1.5261            | 13.3609           |
|                                                       |       | Federated Averaging      | 0.7677   | 0.7117    | 0.6828 | 0.6908   | 0.8814  | 3.0515            | 25.4323           |
|                                                       |       | Federated Proximal       | 0.7621   | 0.7033    | 0.6764 | 0.6837   | 0.8808  | 3.2194            | 28.1648           |
|                                                       |       | Federated Dynamics       | 0.7641   | 0.7051    | 0.6781 | 0.6853   | 0.8817  | 3.0446            | 25.0516           |
|                                                       |       | Fair Federated Averaging | 0.7668   | 0.7114    | 0.6813 | 0.6896   | 0.8810  | 3.0219            | 24.6664           |
|                                                       |       | SCAFFOLD                 | 0.7643   | 0.7068    | 0.6771 | 0.6850   | 0.8803  | 3.0585            | 25.8552           |
| Student Performance                                   | 1     | Local                    | 0.7059   | 0.8586    | 0.7627 | 0.7793   | 0.5386  | 1.6373            | 15.5315           |
|                                                       |       | Federated Averaging      | 0.6441   | 0.8394    | 0.7133 | 0.7442   | 0.4721  | 3.9112            | 24.3836           |
|                                                       |       | Federated Proximal       | 0.7170   | 0.8388    | 0.8201 | 0.8180   | 0.4525  | 3.4416            | 30.7625           |
|                                                       |       | Federated Dynamics       | 0.6876   | 0.8534    | 0.7647 | 0.7935   | 0.5219  | 3.1987            | 29.3857           |
|                                                       |       | Fair Federated Averaging | 0.6698   | 0.8722    | 0.7217 | 0.7448   | 0.5602  | 3.2166            | 27.1354           |
|                                                       |       | SCAFFOLD                 | 0.6718   | 0.8402    | 0.7508 | 0.7721   | 0.4285  | 3.2044            | 26.8719           |
|                                                       | 25    | Local                    | 0.6146   | 0.8534    | 0.6616 | 0.6943   | 0.5907  | 1.6223            | 7.7096            |
|                                                       |       | Federated Averaging      | 0.8473   | 0.8471    | 1.0000 | 0.9131   | 0.7367  | 3.1887            | 16.7583           |
|                                                       |       | Federated Proximal       | 0.8457   | 0.8457    | 1.0000 | 0.9123   | 0.7438  | 3.3952            | 18.6219           |
|                                                       |       | Federated Dynamics       | 0.8473   | 0.8482    | 0.9983 | 0.9130   | 0.7249  | 3.1883            | 16.6143           |
|                                                       |       | Fair Federated Averaging | 0.8442   | 0.8470    | 0.9966 | 0.9114   | 0.7179  | 3.1838            | 16.5591           |
|                                                       |       | SCAFFOLD                 | 0.8473   | 0.8472    | 1.0000 | 0.9131   | 0.7054  | 3.1778            | 16.6714           |
|                                                       | 50    | Local                    | 0.6982   | 0.8577    | 0.7590 | 0.7963   | 0.5435  | 1.5966            | 13.3422           |
|                                                       |       | Federated Averaging      | 0.8503   | 0.8594    | 0.9799 | 0.9123   | 0.7280  | 3.1975            | 24.0247           |
|                                                       |       | Federated Proximal       | 0.8549   | 0.8627    | 0.9844 | 0.9154   | 0.7332  | 3.4386            | 28.5630           |
|                                                       |       | Federated Dynamics       | 0.8533   | 0.8629    | 0.9826 | 0.9146   | 0.7071  | 3.2010            | 26.3206           |
|                                                       |       | Fair Federated Averaging | 0.8580   | 0.8647    | 0.9871 | 0.9176   | 0.7160  | 3.2118            | 25.9878           |
|                                                       |       | SCAFFOLD                 | 0.8518   | 0.8603    | 0.9848 | 0.9141   | 0.7112  | 3.2108            | 27.5250           |
| Students Academic Performance Dataset (xAPI-Edu-Data) | 1     | Local                    | 0.3870   | 0.3918    | 0.3795 | 0.3357   | 0.5465  | 1.5210            | 14.5831           |
|                                                       |       | Federated Averaging      | 0.3646   | 0.3041    | 0.3662 | 0.2986   | 0.5792  | 2.9871            | 26.3849           |
|                                                       |       | Federated Proximal       | 0.2896   | 0.2879    | 0.3118 | 0.2580   | 0.4637  | 3.1902            | 26.5143           |
|                                                       |       | Federated Dynamics       | 0.3937   | 0.3795    | 0.3766 | 0.3303   | 0.5566  | 2.9593            | 24.7766           |
|                                                       |       | Fair Federated Averaging | 0.3812   | 0.3741    | 0.3561 | 0.3261   | 0.5411  | 2.9610            | 24.4776           |
|                                                       |       | SCAFFOLD                 | 0.3271   | 0.3520    | 0.3231 | 0.2847   | 0.4826  | 2.9322            | 25.8198           |
|                                                       | 25    | Local                    | 0.4135   | 0.4183    | 0.4023 | 0.3618   | 0.5861  | 1.6027            | 10.4393           |
|                                                       |       | Federated Averaging      | 0.7146   | 0.7221    | 0.7412 | 0.7193   | 0.8726  | 3.1975            | 21.5247           |
|                                                       |       | Federated Proximal       | 0.6833   | 0.7004    | 0.7132 | 0.6895   | 0.8532  | 3.7225            | 24.9669           |
|                                                       |       | Federated Dynamics       | 0.6625   | 0.6807    | 0.6902 | 0.6680   | 0.8559  | 3.1843            | 19.5151           |
|                                                       |       | Fair Federated Averaging | 0.6833   | 0.6958    | 0.7102 | 0.6893   | 0.8578  | 3.1760            | 21.8112           |
|                                                       |       | SCAFFOLD                 | 0.6917   | 0.7001    | 0.7174 | 0.6985   | 0.8554  | 3.1512            | 21.2430           |
|                                                       | 50    | Local                    | 0.3797   | 0.3961    | 0.3920 | 0.3414   | 0.5965  | 1.4685            | 10.6427           |
|                                                       |       | Federated Averaging      | 0.7292   | 0.7441    | 0.7468 | 0.7321   | 0.8878  | 2.9142            | 22.1044           |
|                                                       |       | Federated Proximal       | 0.7333   | 0.7436    | 0.7605 | 0.7408   | 0.8811  | 3.1331            | 24.0406           |
|                                                       |       | Federated Dynamics       | 0.7083   | 0.7260    | 0.7337 | 0.7162   | 0.8768  | 2.9812            | 17.9292           |
|                                                       |       | Fair Federated Averaging | 0.7396   | 0.7458    | 0.7581 | 0.7413   | 0.8815  | 2.9717            | 21.8029           |
|                                                       |       | SCAFFOLD                 | 0.7271   | 0.7426    | 0.7478 | 0.7314   | 0.8761  | 2.9876            | 19.4117           |

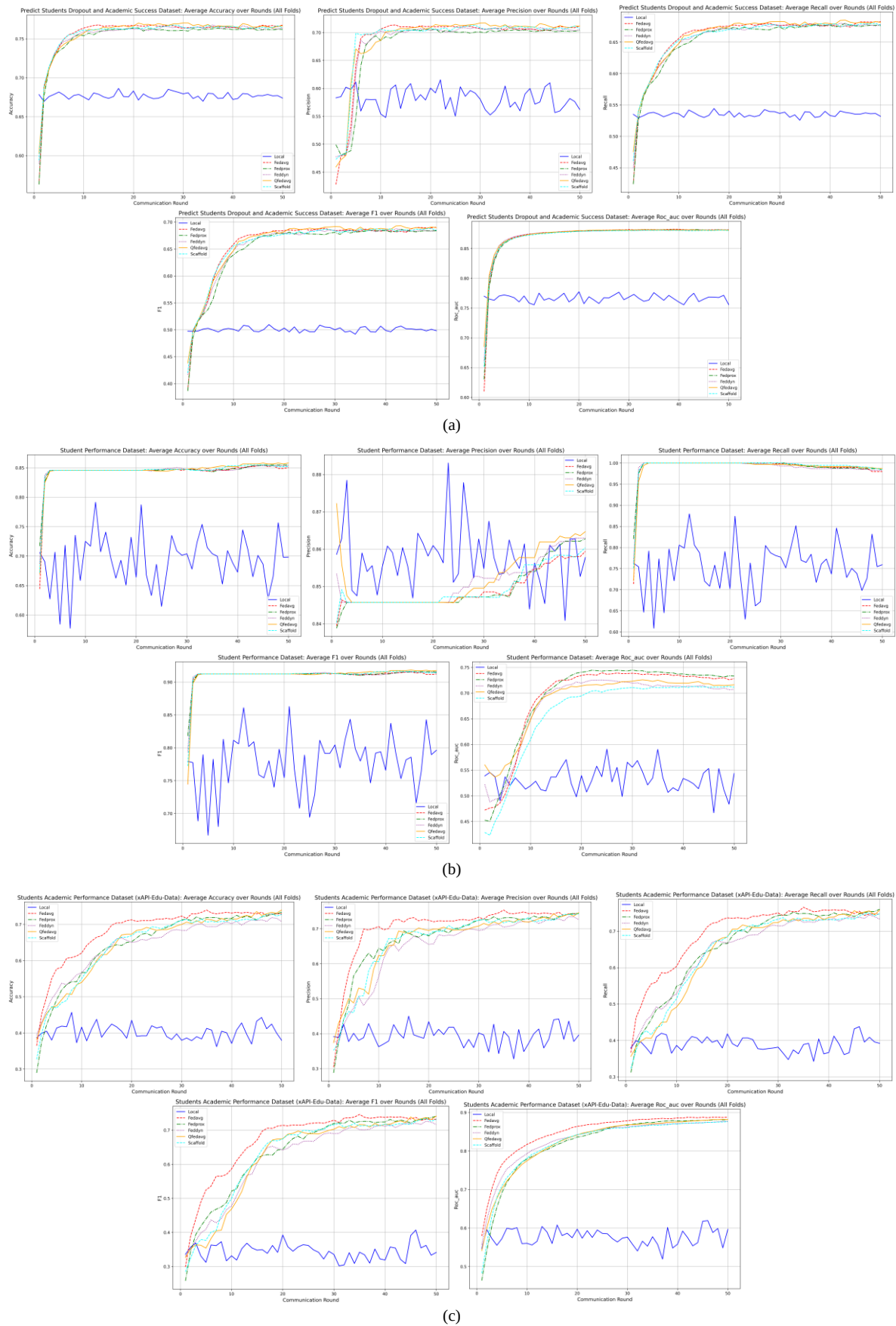


Figure 3. 5-Fold Cross-Validation Performance of Local and Global Average Methods over 50 Communication Rounds across Datasets: (a) Predict Students Dropout and Academic Success, (b) Student Performance, and (c) Students Academic Performance Dataset (xAPI-Edu-Data)

Table 6. Average Performance Metrics Over 5-Fold Cross-Validation and All Rounds (1-50) for Local and Global Average Methods

| Dataset                                               | Method                   | Accuracy | Precision | Recall | F1-Score | ROC-AUC | Training Time (s) | Memory Usage (MB) |
|-------------------------------------------------------|--------------------------|----------|-----------|--------|----------|---------|-------------------|-------------------|
| Predict Students Dropout and Academic Success         | Local                    | 0.6769   | 0.5815    | 0.5352 | 0.5005   | 0.7665  | 1.5921            | 9.8897            |
|                                                       | Federated Averaging      | 0.7570   | 0.6880    | 0.6588 | 0.6599   | 0.8702  | 3.1926            | 19.6515           |
|                                                       | Federated Proximal       | 0.7525   | 0.6793    | 0.6531 | 0.6522   | 0.8700  | 3.6966            | 26.9422           |
|                                                       | Federated Dynamics       | 0.7551   | 0.6847    | 0.6562 | 0.6563   | 0.8715  | 3.1682            | 19.6037           |
|                                                       | Fair Federated Averaging | 0.7575   | 0.6898    | 0.6603 | 0.6613   | 0.8720  | 3.2063            | 19.7224           |
|                                                       | SCAFFOLD                 | 0.7565   | 0.6902    | 0.6572 | 0.6582   | 0.8701  | 3.1681            | 19.5506           |
| Student Performance                                   | Local                    | 0.6929   | 0.8578    | 0.7555 | 0.7782   | 0.5316  | 1.9543            | 11.4601           |
|                                                       | Federated Averaging      | 0.8429   | 0.8490    | 0.9896 | 0.9091   | 0.6923  | 3.5971            | 22.4432           |
|                                                       | Federated Proximal       | 0.8446   | 0.8495    | 0.9919 | 0.9107   | 0.6959  | 3.8242            | 24.7529           |
|                                                       | Federated Dynamics       | 0.8443   | 0.8512    | 0.9893 | 0.9103   | 0.6803  | 3.5852            | 22.2381           |
|                                                       | Fair Federated Averaging | 0.8450   | 0.8521    | 0.9897 | 0.9101   | 0.6870  | 3.5544            | 22.6328           |
|                                                       | SCAFFOLD                 | 0.8444   | 0.8494    | 0.9920 | 0.9104   | 0.6623  | 3.8857            | 22.6534           |
| Students Academic Performance Dataset (xAPI-Edu-Data) | Local                    | 0.4023   | 0.3941    | 0.3894 | 0.3444   | 0.5762  | 1.5318            | 9.6214            |
|                                                       | Federated Averaging      | 0.6812   | 0.7034    | 0.6952 | 0.6723   | 0.8461  | 3.0638            | 19.1479           |
|                                                       | Federated Proximal       | 0.6451   | 0.6729    | 0.6590 | 0.6355   | 0.8186  | 3.2091            | 21.4982           |
|                                                       | Federated Dynamics       | 0.6420   | 0.6488    | 0.6504 | 0.6213   | 0.8275  | 3.0276            | 18.5307           |
|                                                       | Fair Federated Averaging | 0.6462   | 0.6671    | 0.6469 | 0.6262   | 0.8234  | 3.1650            | 22.1208           |
|                                                       | SCAFFOLD                 | 0.6436   | 0.6655    | 0.6490 | 0.6307   | 0.8197  | 3.2750            | 19.7293           |

Table 7. Review of Prior Studies on Machine Learning in Education

| Study      | Year | Approach                                                                                                                                | Dataset(s)                                                                              | Best Accuracy                | Privacy-Preserving | Key Notes                                                                       |
|------------|------|-----------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------|------------------------------|--------------------|---------------------------------------------------------------------------------|
| [25]       | 2025 | Logistic Regression                                                                                                                     | MOOC platform dataset                                                                   | 0.9330                       | No                 | High accuracy on large online dataset, centralized training                     |
| [26]       | 2025 | Multi-model fusion                                                                                                                      | Not specified                                                                           | 0.5100                       | No                 | Lower accuracy, focused on early warning                                        |
| [27]       | 2025 | Logistic regression + RFE                                                                                                               | University of Minho dataset                                                             | 1.0000                       | No                 | Perfect accuracy on small dataset, no privacy measures                          |
| [28]       | 2025 | KNN                                                                                                                                     | 711 high school students (2019–2021)                                                    | 0.9100                       | No                 | Strong performance on domain-specific data                                      |
| [29]       | 2024 | Hybrid (PCA + K-Means + DT)                                                                                                             | OULAD                                                                                   | 0.9370                       | No                 | High accuracy using decision trees, centralized                                 |
| [30]       | 2024 | Random Forest                                                                                                                           | xAPI-Edu-Data                                                                           | 0.9243                       | No                 | Best centralized result on one of the datasets used here                        |
| This Study | 2025 | Federated Learning (best variant):<br>Predict Students Dropout (q-FedAvg)<br>Student Performance (q-FedAvg)<br>xAPI-Edu-Data (q-FedAvg) | Predict Students<br>Dropout and Academic Success, Student<br>Performance, xAPI-Edu-Data | 0.7677,<br>0.8580,<br>0.7396 | Yes                | Privacy-preserving; comparable or superior recall/ROC-AUC on imbalanced classes |

In educational contexts characterized by severe class imbalance and non-IID distributions (evident in Table 2), clients holding minority class samples often exhibit higher loss. By emphasizing these clients, q-FedAvg effectively improves the global model's sensitivity to underrepresented classes (e.g., dropout or failing students), directly enhancing recall and discriminatory power (ROC-AUC) without substantially sacrificing overall accuracy. This mechanism provides a theoretical advantage over standard FedAvg (uniform weighting) and proximal-based methods (FedProx, SCAFFOLD), which may dilute the contribution of informative but high-loss clients, resulting in the observed statistically significant recall gains ( $p < 0.05$ ) for q-FedAvg.

Table 7 positions the performance of the proposed federated learning approaches against prior centralized

machine learning studies on similar or overlapping educational datasets. While traditional centralized methods achieve higher peak accuracies (ranging from 0.9100 to 1.0000 in some cases), they universally neglect privacy considerations, rendering them unsuitable for real-world deployment under regulations such as GDPR and FERPA. In contrast, the federated approaches in this study maintain competitive accuracy—reaching 0.8580 on the Student Performance dataset and 0.7668 on Predict Students Dropout—while strictly preserving data privacy through decentralized training.

Among the FL variants, q-FedAvg consistently emerges as the top performer across datasets (0.7668–0.8580 accuracy), slightly outperforming FedAvg (0.7677 on the largest dataset but lower recall on smaller ones) and clearly surpassing FedProx, FedDyn, and SCAFFOLD

in recall and ROC-AUC (Table 6). This advantage is particularly pronounced on imbalanced and heterogeneous data, where q-FedAvg's fairness-aware aggregation provides greater weight to clients exhibiting higher loss—typically those containing minority class instances—leading to improved detection of at-risk students without requiring raw data exchange.

Thus, the primary contribution of this work extends beyond privacy alone: it demonstrates that advanced FL algorithms can deliver comparable overall accuracy to centralized benchmarks while offering substantial gains in metrics critical for educational intervention (recall and ROC-AUC). These results indicate that the perceived trade-off between privacy and performance is minimal in practice, especially when fairness mechanisms are incorporated, positioning federated learning as a viable and superior paradigm for secure academic analytics.

These findings enhance FL's theoretical framework by emphasizing the need for algorithms tailored to educational data, especially for smaller, heterogeneous datasets prone to model drift. Significant recall improvements ( $p < 0.05$  for FedProx, FedDyn, q-FedAvg, SCAFFOLD) highlight FL's transformative potential in educational analytics, balancing predictive accuracy with privacy. However, the lack of significant differences in accuracy, precision, and F1-score ( $p \geq 0.05$ ) indicates FL's benefits are most evident in metrics sensitive to class imbalance, a key insight for educational applications. Compared to Table 7, where traditional ML prioritizes accuracy over privacy, this FL approach offers a paradigm shift, maintaining performance while ensuring privacy. The evaluation of advanced algorithms like FedDyn and q-FedAvg, previously understudied in education, provides novel insights into managing non-IID data, paving the way for robust FL frameworks in diverse educational contexts.

Despite these advancements, limitations remain, warranting further research. The simulation's use of three clients may not fully reflect real-world educational systems with numerous institutions. Smaller datasets like Student Performance exacerbate heterogeneity challenges, suggesting the need for adaptive FL methods, such as dynamic regularization or client-selection strategies, to improve convergence. Incorporating differential privacy, inspired by healthcare FL studies, could enhance protections against model inversion attacks. Exploring larger client pools and real-world implementations will validate FL's scalability, ensuring compliance with global privacy standards and fostering secure, data-driven educational systems.

#### 4. Conclusions

This study comprehensively evaluates five federated learning algorithms—Federated Averaging (FedAvg), Federated Proximal (FedProx), Federated Dynamics (FedDyn), Fair Federated Averaging (q-FedAvg), and

SCAFFOLD—for privacy-preserving prediction of academic outcomes across heterogeneous educational datasets. The results demonstrate that global FL models, particularly q-FedAvg and FedProx, consistently outperform local centralized models in recall (with statistical significance,  $p < 0.05$  for most variants) and ROC-AUC, achieving competitive accuracies up to 0.8580 while strictly adhering to privacy regulations such as GDPR and FERPA. These findings confirm FL's effectiveness in managing non-IID data distributions without raw data sharing. The key contributions of this work include the first systematic comparative analysis of advanced FL algorithms (notably the under-explored FedDyn and q-FedAvg) in educational contexts, novel insights into the impacts of dataset size and heterogeneity on convergence and performance, and practical evidence that fairness-aware aggregation enhances sensitivity to imbalanced classes critical for identifying at-risk students. Despite these advances, limitations persist, particularly with smaller datasets where high heterogeneity hampers convergence. Future research should focus on adaptive FL strategies, such as dynamic regularization and client selection, alongside larger client pools and differential privacy mechanisms to improve scalability and robustness, ultimately advancing secure, distributed educational analytics for personalized learning and dropout prevention.

#### Acknowledgements

The author declares no conflict of interest. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

#### References

- [1] S. Hilbert *et al.*, "Machine learning for the educational sciences," *Rev. Educ.*, vol. 9, no. 3, p. e3310, 2021, doi: [10.1002/rev3.3310](https://doi.org/10.1002/rev3.3310).
- [2] W. Forero-Corba and F. N. Bennisar, "Techniques and applications of Machine Learning and Artificial Intelligence in education: a systematic review," *RIED-Rev. Iberoam. Educ. Distancia*, vol. 27, no. 1, 2024, doi: [doi.org/10.5944/ried.27.1.37491](https://doi.org/10.5944/ried.27.1.37491).
- [3] L. O. Ofusori and R. Hendradi, "ChatGPT and Its Impact on Students Assessment Practices in the Higher Educational Sector: A Systematic Review," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 11, no. 1, pp. 65–78, Mar. 2025, doi: [10.20473/jisebi.11.1.65-78](https://doi.org/10.20473/jisebi.11.1.65-78).
- [4] L. R. Pelima, Y. Sukmana, and Y. Rosmansyah, "Predicting University Student Graduation Using Academic Performance and Machine Learning: A Systematic Literature Review," *IEEE Access*, vol. 12, pp. 23451–23465, 2024, doi: [10.1109/ACCESS.2024.3361479](https://doi.org/10.1109/ACCESS.2024.3361479).
- [5] M. Taufiq and A. S. Kenyo, "The Legal Protection of Personal Data in the Digital Era: A Comparative Study of Indonesian Law and the GDPR," *Int. J. Bus. Law Educ.*, vol. 6, no. 2, pp. 1260–1268, Aug. 2025, doi: [10.56442/ijble.v6i2.1178](https://doi.org/10.56442/ijble.v6i2.1178).
- [6] S. Radway, K. Quintanilla, C. Ludden, and D. Votipka, "An Investigation of US Universities' Implementation of FERPA Student Directory Policies and Student Privacy Preferences," in *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, New York, NY, USA: Association for Computing Machinery, 2024, pp. 1–35. doi: [10.1145/3613904.3642066](https://doi.org/10.1145/3613904.3642066).
- [7] C. Prince, N. Omrani, and F. Schiavone, "Online privacy literacy and users' information privacy empowerment: the

- case of GDPR in Europe,” *Inf. Technol. People*, vol. 37, no. 8, pp. 1–24, Jan. 2024, doi: [10.1108/ITP-05-2023-0467](https://doi.org/10.1108/ITP-05-2023-0467).
- [8] A. B. F. Mansur and N. Yusof, “The Latent of Student Learning Analytic with K-mean Clustering for Student Behaviour Classification,” *J. Inf. Syst. Eng. Bus. Intell.*, vol. 4, no. 2, pp. 156–161, Oct. 2018, doi: [10.20473/jisebi.4.2.156-161](https://doi.org/10.20473/jisebi.4.2.156-161).
- [9] E. W. T. Darmaningrat, H. M. Astuti, and F. Alfi, “Information Privacy Concerns Among Instagram Users: The Case of Indonesian College Students,” *J. Inf. Syst. Eng. Bus. Intell.*, vol. 6, no. 2, pp. 159–168, Oct. 2020, doi: [10.20473/jisebi.6.2.159-168](https://doi.org/10.20473/jisebi.6.2.159-168).
- [10] R. G. Santosa, Y. Lukito, and A. R. Chrismanto, “Classification and Prediction of Students’ GPA Using K-Means Clustering Algorithm to Assist Student Admission Process,” *J. Inf. Syst. Eng. Bus. Intell.*, vol. 7, no. 1, pp. 1–10, Apr. 2021, doi: [10.20473/jisebi.7.1.1-10](https://doi.org/10.20473/jisebi.7.1.1-10).
- [11] R. Prabaswara, J. Lemantara, and J. Jusak, “Classification of Secondary School Destination for Inclusive Students using Decision Tree Algorithm,” *J. RESTI Rekayasa Sist. Dan Teknol. Inf.*, vol. 7, no. 5, pp. 1009–1017, Aug. 2023, doi: [10.29207/resti.v7i5.5081](https://doi.org/10.29207/resti.v7i5.5081).
- [12] E. Sutoyo and A. Almaarif, “Educational Data Mining for Predicting Student Graduation Using the Naïve Bayes Classifier Algorithm,” *J. RESTI Rekayasa Sist. Dan Teknol. Inf.*, vol. 4, no. 1, pp. 95–101, Feb. 2020, doi: [10.29207/resti.v4i1.1502](https://doi.org/10.29207/resti.v4i1.1502).
- [13] L. Chen, W. Liu, Y. Chen, and W. Wang, “Communication-Efficient Design for Quantized Decentralized Federated Learning,” *IEEE Trans. Signal Process.*, vol. 72, pp. 1175–1188, 2024, doi: [10.1109/TSP.2024.3363887](https://doi.org/10.1109/TSP.2024.3363887).
- [14] H. Chen, H. Wang, Q. Long, D. Jin, and Y. Li, “Advancements in Federated Learning: Models, Methods, and Privacy,” *ACM Comput Surv*, vol. 57, no. 2, Nov. 2024, doi: [10.1145/3664650](https://doi.org/10.1145/3664650).
- [15] K. Hu, S. Gong, Q. Zhang, and others, “An overview of implementing security and privacy in federated learning,” *Artif. Intell. Rev.*, vol. 57, p. 204, 2024, doi: [10.1007/s10462-024-10846-8](https://doi.org/10.1007/s10462-024-10846-8).
- [16] J. Wen, Z. Zhang, Y. Lan, and others, “A survey on federated learning: challenges and applications,” *Int. J. Mach. Learn. Cybern.*, vol. 14, no. 2, pp. 513–535, 2023, doi: [10.1007/s13042-022-01647-y](https://doi.org/10.1007/s13042-022-01647-y).
- [17] M. Ye, W. Shen, B. Du, E. Snezhko, V. Kovalev, and P. C. Yuen, “Vertical federated learning for effectiveness, security, applicability: A survey,” *ACM Comput. Surv.*, vol. 57, no. 9, pp. 1–32, 2025, doi: [10.1145/3720539](https://doi.org/10.1145/3720539).
- [18] H. Guan, P.-T. Yap, A. Bozoki, and M. Liu, “Federated learning for medical image analysis: A survey,” *Pattern Recognit.*, vol. 151, p. 110424, 2024, doi: <https://doi.org/10.1016/j.patcog.2024.110424>.
- [19] Y. Zhang and Y. Liu, “The Research of Predicting Student’s Academic Performance Based on Educational Data,” in *Proceedings of the 2021 5th International Conference on Computer Science and Artificial Intelligence*, in CSAI ’21. New York, NY, USA: Association for Computing Machinery, 2022, pp. 193–201. doi: [10.1145/3507548.3507578](https://doi.org/10.1145/3507548.3507578).
- [20] Tao-Hongli, “Educational data mining for student performance prediction: feature selection and model evaluation,” *J. Electr. Syst.*, vol. 20, pp. 1063–1074, Apr. 2024, doi: [10.52783/jes.3434](https://doi.org/10.52783/jes.3434).
- [21] H. Yuan, F. Zhang, and L. Liu, “Research on Course Recommendation Model Based on MLP and Attention Mechanism,” in *Proceedings of the 2024 International Conference on Intelligent Education and Computer Technology*, in IECT ’24. New York, NY, USA: Association for Computing Machinery, 2024, pp. 99–103. doi: [10.1145/3687311.3687330](https://doi.org/10.1145/3687311.3687330).
- [22] D. C. E. Saputra, A. Azhari, and A. Ma’arif, “K-Nearest Neighbor of Beta Signal Brainwave to Accelerate Detection of Concentration on Student Learning Outcomes,” *Eng. Lett.*, vol. 30, Feb. 2022.
- [23] H. Zhu, J. Xu, S. Liu, and Y. Jin, “Federated learning on non-IID data: A survey,” *Neurocomputing*, vol. 465, Sept. 2021, doi: [10.1016/j.neucom.2021.07.098](https://doi.org/10.1016/j.neucom.2021.07.098).
- [24] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, “Federated Learning: Challenges, Methods, and Future Directions,” *IEEE Signal Process. Mag.*, vol. 37, no. 3, pp. 50–60, 2020, doi: [10.1109/MSP.2020.2975749](https://doi.org/10.1109/MSP.2020.2975749).
- [25] X. Li, S. Zhao, C. Chen, and Z. Zheng, “Heterogeneity-aware fair federated learning,” *Inf. Sci.*, vol. 619, pp. 968–986, 2023, doi: <https://doi.org/10.1016/j.ins.2022.11.031>.
- [26] M. Noble, A. Bellet, and A. Dieuleveut, “Differentially Private Federated Learning on Heterogeneous Data,” in *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, G. Camps-Valls, F. J. R. Ruiz, and I. Valera, Eds., in Proceedings of Machine Learning Research, vol. 151. PMLR, Mar. 2022, pp. 10110–10145. [Online]. Available: <https://proceedings.mlr.press/v151/noble22a.html>
- [27] J. Wang and Y. Yu, “Machine learning approach to student performance prediction of online learning,” *PloS One*, vol. 20, no. 1, p. e0299018, 2025, doi: [10.1371/journal.pone.0299018](https://doi.org/10.1371/journal.pone.0299018).
- [28] W. Zou, W. Zhong, J. Du, and L. Yuan, “Prediction of Student Academic Performance Utilizing a Multi-Model Fusion Approach in the Realm of Machine Learning,” *Appl. Sci.*, vol. 15, no. 7, p. 3550, 2025, doi: [10.3390/app15073550](https://doi.org/10.3390/app15073550).
- [29] S. N. Thakur and B. Rawat, “Prediction of Student Performance in E-Learning Environment using Machine Learning Techniques,” *Procedia Comput. Sci.*, vol. 259, pp. 1416–1425, 2025, doi: [10.51201/JUSST/21/05179](https://doi.org/10.51201/JUSST/21/05179).
- [30] G. L. E. M. Niño et al., “Machine learning model for classifying high school students’ academic performance in mathematics amidst the COVID-19 context,” *JOTSE*, vol. 15, no. 2, pp. 322–334, 2025, doi: [10.3926/jotse.2945](https://doi.org/10.3926/jotse.2945).
- [31] G. Al-Tameemi, J. Xue, I. H. Ali, and S. Ajit, “A hybrid machine learning approach for predicting student performance using multi-class educational datasets,” *Procedia Comput. Sci.*, vol. 238, pp. 888–895, 2024, doi: <https://doi.org/10.1016/j.procs.2024.06.108>
- [32] E. F. Agyemang et al., “Predicting Students’ Academic Performance Via Machine Learning Algorithms: An Empirical Review and Practical Application,” *Comput. Eng. Intell. Syst.*, vol. 15, no. 1, p. 86, 2024, doi: [10.7176/CEIS/15-1-09](https://doi.org/10.7176/CEIS/15-1-09).
- [33] T. Kołodziej and P. Rościszewski, “Towards Scalable Simulation of Federated Learning,” in *International Conference on Neural Information Processing*, Springer, 2021, pp. 248–256, doi: [https://doi.org/10.1007/978-3-030-92307-5\\_29](https://doi.org/10.1007/978-3-030-92307-5_29)
- [34] M. Djukic, I. Prokić, M. Popovic, and others, “Correct orchestration of federated learning generic algorithms: Python translation to CSP and verification by PAT,” *Int. J. Softw. Tools Technol. Transf.*, vol. 27, pp. 21–34, 2025, doi: [10.1007/s10009-025-00795-0](https://doi.org/10.1007/s10009-025-00795-0).
- [35] A. Dureja and P. Pahwa, “Analysis of Non-Linear Activation Functions for Classification Tasks Using Convolutional Neural Networks,” *Recent Patents on Computer Science*, vol. 12, no. 3. Bentham Science Publishers, pp. 156–161, 2019, doi: <https://doi.org/10.2174/2213275911666181025143029>.
- [36] C. Zhang, Y. Xie, H. Bai, B. Yu, W. Li, and Y. Gao, “A survey on federated learning,” *Knowl.-Based Syst.*, vol. 216, p. 106775, 2021, doi: <https://doi.org/10.1016/j.knosys.2021.106775>.
- [37] J. Chen, H. Yan, Z. Liu, M. Zhang, H. Xiong, and S. Yu, “When Federated Learning Meets Privacy-Preserving Computation,” *ACM Comput Surv*, vol. 56, no. 12, Oct. 2024, doi: [10.1145/3679013](https://doi.org/10.1145/3679013).
- [38] A. Khraisat, A. Alazab, S. Singh, T. Jan, and A. Jr. Gomez, “Survey on Federated Learning for Intrusion Detection System: Concept, Architectures, Aggregation Strategies, Challenges, and Future Directions,” *ACM Comput Surv*, vol. 57, no. 1, Oct. 2024, doi: [10.1145/3687124](https://doi.org/10.1145/3687124).
- [39] B. Casella and S. Fonio, “Architecture-Based FedAvg for Vertical Federated Learning,” in *Proceedings of the IEEE/ACM 16th International Conference on Utility and Cloud Computing*, in UCC ’23. New York, NY, USA: Association for Computing Machinery, 2024. doi: [10.1145/3603166.3632559](https://doi.org/10.1145/3603166.3632559).

- [40] J. Cui, Y. Li, Q. Zhang, Z. He, and S. Zhao, "A Federated Learning Framework Using FedProx Algorithm for Privacy-Preserving Palmprint Recognition," in *Biometric Recognition*, S. Yu, W. Jia, X. Shu, X. Yuan, J. Gui, J. Tang, C. Shan, and Q. Liu, Eds., Singapore: Springer Nature Singapore, 2025, pp. 187–196, doi: [https://doi.org/10.1007/978-981-96-1071-6\\_17](https://doi.org/10.1007/978-981-96-1071-6_17)
- [41] Y. Song, H.-H. Chang, and L. Liu, "Federated Dynamic Spectrum Access through Multi-Agent Deep Reinforcement Learning," in *GLOBECOM 2022 - 2022 IEEE Global Communications Conference*, 2022, pp. 3466–3471. doi: [10.1109/GLOBECOM48099.2022.10001688](https://doi.org/10.1109/GLOBECOM48099.2022.10001688).
- [42] S. P. Karimireddy, S. Kale, M. Mohri, S. Reddi, S. Stich, and A. T. Suresh, "SCAFFOLD: Stochastic Controlled Averaging for Federated Learning," in *Proceedings of the 37th International Conference on Machine Learning*, H. D. III and A. Singh, Eds., in *Proceedings of Machine Learning Research*, vol. 119. PMLR, July 2020, pp. 5132–5143. [Online]. Available: <https://proceedings.mlr.press/v119/karimireddy20a.html>
- [43] G. Naidu, T. Zuva, and E. M. Sibanda, "A Review of Evaluation Metrics in Machine Learning Algorithms," in *Artificial Intelligence Application in Networks and Systems*, R. Silhavy and P. Silhavy, Eds., Cham: Springer International Publishing, 2023, pp. 15–25, doi: [https://doi.org/10.1007/978-3-031-35314-7\\_2](https://doi.org/10.1007/978-3-031-35314-7_2)
- [44] I. Issah, O. Appiah, P. Appiahene, and F. Inusah, "A systematic review of the literature on machine learning application of determining the attributes influencing academic performance," *Decis. Anal. J.*, vol. 7, p. 100204, 2023, doi: <https://doi.org/10.1016/j.dajour.2023.100204>.
- [45] S. Bates, T. Hastie, and R. Tibshirani, "Cross-Validation: What Does It Estimate and How Well Does It Do It?," *J. Am. Stat. Assoc.*, vol. 119, no. 546, pp. 1434–1445, 2024, doi: [10.1080/01621459.2023.2197686](https://doi.org/10.1080/01621459.2023.2197686).