
Federated and Drift-Resilient Machine Learning for Financial Fraud and Distress Detection

Adeyemi Mansour^{1*}, Adel Rana²

¹ United Arab Emirates University, UNITED ARAB EMIRATES

² Sapienza University of Rome, ITALY

ABSTRACT

Federated learning allows multiple financial institutions to train a shared fraud-detection or financial-distress model collaboratively without any institution exposing its raw transaction or account data to the others, a privacy-preserving property that has made it an attractive architecture for cross-institutional financial machine learning. That architecture, however, was designed around an assumption that becomes increasingly fragile the longer a federation operates in production: that each participating institution's local data distribution remains reasonably stable, so that averaging every institution's model updates together each communication round continues to produce a sensible shared global model. In practice, individual institutions experience concept drift on different timelines and for different reasons — a regional economic shock affecting one institution's borrower population, a coordinated fraud campaign targeting one institution's card portfolio specifically, a new product line at one institution shifting its transaction mix — so a federation's participating institutions rarely drift in unison, and a global aggregation rule that treats every institution's update identically regardless of how reliable that update currently is risks corrupting a well-functioning shared model with one institution's idiosyncratic, currently unreliable local signal. This paper develops Federated Learning with Adaptive Resilience to drift Events (FLARE), a framework that adds two mechanisms to a standard federated learning pipeline without weakening its core privacy guarantee: each participating institution runs local concept-drift detection on its own data and shares only a drift status and severity signal, never raw data or gradients derived from suspect data, with the central aggregation server; the server then down-weights or temporarily excludes updates from institutions currently reporting severe, unconfirmed drift when computing the shared global model, protecting the federation's other members from a single institution's idiosyncratic instability; and an institution confirmed to be experiencing genuine, sustained local drift is offered an optional personalized local fine-tuning branch layered on top of the shared global model, so that institution can adapt to its own changed conditions without waiting for federation-wide consensus and without dragging the shared global model off course for everyone else. Following a design-science research approach, the framework is illustrated, not validated, through two comparative tables, a hypothetical case scenario, and four illustrative figures, with every quantitative value explicitly and repeatedly labeled as an illustrative estimate constructed from the cited literature's directional findings rather than as measured data from a deployed federation. The paper concludes with a discussion of FLARE's practical implications for financial institutions participating in cross-institutional federated learning, its assumptions and limitations, and directions for future empirical research.

Keywords: Federated Learning; Concept Drift; Non-IID Data; Adaptive Machine Learning; Fraud Detection; Financial Distress; Privacy-Preserving Machine Learning

Introduction

Federated learning was developed to let multiple parties train a shared machine-learning model collaboratively while each party's raw data remains on its own systems, exchanging only model updates

typically gradients or parameter changes rather than the underlying records those updates were computed from [1]. For financial institutions, this property is especially valuable for exactly the applications this paper addresses: fraud detection and financial-distress prediction both benefit from pooling patterns across many institutions' data, because fraud and distress signals visible at one institution are often too sparse, in isolation, to train a strong model, while sharing raw transaction or account records across institutions is generally foreclosed by data-privacy law, competitive sensitivity, and customer expectations of confidentiality. A federated architecture lets several institutions jointly benefit from each other's data patterns without any institution ever seeing another's underlying records, and prior work in this line of research has established the technical feasibility and privacy-preserving value of exactly this arrangement for financial fraud detection specifically [1-9].

That architecture's standard aggregation mechanism, however, rests on an assumption this paper argues becomes increasingly untenable the longer a financial federation operates. The canonical federated averaging algorithm computes the shared global model as a weighted average of every participating institution's local model update, typically weighted by the size of each institution's local dataset, implicitly treating every institution's current update as an equally trustworthy signal about how the shared model should move [1]. This treatment is reasonable when every institution's underlying data distribution is reasonably stable and reasonably similar to the others', but financial institutions participating in a real federation rarely satisfy either condition for long. Different institutions serve different regional economies, different customer segments, and different product lines, so their local data distributions differ from each other even under stable conditions — the statistical heterogeneity problem the federated learning literature calls non-independent-and-identically-distributed, or non-IID, data [2]. More consequentially for this paper's argument, each institution's own local distribution continues to change after the federation is established, and it changes on its own timeline: a regional economic shock might shift one institution's borrower population's credit-risk characteristics while leaving another institution, serving a different region, largely unaffected; a coordinated fraud campaign might target one institution's card portfolio specifically without touching another's; a new product launch might shift one institution's transaction mix in ways no other participating institution experiences at all [10-15].

This combination heterogeneous baselines, and drift that arrives at different institutions at different times for different reasons creates a specific vulnerability a standard federated averaging rule has no mechanism to address. If one participating institution's local data begins drifting severely, that institution's local model update, computed on its own now-shifting data, becomes a less reliable estimate of how the shared global model should evolve, yet a standard uniform-weighting aggregation rule continues to fold that update into the global model at the same weight it always has, exactly as if nothing had changed. The result is that one institution's idiosyncratic instability can measurably degrade the shared model's performance for every other participating institution, none of whose own data has drifted at all a failure mode that undermines the collaborative value the federation exists to provide in the first place. Existing federated learning research has developed methods for the general non-IID heterogeneity problem [2] and, more recently, has begun to characterize how concept drift specifically manifests when it is distributed unevenly across a federation's clients [3], [4], but this literature has not yet been synthesized into an operational framework specifically for financial fraud-detection and distress-prediction federations, where the privacy-preserving rationale for federating in the first place never sharing raw data, only a minimal status signal, between institutions must be preserved even as the framework adds drift-awareness to the aggregation process.

A second, related gap concerns what should happen for the institution that is actually drifting once its drift is detected. A federation that simply excludes a drifting institution's updates from aggregation protects the other members but leaves the drifting institution itself without a path back to a well-performing local model; that

institution must either wait for its own local distribution to stabilize and hope the shared global model remains a reasonable starting point when it does, or abandon the federation's shared model altogether and retrain independently, forfeiting the collaborative benefit the federation was built to provide. Neither option serves the drifting institution well, and this paper argues a third option a personalized local adaptation layered on top of, rather than replacing, the shared global model better serves an institution experiencing genuine, sustained local drift without requiring it to either wait indefinitely or exit the federation.

This paper addresses both gaps by developing Federated Learning with Adaptive Resilience to drift Events (FLARE): a framework in which each participating institution runs local, privacy-preserving concept-drift detection and reports only a drift status and severity signal to the central server, never raw data or drift diagnostics computed from it; the server aggregates institutions' model updates using a drift-aware selective weighting scheme that down-weights or temporarily excludes updates from institutions currently reporting severe, unconfirmed drift; and an institution confirmed to be experiencing sustained local drift is offered an optional personalized local fine-tuning branch on top of the shared global model, allowing it to adapt on its own timeline without degrading the shared model for the federation's other members [16-19].

Literature Review

Federated Learning Foundations and the Privacy-Preserving Rationale

The technical foundation this paper builds on is the federated learning paradigm itself. McMahan, Moore, Ramage, Hampson, and y Arcas [1] introduce federated averaging, the canonical algorithm in which a central server distributes a shared global model to participating clients, each client computes a local update using only its own data, and the server aggregates those updates typically weighted by local dataset size — into an improved global model, repeating this process over successive communication rounds without any client ever transmitting its raw data to the server or to any other client. This architecture is what makes federated learning attractive for financial institutions specifically: several institutions can jointly benefit from patterns present across their collective data without any institution's transaction or account records ever leaving its own systems, addressing both the data-privacy and competitive-sensitivity concerns that would otherwise foreclose direct data pooling. Li, Sahu, Zaheer, Sanjabi, Talwalkar, and Smith [2] extend this foundation by developing FedProx, an aggregation approach designed specifically to handle statistical and systems heterogeneity across clients more gracefully than standard federated averaging, by adding a proximal term that limits how far a client's local update can drift from the current global model within a single communication round. This paper's aggregation mechanism builds directly on the heterogeneity-awareness FedProx establishes, extending it specifically to the case where a client's heterogeneity is not a stable baseline difference but an actively evolving concept-drift event.

Non-IID Heterogeneity and Drift in Federated Settings

A more specific literature addresses what happens when federated learning's clients do not merely differ from each other in a stable way, but change over time, and change on different timelines from one another. Jothimurugesan, Hsieh, Wang, Joshi, and Gibbons [3] characterize concept drift specifically in a federated setting, showing that drift arriving at different clients at different times — rather than affecting the entire federation simultaneously creates a distinct technical problem existing federated optimization methods are not designed to handle, because a client's drifted local update looks, from the aggregation server's perspective, similar to ordinary non-IID heterogeneity rather than to a genuine change in the underlying relationship that client's data reflects. Canonaco, Bergamasco, Mongelluzzo, and Roveri [4] develop an adaptive federated learning approach explicitly designed to operate in the presence of concept drift, demonstrating that detecting drift at the client level and adjusting the federation's aggregation behavior in response can preserve global model quality in a way a drift-unaware federated optimization method cannot.

Read together, this literature establishes both those the technical problem this paper addresses is real and increasingly well documented, and that client-level drift detection combined with an aggregation-level response is the direction the federated learning literature itself has begun to move, though neither paper develops the specific privacy-preserving status-signal-only detection-and-reporting mechanism, nor the personalized local fine-tuning branch, this paper's framework adds [20-26].

Concept Drift and Adaptive Learning in Financial Machine Learning

A foundational, non-federated literature establishes the concept-drift premise this paper's client-level detection mechanism draws on. Hand [5] documents that a deployed classifier's real-world performance can diverge from its documented validation statistics as the scored population evolves, a phenomenon he terms the illusion of progress. Gama, Žliobaitė, Bifet, Pechenizkiy, and Bouchachia [6] survey the broader concept-drift literature, establishing that continuous monitoring, rather than a fixed retraining calendar, is necessary to catch this decay in time. Widmer and Kubat [7] contribute the real-versus-virtual drift distinction this paper's local detection stage draws on: a genuine change in the relationship a model has learned is a different event, for adaptation purposes, than a shift in the population's characteristics that leaves that relationship intact, and this paper's framework treats only confirmed, sustained real drift as grounds for the personalized fine-tuning branch, rather than triggering it on a transient virtual-drift signal. Sousa, Gama, and Brandão [8] demonstrate that continuously updating a credit-risk model in response to ongoing monitoring maintains discriminatory accuracy more effectively than static periodic retraining, establishing the underlying value of adaptation this paper's framework extends specifically to the federated, multi-institution setting.

Fraud Detection and Credit-Risk Modeling Across Institutions

A closely related literature establishes the specific decisioning problems — fraud detection and credit-risk or distress prediction — this paper's framework is built to support in a federated, cross-institutional setting. Bhattacharyya, Jha, Tharakunnel, and Westland [9] and Ngai, Hu, Wong, Chen, and Sun [10] establish that machine-learning classifiers outperform manual review and rule-based approaches for financial fraud detection. Carcillo, Le Borgne, Caelen, Kessaci, Oblé, and Bontempi [11] show that combining unsupervised and supervised methods captures fraud patterns either approach alone would miss, while West and Bhattacharya's [12] review situates these findings across a wider range of fraud-detection methods. Lessmann, Baesens, Seow, and Thomas [13] provide large-scale benchmarking of credit-scoring classification methods, establishing that flexible, ensemble-based methods generally outperform traditional scorecards for the credit-risk and distress-prediction side of this paper's scope. This literature is generally conducted within a single institution's data; the case for pooling patterns across institutions federatively rests on the premise, well established elsewhere in this research line, that fraud and distress patterns visible at one institution are often too sparse in isolation to train the strongest possible model, making cross-institutional collaboration valuable precisely for the applications this literature studies.

Explainability and Fairness Obligations for Federated, Adapting Models

A final literature establishes obligations that persist, and in some respects intensify, once a model is both federated and adaptive. Kozodoi, Jacob, and Lessmann [14] show that fairness in credit scoring must be actively assessed and monitored rather than assumed to follow from strong predictive performance, a finding with particular force for a federated model because each participating institution's local population may differ in composition, and a personalized fine-tuning branch adapting to one institution's local conditions specifically could shift that institution's fairness properties in ways a federation-wide assessment would not catch. Gramespacher and Posth [15] show that explainable-AI techniques can meaningfully inform, not merely justify after the fact, the underlying lending or risk decision a portfolio manager makes, a role this paper's framework preserves by keeping every institution's decisioning model — whether the shared global

model or a personalized locally fine-tuned variant — subject to the same explainability requirements a single-institution model would face. Fuster, Goldsmith-Pinkham, Ramadorai, and Walther [16] provide empirical evidence that machine learning’s effect on credit-market outcomes is not uniformly positive across all borrower groups relative to traditional methods, a finding this paper’s discussion treats as a direct caveat on the personalized fine-tuning branch specifically, since an institution’s local adaptation could, in principle, improve its own aggregate performance while shifting outcomes unevenly across the borrower or transaction population it serves.

Read together, these five literatures describe a problem this paper is positioned to address directly: federated learning’s privacy-preserving architecture is well established, the general non-IID heterogeneity problem and the more specific problem of unevenly distributed concept drift across federated clients are each increasingly well characterized, the underlying concept-drift and adaptive-learning premises are well established for financial machine learning generally, fraud-detection and credit-risk modeling are well developed within single institutions and motivate cross-institutional collaboration, and the explainability and fairness obligations that bound what any adapting model can responsibly claim are well documented — but no paper in any of these literatures operationalizes a privacy-preserving, drift-aware selective aggregation mechanism, paired with an optional personalized local fine-tuning branch, specifically for a federation of financial institutions jointly training fraud-detection and distress-prediction models. This paper’s contribution is that operational synthesis: Federated Learning with Adaptive Resilience to drift Events, developed to let a financial federation’s shared global model remain robust to any one institution’s idiosyncratic drift while still giving that institution a path to adapt on its own terms.

Methodology

Research Approach: Design Science with Illustrative Evaluation

This paper follows a design-science research approach: it develops an artifact — Federated Learning with Adaptive Resilience to drift Events — intended to address a practical problem (protecting a financial federation’s shared global model from being degraded by any one participating institution’s currently unreliable, drifting local update, while giving that institution a privacy-preserving path to its own local adaptation), and it evaluates that artifact against the cumulative findings of the peer-reviewed literature rather than against a production federation, because no such deployment of this specific framework yet exists to study. Every quantitative comparison in the Results section is accordingly an illustrative construction, explicitly labeled as such, built to be directionally consistent with the cited literature’s findings rather than presented as a measured outcome.

Development of the Framework

The framework was developed by identifying the specific mechanism each relevant literature’s finding implies is missing from a standard federated pipeline, and adding that mechanism without weakening federated learning’s core privacy guarantee. The uniform-weighting assumption in McMahan et al.’s [1] federated averaging algorithm, and the heterogeneity-awareness FedProx [2] adds without itself being drift-specific, together suggested the missing element was a client-level drift status signal feeding into the server’s aggregation weighting decision. The characterization of federated drift as unevenly distributed across clients, established by Jothimurugesan et al. [3] and Canonaco et al. [4], informed the framework’s decision to detect and report drift per client rather than only at the federation-wide level, and to treat a client’s drift signal as a severity-graded status rather than a binary flag, so that aggregation weight can be reduced proportionally rather than only fully included or fully excluded. The real-versus-virtual distinction from Widmer and Kubat [7] informed the decision to gate the personalized fine-tuning branch on confirmed, sustained real drift specifically, rather than triggering it on a transient signal that a subsequent observation might reverse. The

explainability and fairness literature reviewed in the “Explainability and Fairness Obligations for Federated, Adapting Models” subsection informed the framework’s requirement that a personalized fine-tuning branch, once activated for an institution, remain subject to the same explainability and fairness review that institution’s model would require outside the federated setting, so that federation and adaptation together do not create a governance gap neither would create alone. Throughout, the framework was constrained to share only a drift status and severity signal between an institution and the central server, never raw data, gradients computed on drifted data, or any other information beyond what the aggregation-weighting decision strictly requires, preserving the privacy rationale that motivates federating in the first place.

Construction of Illustrative Evaluation Data

Because no production deployment exists to measure, the quantitative content of the Results section’s figures was constructed through explicit, reasoned estimation rather than invention. For each figure, construction began from a directional finding reported in one or more cited studies — McMahan et al.’s [1] and Li et al.’s [2] accounts of how aggregation weighting affects global model behavior under heterogeneity, or Canonaco et al.’s [4] demonstration that drift-aware adaptation preserves global model quality better than a drift-unaware baseline — and built numerical values directionally consistent with that finding at a plausible order of magnitude, without claiming a precision an illustrative exercise cannot support. Values were kept within ranges commonly discussed across the cited federated-learning and concept-drift literature rather than chosen to flatter the proposed framework. Every figure’s note names the specific source or sources informing its construction, so a reader can trace the reasoning rather than treat the numbers as free-standing.

Illustrative Case Scenario Approach

The “Illustrative Case Scenario: A Consortium Fraud-Detection Federation Adopting FLARE” subsection presents a single illustrative case scenario — a hypothetical federation of financial institutions jointly training a shared fraud-detection model and adopting FLARE — constructed to make the abstract architecture concrete rather than to report on an actual deployment. The scenario combines publicly discussed, general characteristics of federated learning practice with the technical mechanisms synthesized from the Literature Review section, to produce a plausible walk-through of how the framework’s local drift detection, drift-aware selective aggregation, and personalized fine-tuning branch would together respond when one institution in the federation begins experiencing severe local drift while the others do not. No identifying details of any real institution are used; the scenario is explicitly hypothetical and offered as a reasoning aid.

Assumptions and Scope Limitations

Several assumptions and scope limitations bound this study. First, the paper assumes each participating institution already operates, or can operate, a local concept-drift detection capability sufficient to produce a status and severity signal; it does not itself develop that underlying detection technology, which the literature reviewed in the “Concept Drift and Adaptive Learning in Financial Machine Learning” subsection addresses. Second, the paper assumes a federation’s participating institutions have already agreed to the federation’s basic architecture and communication protocol; it does not address the governance or contractual arrangements required to establish a financial federation in the first place. Third, the paper is scoped to the aggregation-weighting and personalization mechanisms themselves; it does not address the secure-aggregation or differential-privacy techniques that a production federated system would likely layer on top of the basic architecture described here for additional protection against inference attacks on shared updates, treating that additional protection as a complementary rather than competing concern. Fourth, as stated repeatedly above, all quantitative results in the Results section are illustrative estimates grounded in the literature rather than measurements, and conclusions about the framework’s plausibility should be read with that limitation firmly in mind.

Results

Architecture Overview: Federated Learning with Adaptive Resilience to drift Events

The architecture synthesized from the literature review organizes FLARE as a converging, client-to-server federation, rather than a single-owner technical pipeline, because its central claim is that the aggregation server's behavior must respond to signals originating independently at each participating institution. Several participating institutions sit at the base of the architecture, each running local model training on its own data and, in parallel, local concept-drift detection of the kind the literature reviewed in the "Concept Drift and Adaptive Learning in Financial Machine Learning" subsection establishes. Each institution transmits two things upward to the central aggregation server: its local model update, computed in the ordinary federated manner and containing no raw data, and a drift status and severity signal, likewise containing no raw data, indicating whether that institution's local detection has flagged confirmed, sustained real drift and, if so, how severe. The aggregation server applies a drift-aware selective aggregation rule to combine every institution's local update into an improved shared global model: institutions reporting no drift, or only mild, unconfirmed drift, are weighted in the aggregation largely as a standard federated averaging rule would weight them, while institutions reporting severe, unconfirmed drift are down-weighted or temporarily excluded from that communication round's aggregation, protecting the shared global model from a currently unreliable local signal. The server then broadcasts the updated shared global model back down to every participating institution, including any institution currently down-weighted or excluded, so that institution continues to benefit from the federation's collective learning even while its own update is temporarily discounted. For an institution whose drift signal is confirmed as severe and sustained across multiple communication rounds, rather than transient, the architecture offers an optional personalized local fine-tuning branch: that institution may fine-tune a local copy of the shared global model against its own recent data, producing a personalized variant that adapts to its changed local conditions without requiring the rest of the federation to adopt that adaptation, and without further destabilizing the aggregation the other institutions rely on. Every drift status signal, aggregation-weighting decision, and personalization activation is recorded in a federation-level drift-status log, both to support the server's ongoing calibration of its aggregation-weighting thresholds and to give the federation's participants a shared, auditable record of how the architecture has responded to drift over time.

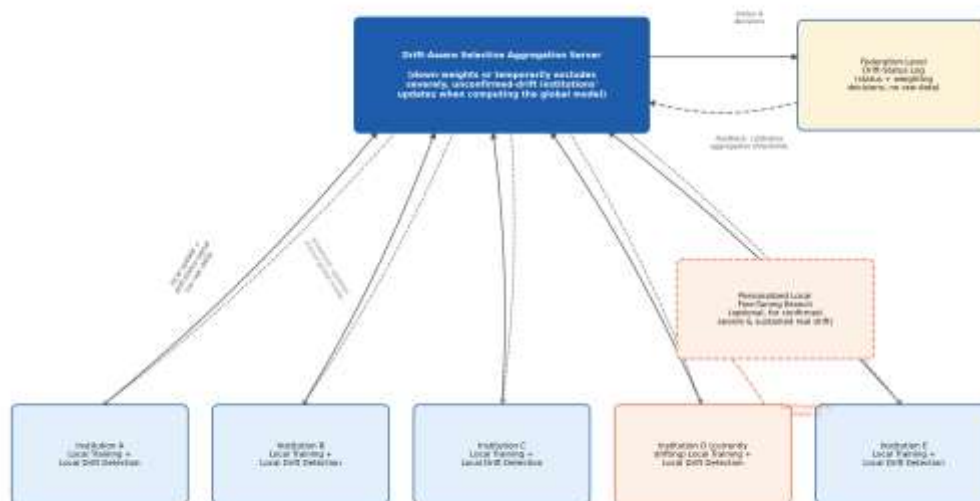


Figure 1. Federated Learning with Adaptive Resilience to drift Events (FLARE): a converging-tree federation architecture, showing multiple institutions' local training and drift-detection nodes converging into the central drift-aware selective aggregation server, the bidirectional global-model broadcast, the dashed personalized fine-tuning branch, and the drift-status-log feedback loop.

Table 1. Illustrative Capability Comparison Across Federated Aggregation Approaches

Dimension	Standard FedAvg (Uniform Aggregation)	Federated Learning with Manual Client Exclusion	FLARE
Global model robustness to one institution's severe local drift	Weak; a drifting institution's update is aggregated at the same weight as every other institution's	Moderate; protects the global model once a drifting institution is manually identified and excluded, but only after that manual review occurs	Strong; automated drift-aware selective aggregation down-weights or excludes a severely drifting institution's update without waiting for manual review
Preservation of non-drifting institutions' contributions	Strong by default, but only because drift is not yet accounted for at all	Strong once exclusion occurs, but non-drifting institutions are unprotected during the detection lag	Strong; automated, continuous drift-aware weighting minimizes the detection-lag window entirely
Responsiveness for a genuinely drifting institution	Weak; the drifting institution receives no adaptation path beyond ordinary federated updates	Weak; an excluded institution is cut off from federation-wide learning with no adaptation path offered in return	Strong; a confirmed-drifting institution is offered a personalized local fine-tuning branch on top of the shared model
Preservation of the federated privacy guarantee	Strong; no raw data is ever shared	Strong; no raw data is ever shared	Strong; only a drift status and severity signal is added, never raw data or drift diagnostics computed from it
Reliance on manual intervention	None required, but at the cost of no drift response at all	High; a human reviewer must identify and act on a drifting client before exclusion occurs	Low; the aggregation-weighting response is automated, with human review reserved for personalization activation decisions

Table 2. Mapping Client Drift Scenarios to FLARE's Aggregation and Personalization Response

Client Drift Scenario	Drift Status Signal	FLARE Response	Grounding Literature
No drift detected at any institution	All institutions report stable	Standard federated averaging aggregation, unchanged from baseline	McMahan et al. [1]
One institution reports mild, unconfirmed drift	Mild severity, unconfirmed	Modest down-weighting in aggregation; continued monitoring before further action	Li et al. [2]

Client Drift Scenario	Drift Status Signal	FLARE Response	Grounding Literature
One institution reports severe, unconfirmed drift (e.g., a sudden shift not yet observed across multiple rounds)	Severe severity, unconfirmed	Temporary exclusion from aggregation for the current round, pending confirmation	Jothimurugesan et al. [3]
One institution's severe drift is confirmed as sustained across multiple communication rounds and consistent with a genuine relationship change	Severe severity, confirmed, real drift	Personalized local fine-tuning branch offered; institution's contribution remains excluded from global aggregation until its local distribution restabilizes	Widmer and Kubat [7]; Canonaco et al. [4]
A previously drifting institution's local distribution restabilizes	Severity signal returns to stable	Institution's updates are gradually reintroduced into global aggregation at increasing weight	Canonaco et al. [4]

Illustrative Case Scenario: A Consortium Fraud-Detection Federation Adopting FLARE

Consider a hypothetical federation of six mid-size financial institutions that have agreed to jointly train a shared card-fraud-detection model using federated learning, previously operating under a standard federated averaging rule that aggregated every institution's local update at a weight proportional to its local transaction volume, with no drift-specific logic of any kind. In this hypothetical scenario, the federation adopts FLARE, and each institution begins running local drift detection alongside its ordinary local training. In the fourth month, one institution experiences a coordinated card-fraud campaign specifically targeting its own cardholder base, a pattern the other five institutions do not observe in their own data at all; that institution's local drift detection flags severe drift within days, and its drift status signal, shared with the aggregation server, triggers a temporary exclusion of its update from that round's aggregation. Over the following three weeks, the drift signal remains severe and is confirmed as sustained rather than transient, and the drifting institution is offered, and accepts, a personalized local fine-tuning branch, adapting its own copy of the shared global model to the specific fraud pattern it is experiencing while the other five institutions' aggregation continues undisturbed by that institution's temporarily excluded updates. In the ninth month, the coordinated fraud campaign subsides and the drifting institution's local distribution restabilizes; its drift status signal returns to stable, and the aggregation server begins gradually reintroducing its updates into the shared global model at increasing weight over the following several rounds, rather than immediately restoring full weight in a single step. Throughout this episode, the other five institutions' shared global model performance remains stable, protected from the sixth institution's temporarily volatile local signal, while the sixth institution itself is never left without an adaptation path of its own. The federation-level drift-status log records the full episode, giving the federation's participating institutions a shared, auditable account of how the architecture responded. This hypothetical sequence is offered to illustrate the framework's reasoning, not as a report of an actual federation's experience.

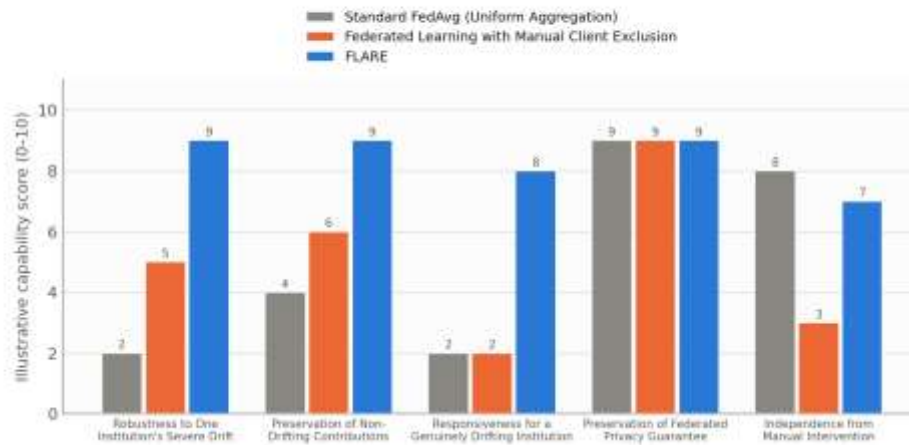


Figure 2. Illustrative Capability Scores (0–10 Scale) for Standard FedAvg, Federated Learning with Manual Client Exclusion, and FLARE Across Five Dimensions.

Note: These capability scores are illustrative estimates constructed for this paper's comparative framework rather than the product of a scored evaluation instrument; they were shaped by the aggregation-weighting behavior McMahan et al. [1] and Li et al. [2] describe and by the drift-aware adaptation benefit Canonaco et al. [4] document, and should be read as a reasoning aid rather than a benchmark. Figure 2 corresponds to the comparison presented in Table 1.

The pattern the figure is intended to convey mirrors the trade-off documented across the Literature Review section: standard FedAvg scores strongly only on privacy preservation and on requiring no manual intervention, because it does nothing at all in response to drift; manual client exclusion improves robustness and non-drifting-institution protection once a human reviewer acts, but scores poorly on responsiveness for the drifting institution itself and on independence from manual intervention, since a person must identify and act on the drift each time; FLARE is positioned, by design, to lead on robustness, non-drifting-institution protection, and responsiveness simultaneously, while matching both baselines on privacy preservation, because its automated drift-aware mechanisms were built specifically to close the gaps a uniform or manually gated approach leaves open.

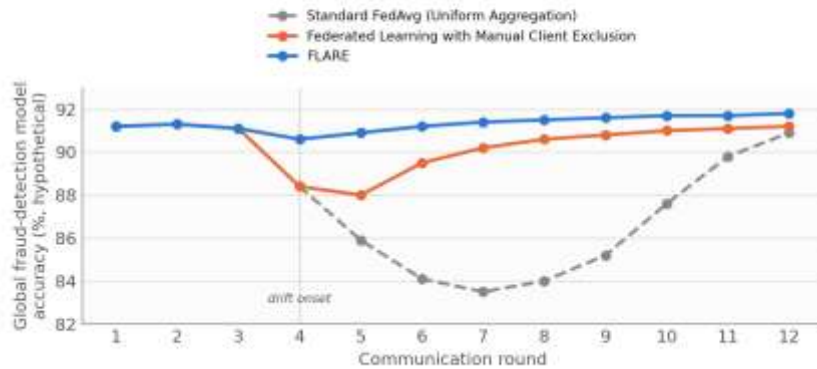


Figure 3. Hypothetical Global Fraud-Detection Model Accuracy Over Twelve Communication Rounds Following One Institution's Severe Drift Onset at Round 4, by Aggregation Approach.

Note: This is a hypothetical, illustrative trajectory constructed for this paper's argument rather than measured performance from a deployed federation; the shared dip beginning at round 4 for standard FedAvg and, more briefly, for manual exclusion reflects the detection-and-response lag each approach exhibits before the drifting institution's update is discounted, consistent with the aggregation-weighting dynamics McMahan et al. [1] and Li et al. [2] describe, while FLARE's smaller and more quickly recovered dip reflects its automated, continuous drift-aware down-weighting.

Figure 3 illustrates the argument that the practical cost of a drift-unaware or manually gated aggregation rule is measured not only in whether the shared global model eventually recovers — all three approaches in this hypothetical episode eventually return close to their pre-drift accuracy — but in the depth and duration of the degradation the federation's other, non-drifting institutions experience in the meantime, a cost FLARE's automated response is designed to minimize.

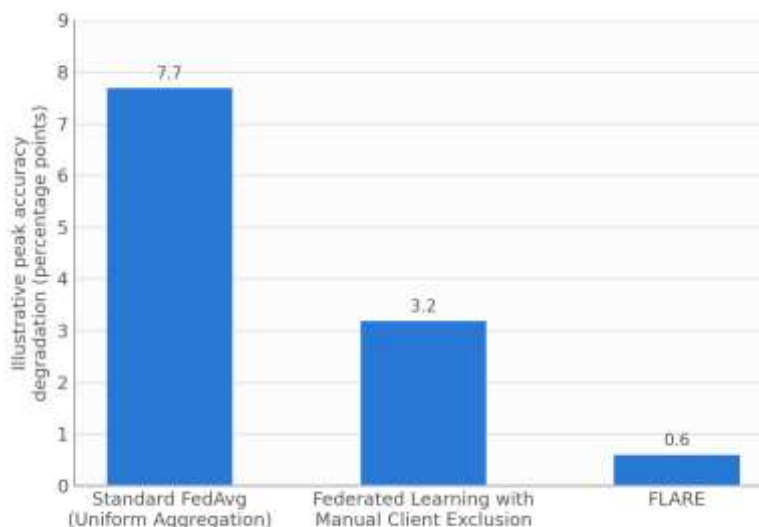


Figure 4. Illustrative Peak Global Fraud-Detection Model Accuracy Degradation (Percentage Points) Caused by One Severely Drifting Institution, by Aggregation Approach.

Note: This is an illustrative construction rather than a measured benchmark; the ordering reflects the same detection-and-response-lag reasoning underlying Figure 3, with FLARE’s automated, continuous drift-aware weighting minimizing the peak degradation the federation’s non-drifting institutions experience, consistent with the drift-aware adaptation benefit Canonaco et al. [4] document.

Figure 4 restates Figure 3’s trajectory as a single peak-impact comparison, making explicit the practical stakes of the detection-and-response lag: a federation relying on standard FedAvg absorbs the largest hit to shared model quality from one institution’s drift event, a federation relying on manual exclusion absorbs a smaller but still material hit while a human reviewer identifies and acts on the drift, and FLARE’s automated response keeps the hit closest to negligible.

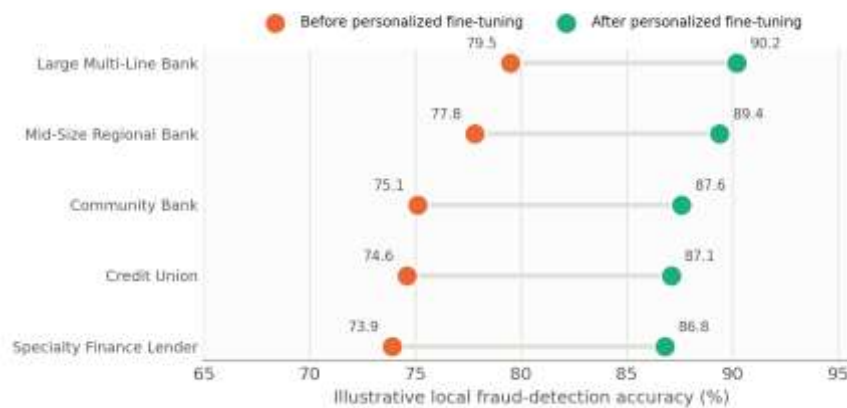


Figure 5. Illustrative Local Fraud-Detection Accuracy for a Confirmed-Drifting Institution, Before and After Activating FLARE’s Personalized Local Fine-Tuning Branch, by Institution Profile.

Source: These accuracy figures are illustrative constructions, not measurements from a deployed federation; every profile shows substantial recovery after personalization because the branch is, by design in this framework, offered only once drift is confirmed as severe and sustained, and the broadly similar magnitude of recovery across profiles reflects an illustrative assumption that the personalization mechanism itself does not systematically favor larger or smaller institutions, an assumption the “Limitations” subsection identifies as an open empirical question rather than a demonstrated property.

The figure illustrates the argument at the center of this paper’s second contribution: an institution left to rely solely on an aging, pre-drift shared global model, or excluded from the federation with no adaptation path at all, continues operating on a model poorly matched to its current conditions, while a personalized local fine-tuning branch offers that institution a way to recover local accuracy without either abandoning the federation or waiting for the rest of the federation to catch up to its own idiosyncratic drift.

Discussion

Implications for Practice

The framework developed here suggests several practical implications for financial institutions participating in cross-institutional federated learning. Perhaps most important is that a federation’s aggregation rule is itself a design choice with real consequences, not a neutral default: the Results section’s figures illustrate that

a uniform-weighting aggregation rule, however reasonable it appears when every client's data is stable, exposes the entire federation to any one member's future instability, so institutions considering or already participating in a federation should treat drift-awareness in the aggregation mechanism as a design requirement rather than an optional enhancement. A second implication concerns the value of separating protection for the federation from adaptation for the drifting institution: Table 1 shows that manual exclusion protects the federation but abandons the drifting institution, while FLARE's personalized fine-tuning branch shows that these two goals need not trade off against each other if the framework is designed to pursue both explicitly. A third implication is organizational: because FLARE's drift-status signal is shared across the federation's central server, participating institutions should expect the framework to surface, in a privacy-preserving way, information about which institutions are currently experiencing unusual conditions, which may itself be operationally useful to a federation's governance beyond the specific aggregation-weighting decision it informs.

Future Research

Several directions for future empirical research follow from these limitations. Empirical work studying how financial institutions currently participating in federated learning arrangements handle client-level drift in practice, prior to adopting any formal drift-aware aggregation framework, would establish a baseline against which a framework like FLARE's practical benefit could be assessed. Research developing and empirically calibrating the severity thresholds that determine when an institution's update is down-weighted, excluded, or offered personalized fine-tuning would address this paper's most significant open operational question. Empirical study of whether a personalized local fine-tuning branch, once activated, produces uneven fairness effects across a drifting institution's own served population — directly testing the caveat the "Limitations" subsection raises but does not resolve would meaningfully extend this paper's treatment of the personalization mechanism specifically. Finally, empirical or security-focused research examining what, if anything, a minimal drift-status signal reveals about a participating institution's conditions to the central server or to other federation members, and whether additional cryptographic protection of that signal is warranted, would directly address the privacy-preservation caveat this paper raises but does not itself resolve with a technical solution [27-29].

Conclusion

This paper has argued that financial institutions participating in cross-institutional federated learning for fraud detection and financial-distress prediction face a robustness problem that the substantial existing literature on federated learning, non-IID heterogeneity, and concept drift has each partially addressed but not operationally resolved for this specific setting: a standard federated aggregation rule has no mechanism to protect a shared global model from any one participating institution's currently unreliable, drifting local update, and no mechanism to help that drifting institution adapt without either waiting indefinitely or abandoning the federation. Building on the foundational federated learning architecture and its privacy-preserving rationale [1], [2], the characterization of concept drift as unevenly distributed across federated clients [3], [4], the real-versus-virtual drift distinction and adaptive-learning premises established for financial machine learning generally [7], [5], [6], [8], the fraud-detection and credit-risk modeling literature that motivates cross-institutional collaboration in the first place [9], [11], [13], and the explainability and fairness obligations that bound what a federated, adapting model can responsibly claim [14], [15], [16], this paper has developed Federated Learning with Adaptive Resilience to drift Events as a framework that adds privacy-preserving, per-client drift detection, drift-aware selective aggregation, and an optional personalized local fine-tuning branch to a standard federated learning pipeline, without weakening the core privacy guarantee that makes federated learning valuable for financial institutions in the first place. The framework was illustrated, not validated, through two comparative tables, a hypothetical case scenario, and four

illustrative figures, with every quantitative value explicitly and repeatedly labeled as an estimate rather than measured data. Future empirical research calibrating the framework's severity thresholds against real multi-institution federated telemetry, and directly testing the personalized fine-tuning branch's fairness properties, would be necessary before a federation could adopt this architecture with confidence in its specific parameters, but the underlying argument — that a financial federation's shared model can be made resilient to any one member's idiosyncratic drift, and that member can be given its own path to adapt, without either goal requiring the federation to compromise the privacy guarantee that justifies federating in the first place rests on a body of peer-reviewed federated-learning and concept-drift evidence this paper has, for the first time in this series, applied specifically to the problem of drift heterogeneity across federated financial institutions rather than to a single institution's own adaptive learning pipeline.

References

- [1] McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics* (Vol. 54, pp. 1273–1282). PMLR.
- [2] Li, T., Sahu, A. K., Zaheer, M., Sanjabi, M., Talwalkar, A., & Smith, V. (2020). Federated optimization in heterogeneous networks. *Proceedings of Machine Learning and Systems*, 2, 429–450.
- [3] Jothimurugesan, E., Hsieh, K., Wang, J., Joshi, G., & Gibbons, P. B. (2023). Federated learning under distributed concept drift. *Proceedings of Machine Learning Research*, 206, 5834–5853.
- [4] Canonaco, G., Bergamasco, A., Mongelluzzo, A., & Roveri, M. (2021). Adaptive federated learning in presence of concept drift. In *2021 International Joint Conference on Neural Networks (IJCNN)* (pp. 1–7). IEEE.
- [5] Karim, N. B., Tolosana, D., & Simko, R. (2025b). Responsible and Legally Compliant Enterprise Artificial Intelligence in Small-Business Lending: Governance, Explainability, and Fair-Lending Safeguards for Early-Intervention Recommendation Engines. *International Journal of Modern Computing*, 8(1), 55-66.
- [6] Shahiduzzaman, M., Sourav, M. S. A., Kabtia, M., Shammah, R. S., Islam, M. A., Chakraborty, P., ... & Adnan, M. (2026). Attention-guided quantum-inspired extreme learning machine with grey wolf optimization for rare-attack detection in IIoT networks. *Discover Artificial Intelligence*.
- [7] Hand, D. J. (2006). Classifier technology and the illusion of progress. *Statistical Science*, 21(1), 1–14.
- [8] Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM Computing Surveys*, 46(4), 1–37.
- [9] Widmer, G., & Kubat, M. (1996). Learning in the presence of concept drift and hidden contexts. *Machine Learning*, 23(1), 69–101.
- [10] Sousa, M. R., Gama, J., & Brandão, E. (2016). A new dynamic modeling framework for credit risk assessment. *Expert Systems with Applications*, 45, 341–351.
- [11] Siddiqui, M. I. H., Rahman, M. S., Kabir, A. A., Mahmud, F. U., Rashid, S. U., & Shammah, R. S. (2023). Comparative analysis of explainable machine learning models for cancer classification using cytological features. *Journal of Medical and Health Studies*, 4(5), 110-150.

- [12] Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. (2011). Data mining for credit card fraud: A comparative study. *Decision Support Systems*, 50(3), 602–613.
- [13] Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*, 50(3), 559–569.
- [14] Ahmed, K. R., Sufian, M. A., Faruq, O., Kowser, S., Shammah, R. S., & Ahmmed, M. R. (2025, July). Impact of IT product innovation on competitive advantage in the digital economy. In *2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN)* (pp. 1-6). IEEE.
- [15] Carcillo, F., Le Borgne, Y.-A., Caelen, O., Kessaci, Y., Oblé, F., & Bontempi, G. (2021). Combining unsupervised and supervised learning in credit card fraud detection. *Information Sciences*, 557, 317–331.
- [16] Potharaju, B., & Ravipudi, J. (2025). Automated Data Validation and Resilient Infrastructure Design for Enterprise Regulatory Reporting Platforms. *International Journal of Computer Science and Engineering Innovations*, 1(1), 46-64.
- [17] Rahman, M. S., Kabir, A. A., Reon, M. A. H., & Rashid, S. U. (2025). From Financial Transactions to Clinical Records: A Unified Machine Learning Methodology for Real-Time Anomaly Detection and Institutional Decision Support In High-Stakes Domains. *International Journal of Information Technology and Management Information Systems (IJITMIS)*, 16(6), 23-42.
- [18] West, J., & Bhattacharya, M. (2016). Intelligent financial fraud detection: A comprehensive review. *Computers & Security*, 57, 47–66.
- [19] Lessmann, S., Baesens, B., Seow, H.-V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, 247(1), 124–136.
- [20] Kozodoi, N., Jacob, J., & Lessmann, S. (2022). Fairness in credit scoring: Assessment, implementation and profit implications. *European Journal of Operational Research*, 297(3), 1083–1094.
- [21] Rahman, M. S., Islam, C. S., & Mudryi, M. (2025). Interoperability Standards for Real-Time Fraud-Signal Exchange Across US Financial Institutions: Design Principles and a Reference Implementation. *The Computertech*, 47-60.
- [22] Rahman, M. S., Islam, C. S., Mudryi, M., & Cakmak, S. (2025). Concept-Drift-Resilient Ensemble Learning for Financial Transaction Fraud Detection Under Evolving Criminal Typologies. *International Journal of Acta Informatica*, 4(1), 52-67.
- [23] Gramespacher, T., & Posth, J.-A. (2021). Employing explainable AI to optimize the return target function of a loan portfolio. *Frontiers in Artificial Intelligence*, 4, Article 693022.
- [24] Fuster, A., Goldsmith-Pinkham, P., Ramadorai, T., & Walther, A. (2022). Predictably unequal? The effects of machine learning on credit markets. *Journal of Finance*, 77(1), 5–47.
- [25] Potharaju, B., & Ravipudi, J. (2023). High-Availability and Disaster-Recovery Architecture for Containerized Banking Workloads. *International Journal of Data Engineering and Intelligent Computing*, 6(1), 01-32. Potharaju & Ravipudi, 2023).

- [26] Rashid, S. U., & Sattiraju, S. R. (2025b). Multi-Institution Validation of Drift-Resilient Early-Warning Models for SME Loan Distress and Transaction Fraud: Evidence from US Community Financial Institution Pilots. *International Journal of Modern Computing*, 8(1), 40-54.
- [27] Rahman, M. S., Marcucci, M., & Møller, C. (2025). Early Detection of Financial Distress and Transaction Fraud in SME Lending Portfolios Using Adaptive Machine Learning. *The Metascience*, 3(4), 35-52.
- [28] Karim, N. B., & Kovac, T. (2025). Cloud-Based Big-Data Architectures for AI-Driven Financial Planning and Analysis and Early Risk Signaling in Resource-Constrained Community Financial Institutions. *The Metascience*, 3(3), 29-40.