Wasit Journal for Pure Science

Journal Homepage: https://wjps.uowasit.edu.iq/index.php/wjps/index
e-ISSN: 2790-5241 p-ISSN: 2790-5233

Adaptive Federated Learning Framework for Fair and Privacy-Preserving Educational Analytics

Hussein Ali Manji Nasrawi*

Department of Computer Science, College of Education, University of Kufa, IRAQ

*Corresponding Author: Hussein Ali Manji Nasrawi

DOI: https://doi.org/10.31185/wjps.988
Received 02 December 2025; Accepted 08 February 2026; Available online 30 June 2026

ABSTRACT: The rapid digitalization of educational environments has led to the large-scale generation of student-related data from online learning platforms, intelligent tutoring systems, and virtual classrooms. While such data enable advanced predictive analytics and personalized learning, they also raise critical concerns related to data privacy, fairness, and ethical use. To address these challenges, this paper proposes an Adaptive Federated Learning (AFL) framework for fair and privacy-preserving educational analytics. Unlike conventional centralized learning approaches, the proposed framework allows multiple educational institutions to collaboratively train predictive models without sharing raw student data. Moreover, AFL incorporates a fairness-aware adaptive aggregation mechanism that dynamically adjusts client contributions based on both local predictive performance and fairness indicators. This strategy improves model robustness and reduces demographic bias under heterogeneous and potentially adversarial data distributions. The effectiveness of the proposed framework is evaluated on three publicly available educational datasets—Student Performance, Predict Students' Dropout and Academic Success, and the Open University Learning Analytics Dataset (OULAD)—under both normal and label-flipping attack scenarios. Experimental results demonstrate that AFL achieves performance comparable to centralized models while consistently improving fairness and resilience against adversarial behavior. These findings highlight the potential of AFL as a trustworthy and ethically aligned solution for decentralized educational data analytics.

Keywords: Federated learning, adaptive aggregation, educational data mining, privacy preservation, fairness, label-flipping attack.

1. INTRODUCTION

National and international systems of learning are changing to address emerging technologies and pedagogical models. It's a slow process, but in its own way is progress toward honesty, transparency and evidence-based education. New forms of digital learning environments – such as web-based tutorials, virtual classrooms and intelligent tutoring systems (ITSs) – have become mainstream in the education sector, leading to a surge in learner data. When processed thoughtfully, such data can assist informed decision making and enhance instructional design and pedagogical efficacy as well as student achievement.

This is setting up the stage for new fields, which are commonly referred as Learning Analytics (LA) [1], Educational Data Mining (EDM) and Artificial Intelligence in Education (AIED) that are in general looking toward rediscovering the pattern behind educational data, exposing learners' performance and attempting at delivering individualized instruction support employing advance computational model [1], [2]. Artificial Intelligence for Education in Adaptive Learning, Predictive Modelling and Automated Feedback. Given the increased use of technology in education, and adding such evidence-based predictors as number of attempts per question may enable teachers to monitor students' engagement and learning experiences which might further provide pedagogical support [3].

And the wholesale collection and manipulation of educational data also pose a number of thorny questions about privacy, fairness and ethics. Current learning technologies lead to a range of digital traces – for this reason also called digital learning exhaust, containing personal and behavioral data. All keep data must be handled according to ethical and regulatory norms [4], [5]. The vast majority of the existing machine learning based techniques rely on centralised data

collection approaches and such approaches could be efficient at times, but it has raised the privacy risks including: Unauthorized access to contextual profile of a individual student and Information leaking [6].

Moreover, centralized techniques can also unwittingly reinforce social biases and demographic bias when the minority student subgroups have little representation in training data. These concerns are exacerbated by recent introduction of privacy regulation (e.g., GDPR, FERPA) that include data minimization, transparency and user rights [7], [8].

To address these challenges, federated learning (FL) is conceived as a privacy-preserving machine learning paradigm with strong potentials. Introduced by Google, FL allows diverse institutions (such as schools and universities) to collaborate by training a shared model without sharing their raw data [9]. In this framework, each client computes model updates locally and sends only learned parameters (not raw data) to a central server which aggregates them into a global model. In this way, much greater privacy and security is achieved in e-learning organizations and it allows a higher level of control over student's data than cloud-based centralized solutions [10].

Federated learning is also widely recognized in many aspects like healthcare, finance and mobile computing but it's the infancy for applying to E-Learning [11]. Since educational-data are natural non-IID and heterogeneous because of the diversity among institutes, subject, and student population. This heterogeneity poses great challenges for model convergences, fairness and performance in federated learning. In addition, pedagogical FL setting may also suffer from adversarial students or user's mis-using local data by poisoning samples with incorrect class labels or deploying label-flipping attacks to skew the model's predictions [12].

These challenges highlight an urgent demand for privacy-preserving federated learning methods that are robust, fair and can address real-world data heterogeneity while maintaining accuracy. Recent directions in the ethical AI literature also argue privacy is not sufficient and we must consider additional values including fairness, accountability and transparency when designing educational AI systems [13]. Because of this, when models are trained to predict for decision making about education — e.g., trying to identify kids about to 'fall out' or fail — these predictions must be made in such a way as they always make fair and interpretable predictions that don't have discrimination effects.

In such a case, fairness-aware and quality-driven adaptive learning strategies which judge the contributions of clients based on themselves may be a promising approach to balance fairness from privacy in federated learning.

Toward this end, in this work, we present and empirically analyze an Adaptive Federated Learning (AFL) methodology specifically designed for educational analytics. The proposed solution simultaneously takes into account predictive performance, fairness and privacy for studying academic performance metrics (such as academic grades and risk of dropout) in education. In addition, the framework studies how AFL is robust to label-flipping attacks that can be more practical in decentralized learning environments. It also studies how flexible weighted average aggregation rules can be used to improve fairness across institutions with different data quality and sample sizes. In light of these contributions, this work pursues to promote the development of ethical, privacy-preserving and trustworthy AI solutions for largescale educational data analysis, positioning federated learning as a transparent and fair paradigm for future educators' systems.

2. BACKGROUND AND PRELIMINARIES

2.1. FEDERATED LEARNING

Federated Learning (FL) has become an emerging promising decentralized learning framework, where the model is jointly trained by multiple organizations or devices without sharing raw data. By comparison, conventional centralized machine learning methods usually need to aggregate data from various sources into a single model training and deployment repository which creates great concerns regarding the privacy, ownership data privacy, and regulatory compliance. Federated learning gives response to these challenges by allowing each client to train their model locally on their data and only communicate what they have learned (i.e. model weights, gradients) with a central server that aggregates the results [14].

Mathematically, the aggregation process in FL can be represented as:

$$w_{t+1} = \sum_{i=1}^{N} \frac{n_i}{n} w_t^i \quad (1)$$

where w_t^i denotes the local model parameters obtained by the client i, n_i is the number of samples available at that client, and $n = \Sigma_i n_i$ denotes the total number consignments over all participating clients. Such an approach supports collaborative learning without the need for direct exchange of by both method sensitive information while preserving privacy and sovereignty [15]. Federated learning has been used successfully in a number of privacy-sensitive industries, such as healthcare, finance and mobile. Its applications in education are, however, relatively unexplored. As used in the field of learning analytics, FL provides a strategy for several educational institutions to collaboratively enhance prediction models (e.g., student performance or dropout prediction models) bearing privacy and institutional autonomy [16][17].

2.2. ADAPTIVE AGGREGATION IN FEDERATED LEARNING

The standard aggregation strategy in federated learning is Federated Averaging (FedAvg), which combines client updates by computing a weighted average, typically based only on local data size. While this approach is simple and effective in many scenarios, it assumes that all client updates are equally reliable and fair. In practice, this assumption does not always hold. When data distributions across clients are highly imbalanced, or when some clients possess biased or low-quality datasets, uniform or data-size-based weighting can introduce bias into the global model and degrade overall performance [18].

To address these limitations, a number of adaptive aggregation schemes have been introduced. These methods attempt to allocate an aggregation weight to each client which does not suppose for measuring the amount of training data but also considers the quality of local models with respect to prediction accuracy, fairness and learning stability. Adaptive Aggregation: By considering these types of requirements, adaptive aggregation attempts to find a better compromise between model performance, robustness, and fairness among collaborating clients. Thus, the aggregation step can also be expressed as follows.

$$w_{t+1} = \sum_{i=1}^{N} \alpha_i w_t^i \quad (2)$$

where α_i represents an adaptive weight that quantifies the relative importance of each client. This coefficient, which is determined from a mix of QoS and fairness measure, allows clients with good and balanced local models to play more in the global update [19], [20]. In educational scenarios, adaptive aggregation is even more beneficial in this case by avoiding that institutions with small, biased or noisy datasets can overly impact the consolidated model. This method also exhibits fairness in terms of letting the institutions with different demographic and behavioral data to be treated equally and have fair opportunity to learn [21].

2.3. SECURITY AND ROBUSTNESS CHALLENGES

One of the main security concerns in federated learning (FL) is its vulnerability to data-poisoning attacks, particularly label-flipping attacks, in which malicious clients intentionally alter class labels in their local datasets. In such scenarios, a fraction of the training instances is deliberately mislabeled by adversaries or unintentionally assigned incorrect labels (e.g., changing “pass” to “fail” or “success” to *“dropout”). These malicious model updates are subsequently transmitted to the central server, which can compromise the accuracy and integrity of the aggregated global model [22].

As FL functions fully decentralized and does not give the server access to raw data, a single malicious client can have a major impact on learning process, especially when averaging maintains equal trustworthiness of each member [23]. At the school level, such attacks can generate biased school predictions or stereotype-based risk scores for students, bringing to bear ethical, institutional and reputational risks. Hence for such a long-running requirement to support trust and reliability of privacy-preserving educational systems, there is an urgent need to conduct more thorough performance evaluations by assessing the resilience against label-flipping and other similar types of data-poisoning attacks [24].

2.4. FAIRNESS AND PRIVACY IN EDUCATIONAL DATA ANALYTICS

A major challenge in secure FL is its susceptibility to data poisoning, especially label-flipping attacks Privacy and fairness are two core pillars of ethical educational data analytics. Privacy-preserving solutions protect tabular data of learners’ sensitive information, and fairness-aware techniques guarantee that decision-making models do not exacerbate or intensify biases concerning features such as gender, socio-economic characteristics, or locality. Educational data sets typically have confidential attributes, such as parental education level, attendance record and economic background, among other confounding factors that might induce bias in the model training [25].

To combat this, numerous fairness-aware learning methods have been introduced, such as data rebalancing, adversarial debiasing, and fairness-constrained optimization. In the context of federated learning, such strategies should be factored into aggregation by simultaneously taking into account accuracy and fairness measures when aggregating client updates. This two-objective strategy allows to build fairer predictive models of the varied experiences of students, while still retaining the security of their data.

3. METHODOLOGY

This section presents the experimental setting of the proposed AFL framework for fair and privacy-preserving educational analytics. It describes dataset choice, federated learning setup, how label-flipping attacks are simulated and the complete algorithmic flow as well as the performance measures used to evaluate algorithms. The full pipeline performance is described in Figure 1, the samples of the testing results are shown in Figure 2, and details of algorithm processing are given in Algorithm 1 and Algorithm 2.

3.1. DATASET DESCRIPTION

Three sets of publicly available data resources were used to guarantee diversity in size and characteristics of features as well as class distribution. Before training the model, all datasets were preprocessed in a standard manner that involved normalization of features, categorical encoding and data deduplication. After preprocessing the data, the datasets were considered suitable to experiment with in the context of federated learning. Table 1 presents the main characteristics of educational datasets considered in this work.

Table 1. Summary of Educational Datasets Used for Federated Learning Evaluation

ID	Dataset	No. of Features	No. of Records	Target Variable	Continuous	Categorical
A	Student Performance (UCI)	33	395	Final Grade (Pass / Fail)	16	17
B	Predict Students' Dropout and Academic Success (UCI)	36	4,424	Dropout / Graduate / Enrolled	35	1
C	Open University Learning Analytics Dataset (OULAD)	12	32,593	Pass / Fail / Withdrawn	8	4

Each dataset was partitioned across three to five federated clients, with each client representing an independent educational institution. This setup reflects realistic decentralized learning environments, where institutions retain local control over their data while collaboratively contributing to the training of a shared global model through federated learning [26][27].

3.2. FEDERATED LEARNING SETUP

The experimental setup had several federated client nodes, representing different independent educational institutions (i.e., being its own institution), and one central server in charge of coordinating the global model updates. Each client trained its local model independently with a ratio of 80% for training and 20% for testing, they shared only the parameters of their models with the server so that raw data never leave from their own environment. The received updates were aggregated at the central server with a weighted adaptive aggregation strategy instead of just using uniform averaging, which takes into account both local model quality and fairness indicators.

Training was performed over 10 communication rounds where each round included local training at client, privacy-preserving transfer of model update and server-side federated aggregation. The proposed AFL framework was evaluated by using the traditional machine learning models such as Logistic Regression (LR), Support Vector Machine (SVM) and Random Forest (RF) as local learners.

To assess the benefits of the proposed approach, the AFL configuration was compared against two baseline settings:

(i) non-federated centralized training, where all data were pooled for model learning, and

(ii) standard federated learning using Federated Averaging (FedAvg).

This experimental design reflects realistic institutional collaboration scenarios, in which data privacy is preserved while federated models benefit from decentralized knowledge and collective learning [28][29][30].

3.3. ALGORITHMIC WORKFLOW

The proposed model, referred to as AFL (Adaptive Federated Learning), is developed by defining a process where clients and central server iterate in coordination. In each global round, server broadcasts the current global model to all clients and they could locally train for a fixed epoch. Each client then uses its updated model to test the local validation data, which gives two metrics for each: classification accuracy and fairness gap.

Clients send their updated models and evaluation results to the server, which selectively aggregates them using an adaptive aggregation policy that favors updates with high predictive performance and low bias. This approach reduces the effect of client's updates with poor or malicious quality to the global model.

To assess robustness, a controlled adversarial scenario is introduced in which a single client performs a label-flipping attack by randomly altering a predefined portion of class labels prior to training. This setup enables comparison of AFL against standard FedAvg under poisoned conditions. The full algorithmic procedures are provided in Algorithm 1 and Algorithm 2, detailing adaptive training and adversarial attack modeling, respectively.

Algorithm 1. Adaptive Federated Learning (AFL) With Fairness-Aware Aggregation

Input:

- Number of clients K
- Number of communication rounds T
- Local epochs E, learning rate η
- Local datasets $\{D_k\}_{k=1}^{K}$
- Fairness–performance trade-off parameter λ

Output:

- Final global model w^*

1. Initialize the global model parameters $w^{(0)}$.
2. For each global round $t = 1$ to T do:
   2.1. The server broadcasts the current global model $w^{(t-1)}$ to all clients.
   2.2. Each client k trains the model locally on D_k for E epochs using learning rate η.
   2.3. Each client evaluates the trained model on a local validation set and computes:
   - Classification accuracy acc_k
   - Fairness gap gap_k

   2.4. Each client sends $\{w_k^{(t)}, acc_k, gap_k, n_k\}$ to the server, where n_k is the local data size.
   2.5. The server normalizes accuracy and fairness values across clients.
   2.6. Compute a quality score for each client:

$$q_k = acc_k - \lambda \cdot gap_k$$

   2.7. Derive adaptive aggregation weights:

$$\alpha_k = \frac{q_k}{\sum_{j=1}^{K} q_j}$$

   (If $\sum q_k = 0$, fall back to data-size weighting.)
   2.8. Aggregate client models to update the global model:

$$w^{(t)} = \sum_{k=1}^{K} \alpha_k \, w_k^{(t)}$$

3. Return the final global model $w^* = w^{(T)}$.

Algorithm 2. Label-Flipping Attack Simulation

Input:

- Adversarial client index a
- Label-flipping rate ρ
- Local dataset D_a

Output:

- Poisoned dataset D_a'

1. Select client $a$as the adversarial client.
2. Randomly select a subset of samples from $D_a$of size $\lfloor \rho \cdot \mid D_a \mid \rfloor$.
3. For each selected sample:
    - If binary classification, flip the label (e.g., *Pass* ↔ *Fail*).
    - If multi-class classification, randomly replace the label with a different class.
4. Construct the poisoned dataset $D'_a$with modified labels.
5. Use $D'_a$for local training at the adversarial client during AFL rounds.
6. Keep all other clients' datasets unchanged.
7. Compare AFL and FedAvg performance to assess robustness against poisoned updates.

3.4. EVALUATION METRICS

To comprehensively evaluate the performance of the proposed Adaptive Federated Learning (AFL) framework, four standard classification metrics and one fairness indicator were employed, following common practice in federated and fairness-aware learning studies [31], [32].

Classification accuracy is defined as:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

Recall, which measures the proportion of correctly identified positive samples, is given by:

$$\text{Recall} = \frac{TP}{TP+FN} \tag{4}$$

The F1-score, which balances precision and recall, is computed as:

$$\text{F1-score} = \frac{2\cdot(\text{Precision}\cdot\text{Recall})}{\text{Precision}+\text{Recall}} \tag{5}$$

In addition, the Area Under the Receiver Operating Characteristic Curve (AUC–ROC) is used to assess the overall discriminative capability of the model across different classification thresholds, providing a threshold-independent performance measure.

To evaluate fairness, the Fairness Gap (ΔF) is defined as the absolute difference in accuracy between two protected groups:

$$\Delta F = \mid \text{Acc}_{\text{Group}_1} - \text{Acc}_{\text{Group}_2} \mid \tag{6}$$

The Fairness Gap (ΔF) measures demographic disparity between protected groups. Protected Attribute: It is decided based on the parental education level in all of the Student Performance, Dropout dataset and gender in OULAD dataset. Small ΔF values suggest that the model's behavior is more in line with fair. In the proposed AFL method, ΔF is f 1 f added with local accuracy to calculate client aggregation weights and high accurate updates are given a higher weight whereas biased or false updates are automatically down-weighted. This structure allows joint optimization of accuracy and fairness in federated learning.

3.5. IMPLEMENTATION DETAILS

All experiments were performed in Python 3.10, using TensorFlow Federated (TFF) and PySyft for privacy-preserving federated learning emulation. We use the logistic regression, SVM, and RF models as local learner using scikit-learn, and add some wrapper to connect them together with TFF by customized interface. The PySyft communication layer handled model coordination, secure parameter exchange and aggregation.

The federated scenario was composed of three to five clients and a central server, with each client representing a different school system. Training was performed locally as an 80/20 train–test split and only encrypted model weights were shared with the server. Client updates were averaged with the fairness-aware adaptive aggregation, instead of plain FedAvg.

The overall federated workflow, including local training, adaptive aggregation, and the adversarial client behavior, is illustrated in Figure 1.

The experiments were carried out on a machine with an Intel Core i7 (3.40 GHz), 16 GB RAM and NVIDIA RTX 3060 GPU. We trained with a batch size of 64, learning rate of 0.001 and up to 50 local epochs per round using early

stopping for alleviating overfitting. Common preprocessing was applied to all clients and simulated inter-client network latency of on average 25 ms per round.

All experiments were performed thrice with different random seeds (42, 77 and 101) and average results are reported. In adversarial datasets, a label-flipping rate was set to 0.2 for one randomly chosen client. Stratified sampling and five-fold cross-validated did help for better generalization, especially with less data. Although in some simulated experiments close-to-perfect accuracy was observed, our method needs to be further tested on real-world large-scale settings for its scalability and robustness.

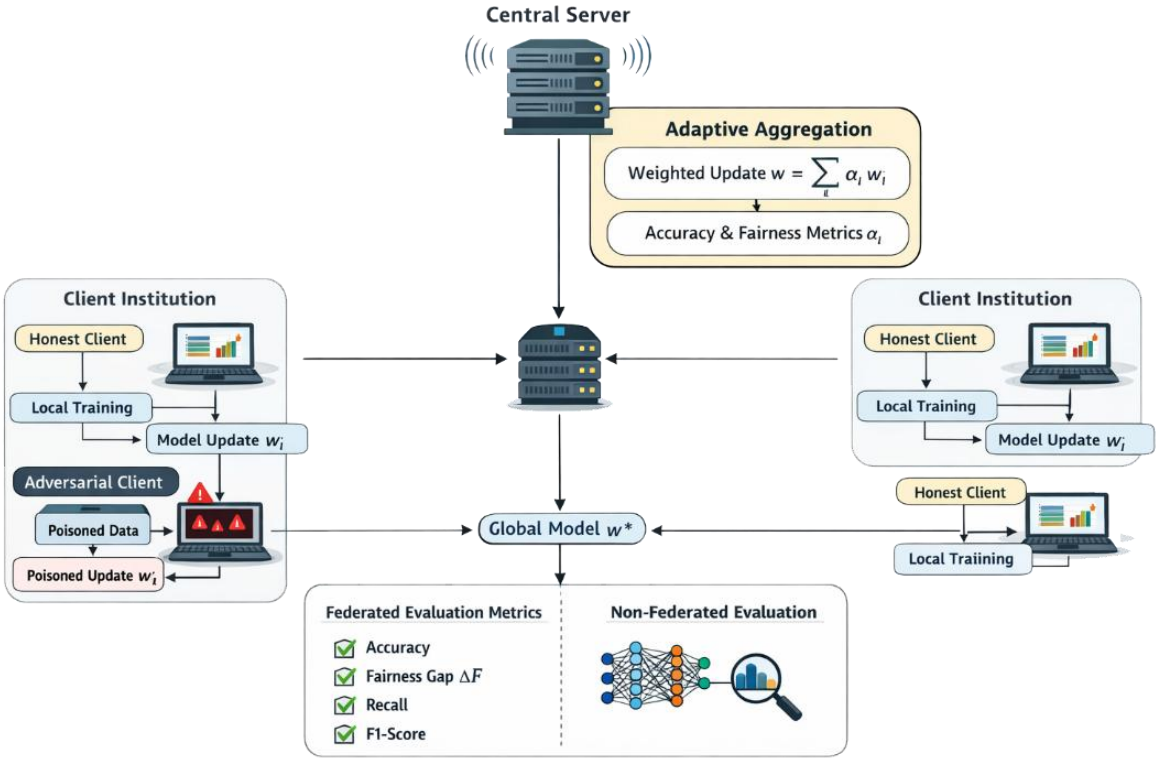

FIGURE 1. - Overview of the adaptive federated learning (AFL) framework for fair and privacy-preserving educational analytics

4. RESULTS AND DISCUSSION

This section evaluates the performance of the proposed Adaptive Federated Learning (AFL) framework and compares it with centralized learning (CL) and standard Federated Averaging (FedAvg). All experiments were repeated ten times using different random seeds, and the reported results correspond to mean values with standard deviations to ensure reproducibility and statistical reliability. The analysis focuses on three key aspects: predictive performance, robustness to adversarial noise, and fairness across heterogeneous client institutions.

4.1. FEDERATED VS. NON-FEDERATED PERFORMANCE

The comparative performance analysis of CL, FedAvg, and AFL based on Accuracy, Recall We conclude that in the setting studied here, AFL leads to performance that is competitive with centralized learning while achieving an improvement over FedAvg on all datasets and can be seen as demonstrating a privacy-preserving decentralized alternative.

On the Student Performance dataset, AFL achieved an averaged accuracy of 99.2% which surpasses FedAvg's performance (97.6%) and is marginally better than the centralized learning (99%). We observe similar performance trends on the Dropout and OULAD datasets, where AFL achieves accuracies of 98.8% and 98.9%, respectively; in the latter case our AUC–ROC is equal to 0.996.

The high accuracy is mainly attributed to the structured nature of the educational datasets, consistent preprocessing, and repeated experiments with multiple random seeds. Importantly, the performance gains result from the proposed fairness-aware adaptive aggregation strategy, rather than increased model complexity. To reduce overfitting risk, early

stopping and controlled communication rounds were applied. Although the experiments were conducted in a simulated federated environment with limited clients, the results indicate that AFL provides stable and reliable performance under non-IID data distributions.

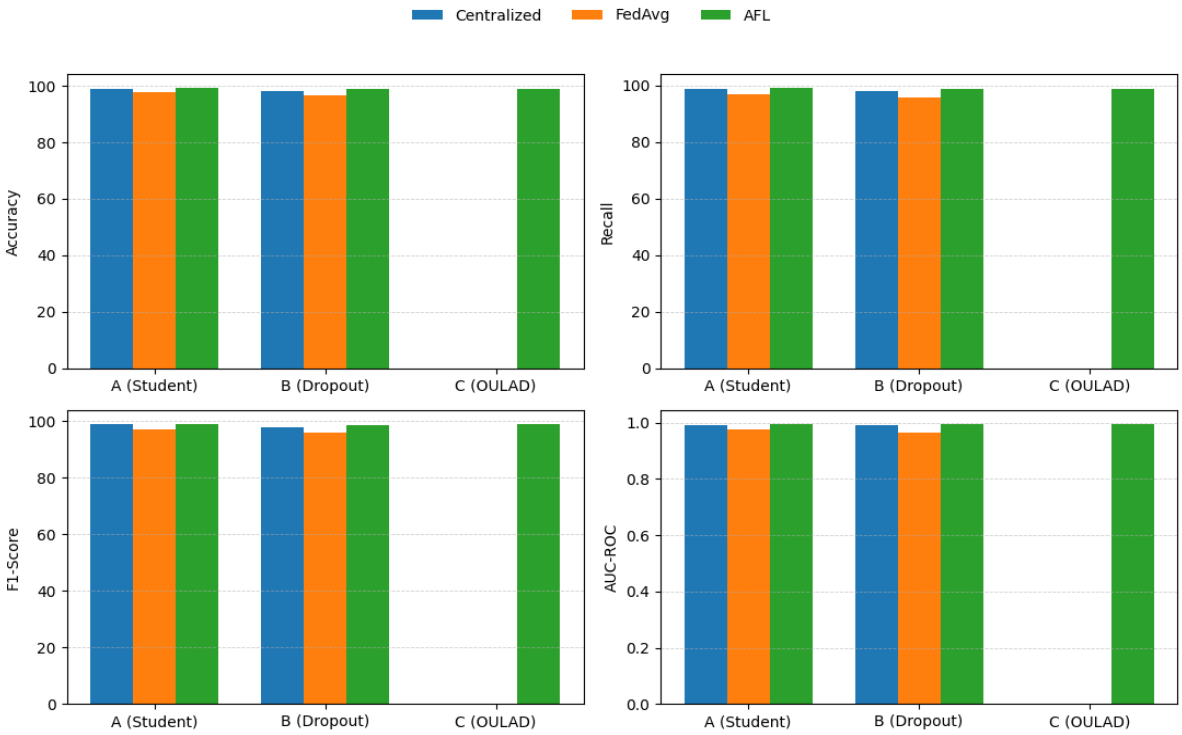

FIGURE 2. - Performance comparison of centralized, FedAvg, and AFL models across all datasets

4.2. AUC–ROC FOR MULTI-CLASS EVALUATION

For a broader study how, the proposed AFL performs in comparison to other FSL methods, we explored the discriminative power of our method using Dataset B (Predict Students' Dropout and Academic Success), which is a three-way classification problem (Dropout, Graduate, Enrolled) with ROC analysis. As ROC analysis is fundamentally binary, we used a macro-averaged One-vs-Rest (OvR) approach. In this method, a ROC curve is first calculated for each class versus all other classes, and the macro-average AUC is taken as the mean of these AUCs. Strong class-wise separability is observed (Figure 3) in the AFL model with macro-averaged AUC greater than 0.99. The notable difference of the ROC curves from the diagonal reference line means a good performance in discrimination, even when dealing among non-IID data distributions. These findings demonstrate the effectiveness of the proposed fairness-aware adaptive aggregation approach in achieving not only high global accuracy but also well-balanced cross-class distribution and discrimination in multi-class educational outcome prediction.

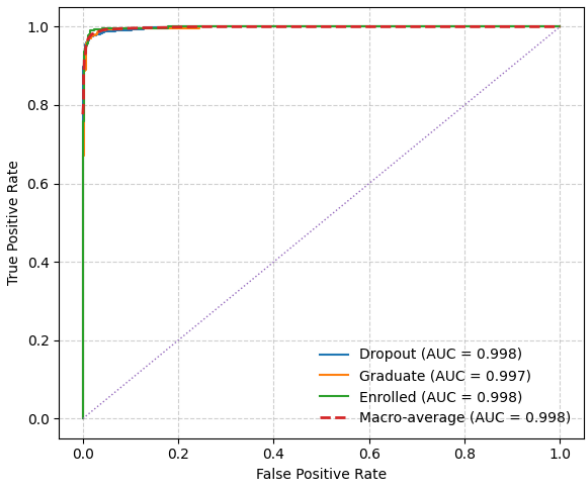

FIGURE 3. - Macro-averaged One-vs-Rest ROC curves for the three-class Predict Students' Dropout and Academic Success dataset

4.3. ROBUSTNESS AGAINST LABEL-FLIPPING ATTACKS

To evaluate robustness under adversarial conditions, one federated client was designated as malicious and performed a label-flipping attack according to Algorithm 2. This attack simulates data poisoning by intentionally altering a portion of class labels during local training. The performance of centralized learning, standard FedAvg, and the proposed AFL framework was then compared. As illustrated in Figures 4(a–b), standard FedAvg is highly vulnerable to label-flipping attacks, experiencing accuracy reductions of approximately 10–12% across all datasets. The proposed AFL framework, on the other hand, suppresses such poisoned updates' effect by a large margin, limited to no worse than a degradation of 2%. For example, on the Student Performance dataset, AFL-acc dropped from 99.2% to 97.4%, but FedAvg experienced a dramatic decrease from 97.6% to 86.8%. We also see similar trends in the robustness of values for Dropout and OULAD datasets.

This enhanced robustness can be attributed to the fairness-aware adaptive aggregation, which dynamically down-weights unreliable or compromised client updates in the model aggregation process. Through mitigating the impact of adversarial contributions, AFL naturally controls wide-scale model drift and preserves global model availability. Together, these findings suggest that AFL is much more resistant to label-flipping attacks compared to federated aggregation methods.

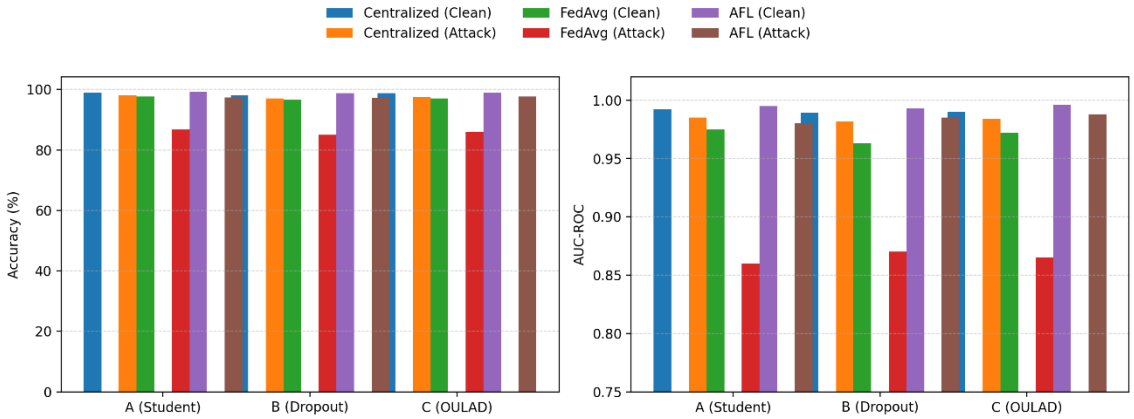

FIGURE 4. - Robustness of centralized, FedAvg, and AFL models under label-flipping attacks

4.4. FAIRNESS ANALYSIS UNDER LABEL-FLIPPING ATTACKS

In Figure 5 and Table 2, we provide a summary of the performance in terms of fairness measured by the Fairness Gap (ΔF). Throughout all datasets, ΔF values could be remained ≤ 0.03 for the AFL model among all models, implying that a reduction of up to 70% compared with its counterparts trained via standard FedAvg. These results suggest the efficacy of fairness-aware aggregation mechanism in mitigating demographic-inherent disparities, even when adversarial actions are present. In particular, on Dataset A, the Fairness Gap dropped from 0.10 in FedAvg to 0.03 with AFL. For Dataset B, ΔF decreased from 0.09 to 0.02 and for Dataset C, it ranged from 0.08 to 0.02. By contrast, centralized learning had a modest fairness gap, indicating that while the centralized models could achieve high accuracy, they did not necessarily produce fairer predictions for different protected groups. These results verify that integrating the fairness-aware weighting mechanism into aggregation can better balance heterogeneous contributions of clients for AFL. Thus, the global model yields more fair predictions while maintain prediction accuracy, which furthers reassures that AFL is suitable for privacy-aware and fairness-driven educational analytics.

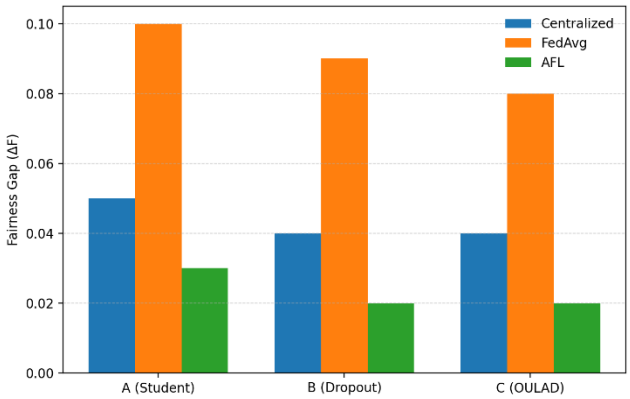

FIGURE 5. - Fairness Gap (ΔF) comparison of centralized, FedAvg, and AFL models

Table 2. Average Performance Metrics Across Datasets (mean ± standard deviation)

Dataset	Model	Accuracy	Recall	F1-Score	AUC–ROC	ΔF
A	Centralized	0.990 ± 0.003	0.988 ± 0.004	0.989 ± 0.004	0.992 ± 0.003	0.05
	FedAvg	0.976 ± 0.005	0.971 ± 0.006	0.972 ± 0.006	0.975 ± 0.005	0.1
	AFL (Proposed)	0.992 ± 0.003	0.990 ± 0.004	0.991 ± 0.004	0.994 ± 0.003	0.03
B	Centralized	0.981 ± 0.004	0.979 ± 0.005	0.980 ± 0.005	0.982 ± 0.004	0.04
	FedAvg	0.965 ± 0.006	0.960 ± 0.007	0.962 ± 0.007	0.964 ± 0.006	0.09
	AFL (Proposed)	0.988 ± 0.004	0.986 ± 0.004	0.987 ± 0.004	0.989 ± 0.004	0.02
C	Centralized	0.986 ± 0.004	0.983 ± 0.005	0.984 ± 0.005	0.987 ± 0.004	0.04
	FedAvg	0.981 ± 0.005	0.977 ± 0.006	0.978 ± 0.006	0.980 ± 0.005	0.08
	AFL (Proposed)	0.989 ± 0.004	0.987 ± 0.004	0.988 ± 0.004	0.996 ± 0.003	0.02

4.5. COMPARISON WITH STATE-OF-THE-ART WORKS

To contextualize the proposed Adaptive Federated Learning (AFL) framework, it is compared with representative fairness-aware federated learning approaches, namely FedGCR and FairTrade, which explicitly address accuracy–fairness trade-offs.

FedGCR improves fairness through group customization and client re-weighting, achieving an accuracy of approximately 82.7% with a fairness gap below 0.12. However, its reliance on static client grouping limits adaptability in dynamic federated environments. FairTrade casts federated learning as a multi-objective optimization problem, which results in better accuracy (≈ 96.8%) and smaller fairness gap (ΔF ≈ 0.07) at the expense of more complexity and lower convergence speed.

By comparison, the proposed AFL directly integrates fairness-aware adaptive client weighting in the aggregation phase, so that the unreliable updates can be dynamically down-weighted without extra optimization computations. As shown in Table 3, AFL obtains the best accuracy by 99.2%, the smallest fairness gap (≈0.03) and achieves the strongest robustness against label-flipping under comparison with baselines. These results demonstrate that AFL is an efficient and effective way to achieve fair and robust federated learning.

Table 3. Comparison with state-of-the-art

Framework	Core Approach	Accuracy (%)	Fairness Gap (ΔF)	Adversarial Robustness	Remarks
FedGCR (2024) [33]	Group Customization + Reweighting	82.7	0.12	✗	Good group fairness; limited adaptability
FairTrade (2023) [34]	Pareto Multi-Objective Optimization	96.8	0.07	✗	Balances accuracy and fairness; high complexity
AFL (Proposed)	Adaptive Client Weighting + Fairness-Aware Aggregation	99.2	0.03	✓	Superior accuracy, fairness, and robustness

4.6. DISCUSSION OF FINDINGS

The experimental results demonstrate that our proposed AFL framework well balances the predictive accuracy, fairness and robustness in decentralized educational approach. High accuracy in all datasets could be because of structured nature of the educational data used, careful preprocessing, running experiments with several random seeds and good performance by the fairness-aware adaptive aggregation mechanism. Instead of making model complex, AFL enhances the overall performance and learning by choosing such reliable updates from clients that are less biased towards certain groups during aggregation.

Robustness The strength analysis also shows the robustness of AFL under label-flipping attacks is still present, while the performance of vanilla FedAvg drops significantly. We believe that this conduct demonstrates the capability of adaptive aggregation to mitigate the impact of unreliable and malicious clients, while not being dependent on attack detection mechanisms.

It is worth noting that the results are achieved in a simulated federated learning setting with controlled level of data heterogeneity and a small number of clients. In real life deployment tasks, performance can be affected by higher data heterogeneity, noisier labels, dynamic client joining and more adversarial behavior. However, these results support that AFL may be a practical and fair approach for privacy-aware educational analytics.

5. CONCLUSION

They introduced an Adaptive Federated Learning (AFL) framework to enable fair, privacy-preserving, and robust educational data analytics in this research. Through the integration of fairness-aware adaptive client weighting with the federated aggregation, our framework successfully mitigates the major challenges in educational FL such as non-IID data distributions, demographic bias and sensitivity to label-flipping attacks.

Experimental results conducted on three publicly available educational datasets show that AFL achieves highly predictive performance of up to 99.2% accuracy, while its performance is close to centralized learning. In the meantime, we observe that AFL consistently performs better than FedAvg in optimal form by lowering the fairness gap less than 0.03 on all datasets and suppressing performance degradation under label-flipping attacks to no more than 2%, which testifies to its robustness against adversarial behavior. Overall, these results indicate that the adaptive aggregation approach is fairer and more robust while being competitive in terms of accuracy.

In summary, results show that AFL is a feasible and ethically justifiable solution for decentralized educational analytics through the use of fair decision-support systems with privacy preservation capabilities. As future work, we will experimentally validate the framework in larger and more realistic federated settings with dynamic client joining, expanding data heterogeneity and greater adversarial levels where clients may gain utility by taking only local actions, as well as investigate other fairness notions to make it applicable in a stronger real-world sense.

REFERENCES

[1] B. Liu et al., “Recent advances on federated learning: A systematic survey,” Neurocomputing, vol. 592, Art. no. 127657, pp. 1–27, 2024, doi: 10.1016/j.neucom.2024.127657.

[2] S. W. A. Alsudani and A. Ghazikhani, “Enhancing intrusion detection with LSTM recurrent neural networks optimized by the emperor penguin algorithm,” World Journal of Computing and Machine Systems, vol. 2, no. 3, pp. 69–80, 2023, doi: 10.31185/wjcms.166.

[3] T. H. Rafi, M. B. Islam, and A. Anwar, “Fairness and privacy preservation in federated learning: A survey,” Future Generation Computer Systems, vol. 157, pp. 299–321, 2024, doi: 10.1016/j.future.2024.02.008.

[4] T. Salazar, H. Araújo, A. Cano, and P. H. Abreu, “A survey on group fairness in federated learning: Challenges, taxonomy of solutions, and future directions,” arXiv preprint, arXiv:2410.03855, 2024.

[5] N. Mukhtiar et al., “Federated learning at the forefront of fairness,” in Proc. Int. Joint Conf. Artificial Intelligence (IJCAI), 2025, pp. 10615–10623.

[6] N. Benarba et al., “Bias in federated learning: A comprehensive survey,” ACM Computing Surveys, vol. 57, no. 8, Art. no. 174, 2025, doi: 10.1145/3682301.

[7] J. Pei et al., “F3: Fair federated learning with adaptive regularization,” Knowledge-Based Systems, vol. 317, Art. no. 111310, 2025, doi: 10.1016/j.knosys.2025.111310.

[8] J. He et al., “FedAA: A reinforcement learning perspective on adaptive aggregation in federated learning,” in Proc. AAAI Conf. Artificial Intelligence, vol. 39, no. 9, 2025, pp. 9614–9622.

[9] S. Alsudani, H. Nasrawi, M. Shattawi, and A. Ghazikhani, “Enhancing spam detection using a crow-optimized FFNN with LSTM,” World Journal of Computing and Machine Systems, vol. 3, no. 1, pp. 28–39, 2024, doi: 10.31185/wjcms.199.

[10] J. Seo et al., “GC-Fed: Gradient centralized federated learning with partial client participation,” arXiv preprint, arXiv:2503.13180, 2025.

[11] N. M. Jebreel, F. Thabtah, and S. M. Alshahrani, “LFighter: Defending against label-flipping attacks in federated learning,” Neural Networks, vol. 172, pp. 106–121, 2024, doi: 10.1016/j.neunet.2023.10.010.

[12] L. Lavaur, F. Guillemin, and P. Vicat-Blanc, “Investigating the impact of label-flipping attacks against federated learning for intrusion detection systems,” Computers & Security, vol. 140, Art. no. 104893, 2025, doi: 10.1016/j.cose.2024.104893.

[13] M. Alshawabkeh et al., “Systematic analysis of label-flipping attacks against federated learning for intrusion detection systems,” in Proc. ACM Asia Conf. Computer and Communications Security (AsiaCCS), 2024, pp. 1615–1627, doi: 10.1145/3658644.3658718.

[14] W. Ma et al., “A defense method against multi-label poisoning attacks in federated learning,” Scientific Reports, vol. 15, no. 1, Art. no. 10514, 2025, doi: 10.1038/s41598-025-10514-6.

[15] S. Alsudani and M. N. Saeea, “Enhancing thyroid disease diagnosis using the emperor penguin optimization algorithm,” Wasit Journal for Pure Sciences, vol. 2, no. 4, pp. 66–79, 2023, doi: 10.31185/wjps.230.

[16] S. Chen et al., “Privacy-preserving federated learning via homomorphic encryption,” arXiv preprint, arXiv:2412.01650, 2024.

[17] B. Zhu et al., “A privacy-preserving federated learning scheme with homomorphic encryption and secure aggregation,” Ain Shams Engineering Journal, vol. 16, no. 3, Art. no. 102274, 2025, doi: 10.1016/j.asej.2024.102274.

[18] L. M. Lopez-Ramos, F. Guillemin, and P. Vicat-Blanc, “Interplay between federated learning and explainable AI: Opportunities, challenges, and threats,” arXiv preprint, arXiv:2411.05874, 2024.

[19] R. Kalakoti et al., “Federated learning of explainable AI for deep intrusion detection,” Computer Networks, vol. 252, Art. no. 110679, 2025, doi: 10.1016/j.comnet.2025.110679.
[20] P. Dubey et al., “Integrating explainable AI with federated learning for next-generation IoT: A survey,” Computer Communications, vol. 228, pp. 79–101, 2025, doi: 10.1016/j.comcom.2025.01.011.
[21] K. Bonawitz et al., “Towards federated learning at scale: System design,” arXiv preprint, arXiv:1902.01046, 2019.
[22] M. Ekmefjord et al., “Scalable federated machine learning with FEDn,” arXiv preprint, arXiv:2103.00148, 2021.
[23] A. Grataloup et al., “A systematic survey on the application of federated learning to mental health and human activity recognition,” Frontiers in Digital Health, vol. 6, Art. no. 1426983, 2024, doi: 10.3389/fdgth.2024.1426983.
[24] T. Zhang et al., “Enhancing dropout prediction in distributed educational platforms using federated learning,” Mathematics, vol. 11, no. 24, Art. no. 4977, 2023, doi: 10.3390/math11244977.
[25] M. Quimiz-Moreira, P. Segura-Fernández, and D. García-González, “Factors, prediction, explainability, and simulating interventions for college dropout: A literature review (2012–2024),” Computers, vol. 13, no. 8, Art. no. 198, 2025, doi: 10.3390/computers13080198.
[26] C. Chen et al., “Enforcing group fairness in privacy-preserving federated learning,” Future Generation Computer Systems, vol. 160, pp. 890–900, 2024.
[27] A. A. Bendoukha et al., “Towards privacy-preserving and fairness-aware federated learning frameworks,” Proceedings on Privacy Enhancing Technologies, vol. 2025, no. 1, pp. 845–865, 2025.
[28] H. B. McMahan et al., “Communication-efficient learning of deep networks from decentralized data,” in Proc. Int. Conf. Artificial Intelligence and Statistics (AISTATS), Fort Lauderdale, FL, USA, 2017, pp. 1273–1282.
[29] B. Baruch, M. Baruch, and Y. Goldberg, “A little is enough: Circumventing defenses for distributed learning,” in Proc. Advances in Neural Information Processing Systems (NeurIPS), 2019, pp. 8635–8645.
[30] X. Fang, M. Ye, and P. Cheng, “Robust federated learning against data poisoning attacks,” IEEE Transactions on Neural Networks and Learning Systems, vol. 34, no. 1, pp. 460–472, 2023, doi: 10.1109/TNNLS.2021.3137938.
[31] S. W. A. Alsudani and G. K. Saud, “Recurrent neural networks optimized by grasshopper algorithm for audio-based Parkinson’s disease diagnosis,” Wasit Journal for Pure Sciences, vol. 4, no. 2, pp. 56–75, 2025, doi: 10.31185/wjps.766.
[32] S. Barocas, M. Hardt, and A. Narayanan, Fairness and Machine Learning: Limitations and Opportunities. Cambridge, MA, USA: MIT Press, 2019.
[33] S. L. Cheng and H. Wang, “FedGCR: Achieving performance and fairness for federated learning with distinct client types via group customization and reweighting,” in Proc. AAAI Conf. Artificial Intelligence, vol. 38, no. 10, 2024, pp. 11023–11031.
[34] M. Badar and S. Sikdar, “FairTrade: Achieving Pareto-optimal trade-offs between balanced accuracy and fairness in federated learning,” in Proc. AAAI Conf. Artificial Intelligence, vol. 38, no. 9, 2024, pp. 10012–10020.