

Federated Learning For Fraud Detection: A Review

Shivam Babasaheb Rakh¹, Prof. A. P. Banger², Prof. S. B. Bhosale³, Prof. V. A. Karad⁴
Research Scholar¹, Professor^{2,3,4}

Department of Computer [JCOE]
Jaihind College of Engineering, Kuran, Maharashtra, India

Abstract- Fraud detection in banking and payments depends on learning from transaction data that institutions are often unable or unwilling to pool, because of privacy, regulatory, competitive, and security constraints. Federated learning (FL) offers an alternative in which institutions train a shared model by exchanging model updates rather than raw records. This review synthesizes selected scholarly works on FL for fraud detection, covering credit-card and payment fraud, financial-crime detection with graph learning, aggregation methods, non-IID data, class imbalance, differential privacy, secure aggregation, poisoning attacks, communication overhead, concept drift, and real-world deployment. Fraud-specific federated studies are compared with foundational FL, privacy, security, and fraud-detection literature. The literature suggests that FL is a plausible route to collaborative fraud detection, but that its benefits in the fraud setting depend on how well it handles extreme class imbalance, heterogeneous institutional data, evolving fraud patterns, and layered privacy and security protections. Many enabling results were obtained in non-fraud settings and should not be assumed to transfer unchanged. Because the review is based on a selected set of publications rather than an exhaustive database search, it is presented as a focused narrative review and not as a systematic review.

Keywords— Federated learning, fraud detection, credit-card fraud, privacy-preserving machine learning, differential privacy, secure aggregation, non-IID data, class imbalance, poisoning attacks, concept drift.

I. INTRODUCTION

Fraud detection is the task of identifying transactions, accounts, or behaviours that are deceptive or unauthorised. In financial services it spans credit-card and payment fraud, insurance and securities fraud, and money-laundering-related activity, and it has long been approached through data mining and machine learning (Ngai et al., 2011; West & Bhattacharya, 2016; Abdallah et al., 2016). More recent reviews continue to treat financial fraud as a setting in which anomaly detection, supervised learning, and their combinations are applied to large and highly imbalanced data (Hilal et al., 2022).

A recurring difficulty is that the data most useful for detecting fraud are distributed across organisations. A single bank observes only its own customers and merchants, whereas a fraudster may probe several institutions. Pooling transaction records in one place is often impractical because of data-protection requirements, competitive concerns, and security risk. Federated learning (FL)

addresses this tension by training a global model across participants who keep their data locally and share only model updates, as introduced with the Federated Averaging algorithm by McMahan et al. (2017). Conceptual treatments of federated machine learning describe horizontal, vertical, and transfer variants and position FL as a privacy-oriented approach to data silos (Yang et al., 2019a). For fraud detection among banks, the relevant setting is typically cross-silo FL, with a small number of reliable institutions, in contrast to the cross-device setting involving very many unreliable clients. FL does not remove the difficulties of fraud detection and introduces new ones. Fraud labels are extremely scarce relative to legitimate transactions and are often delayed (Dal Pozzolo et al., 2018).

Institutions differ in customers, products, and fraud patterns, so local data are not independent and identically distributed (non-IID) (Zhao et al., 2018; Hsieh et al., 2020). Shared updates can still leak information (Zhu et al., 2019; Nasr et al., 2019) and can be manipulated by malicious participants (Bagdasaryan et al., 2020; Bhagoji et

al., 2019). Fraud behaviour also changes over time as attackers adapt (Gama et al., 2014; Dal Pozzolo et al., 2018). The limitations of FL for fraud detection are therefore statistical, security-related, operational, and organisational, not only technical. The selected literature approaches these issues from different directions. Fraud-specific federated studies are represented by Yang et al. (2019b), Zheng et al. (2020), Suzumura et al. (2019), and Awosika et al. (2024). Foundational work on aggregation, heterogeneity, privacy, security, and communication supplies the methods these studies build on, and fraud-detection literature outside FL, notably Dal Pozzolo et al. (2018), supplies realistic problem formulations.

The objectives of this review are to: (i) summarize the selected studies and their reported contributions; (ii) compare their methods, scope, and boundaries; (iii) identify common challenges and potential gaps suggested

II. REVIEW METHODOLOGY AND LITERATURE SELECTION CRITERIA

This paper uses a focused narrative review of a selected set of published works on federated learning and fraud detection. The works were chosen for direct relevance to the topics listed in the title and abstract rather than through a documented, exhaustive search. They fall into two groups: (a) a small set of studies that apply FL, or closely related collaborative learning, to fraud or financial-crime detection; and (b) foundational and enabling studies on FL algorithms, non-IID data, privacy, security, communication efficiency, class imbalance, and concept drift, including surveys and non-federated fraud-detection papers that define the problem setting.

The selection criteria were: (1) direct relevance to FL, privacy-preserving learning, or fraud and financial-crime detection; (2) an identifiable research question, algorithm, analysis, or survey objective; (3) publication in a scholarly journal, conference, or book series, or, for a small number of widely cited works, availability as an archived preprint (these are labelled as preprints in the

reference list); and (4) contribution to at least one of the review themes. No papers were added beyond the set cited in this review.

Each study is described according to its research focus, method, principal reported findings, and boundaries. Findings are attributed to the original authors and are reported at a qualitative level. Numerical results are intentionally not reproduced, because the studies use different datasets, partitioning schemes, and metrics and are not directly comparable. This review did not re-run any experiments. Statements that go beyond what a cited study reports are identified as the present review's own observations. Where the fraud-specific federated literature in the selected set is thin, the review says so and relies on non-fraud FL studies, whose results should not be assumed to hold for fraud data.

Because the work is based on a selected set of publications and does not document a comprehensive database search, search strings, or formal screening stages, it is not presented as a systematic review. No exhaustive-coverage or PRISMA claim is made. Author, year, venue, and DOI details in the reference list should be checked against publisher or preprint-server records before submission.

III. LITERATURE REVIEW

Fraud detection and the data-sharing problem

Early and survey literature organises financial fraud detection by fraud type and by technique. Ngai et al. (2011) present a classification framework and academic review of data-mining techniques applied to financial fraud, and West and Bhattacharya (2016) review computational-intelligence and statistical approaches. Abdallah et al. (2016) survey fraud detection systems more broadly, and Hilal et al. (2022) review anomaly-detection techniques and more recent advances for financial fraud. These surveys document a common problem structure: few fraudulent cases, evolving behaviour, and a need for timely detection. They are not federated studies, and their relevance here is that they define what a

federated fraud model must still solve. As a boundary, survey coverage reflects the periods and databases each survey searched.

Federated learning and privacy-preserving machine learning

McMahan et al. (2017) introduced Federated Averaging, in which clients perform local training and a server averages the resulting models, and evaluated it on image and language tasks. Yang et al. (2019a) give a conceptual overview of federated machine learning and its categories, while Li et al. (2020a) and Kairouz et al. (2021) survey open problems, including heterogeneity, privacy, robustness, and efficiency. Li et al. (2023) survey FL from a systems perspective. Bonawitz et al. (2019) describe a production-scale system design built for mobile devices. Together these works establish FL as a privacy-oriented training paradigm; they also stress that keeping data local is not by itself a complete privacy guarantee, which motivates the privacy mechanisms in Sections 3.8 and 3.9. Most of this foundational work was not developed for fraud data.

Federated learning for credit-card and payment fraud

Yang et al. (2019b) propose FFD, a federated-learning-based method for credit-card fraud detection in which multiple institutions collaboratively train a shared detection model without exchanging their transaction data. The authors report, from their own experiments, that the federated approach is effective relative to alternatives trained on limited local data.

Zheng et al. (2020) propose federated meta-learning for fraudulent credit-card detection, motivated by the observation that different banks hold data with different characteristics, and report improved detection in their experiments. Both papers support the feasibility of collaborative card-fraud models. Both results are bounded by the datasets, institutional partitions, and metrics chosen by the authors, and readers should consult the original papers for experimental configurations before generalising.

Federated graph learning, financial crime, and explainability

Fraud and money-laundering signals are often relational, appearing in networks of accounts and transfers. Suzumura et al. (2019) discuss federated graph learning for collaborative financial-crimes detection across institutions, an early-stage, workshop-level contribution that extends FL beyond tabular transaction features. Awosika et al. (2024) examine the role of explainable AI together with FL in financial fraud detection, addressing the need for transparency alongside privacy. These works broaden the discussion from model accuracy to relational data and to the interpretability expected in regulated settings. Their evidence is narrower than that of the foundational literature, and graph-based and explainable federated fraud systems remain relatively less studied in the selected set.

Aggregation methods and heterogeneous optimisation

FedAvg (McMahan et al., 2017) is the usual baseline aggregation rule. Later work targets its weaknesses under heterogeneity. Li et al. (2020b) propose FedProx, which adds a proximal term to local objectives and tolerates variable amounts of local work. Karimireddy et al. (2020) analyse client drift and propose SCAFFOLD, which uses control variates to correct local updates. Wang et al. (2020) identify an objective-inconsistency problem when clients perform different numbers of local steps and propose a normalised averaging scheme. These studies were evaluated mostly on standard benchmarks, not on fraud data. Their relevance to fraud is that institutions differ in data volume and compute, so the choice of aggregation rule is likely to matter; this is the present review's inference, not a finding of those papers.

Non-IID data across institutions

Zhao et al. (2018) show that highly skewed, non-IID local label distributions can substantially degrade FedAvg accuracy, relate this to weight divergence, and propose sharing a small globally available data set to mitigate the effect. Hsieh et al. (2020) study non-IID data in decentralised

learning more broadly, report that the degree of skew matters, and propose a mitigation. Kairouz et al. (2021) discuss heterogeneity as a central open problem. For fraud detection, institutions differ in customer mix, merchant types, geography, and fraud typologies, so non-IID data should be expected. The proposal of sharing a global data subset may be unsuitable where data cannot be shared, and the benchmarks used in these studies are not fraud datasets.

Class imbalance

Fraud is a rare-event problem. Dal Pozzolo et al. (2015) show that undersampling, a common remedy for imbalance, changes the class distribution and can bias estimated probabilities, and propose a calibration correction. Chawla et al. (2002) introduce SMOTE, an oversampling method that synthesises minority-class examples. Carcillo et al. (2021) combine unsupervised and supervised learning for credit-card fraud detection using real transaction data. Dal Pozzolo et al. (2018) treat imbalance together with drift and delayed labels in a realistic fraud-detection setting. These studies are centralised. In FL, each institution may see a different fraud rate, and rebalancing performed locally may interact with aggregation in ways these papers do not examine; this interaction is identified here as a gap and not as a finding of the cited work.

Differential privacy

Differential privacy provides a formal, quantifiable privacy guarantee (Dwork & Roth, 2014). Abadi et al. (2016) introduced differentially private stochastic gradient descent with a privacy-accounting method for deep learning. McMahan et al. (2018) and Geyer et al. (2017) adapt differential privacy to federated training at the user or client level, reporting that the accuracy cost can be small when many participants are involved. Wei et al. (2020) analyse FL with differential privacy, add noise before aggregation, and derive a convergence bound that exposes a trade-off between privacy and convergence performance, including the effect of the number of aggregation rounds. The review's observation is that the favourable results with many clients may

not carry over to cross-silo fraud detection, where only a handful of institutions participate and each contribution is large; this should be tested rather than assumed. Gradient-leakage and membership-inference studies (Zhu et al., 2019; Nasr et al., 2019) show why protection beyond data locality is considered necessary.

Secure aggregation

Bonawitz et al. (2017) propose a practical secure aggregation protocol in which the server learns the sum of client updates without seeing individual updates, using masking and secret sharing and tolerating participants that drop out. Secure aggregation reduces exposure of individual updates to the aggregator but does not by itself provide differential privacy for the aggregated result, and it does not defend against malicious updates. The two are therefore commonly considered complementary. The protocol was designed with large-scale, cross-device settings in mind, and its cost and suitability in a small consortium of banks are practical questions not settled by the original paper.

Poisoning and Byzantine attacks

Blanchard et al. (2017) and Yin et al. (2018) study Byzantine-tolerant aggregation, introducing robust rules for combining gradients. Bhagoji et al. (2019) show that model-poisoning attacks can succeed in FL while remaining difficult to detect, and Bagdasaryan et al. (2020) show that a participant can implant a backdoor through model replacement and discuss that secure aggregation can make anomaly detection of updates harder. For fraud detection, a malicious or compromised participant could attempt to make the global model overlook certain fraud patterns; this scenario is the present review's reading of the cited attacks and has not, to the extent covered by this selection, been evaluated on fraud data. Robust aggregation rules are typically analysed under assumptions about honest participants, and their behaviour under strongly non-IID institutional data requires care.

Communication overhead

Konečný et al. (2016) propose structured and sketched updates to reduce uplink communication in FL. Sattler et al. (2020) propose sparse ternary compression designed to remain effective with non-IID data. Bonawitz et al. (2019) report system-level design considerations for scale. In cross-silo fraud detection, a handful of institutions can often rely on better connectivity than mobile devices, so communication may be less limiting than in cross-device settings, although large models, frequent retraining, and added cryptographic or noise-related overhead can still matter. This is an observation of the present review and depends on deployment.

Concept drift and delayed feedback

Gama et al. (2014) and Lu et al. (2019) survey concept drift and its detection and adaptation. Dal Pozzolo et al. (2018) formalise a realistic card-fraud-detection workflow in which investigators confirm only a limited number of alerts, so that fresh feedback is scarce and other labels arrive with delay, and propose a learning strategy suited to this structure and to class imbalance and drift. These works are centralised. The selected fraud-specific federated studies do not centre on drift, so how institutions should coordinate model updates, retraining schedules, and label delay in a federation is an open question.

Real-world deployment

Deployment evidence in the selected literature is limited. Bonawitz et al. (2019) document a production FL system for mobile devices, and Kairouz et al. (2021) and Li et al. (2023) discuss practical, systems, and governance challenges. Awosika et al. (2024) emphasise explainability and privacy, which are central to regulated financial deployments.

Federated fraud detection among banks would additionally involve legal agreements, data-schema alignment, feature harmonisation, model governance, auditability, and integration with real-time scoring and investigation workflows. These requirements are generally recognised, but the selected fraud-specific studies do not report

production deployments, and this review does not claim that any exist.

IV. COMPARATIVE ANALYSIS OF EXISTING STUDIES

The selected studies address connected problems with different methods and units of analysis, and their findings are complementary rather than directly interchangeable. Table 1 compares the fraud-related studies, and Table 2 maps enabling techniques to the fraud-detection challenges they address.

Table 1. Comparative summary of fraud-related studies

Study	Method / scope	Main contribution and boundary
Yang et al. (2019b)	Federated framework (FFD) for credit-card fraud detection across institutions	Shows a federated route to collaborative card-fraud models; results are author-reported under their chosen data and partitioning.
Zheng et al. (2020)	Federated meta-learning for credit-card fraud detection	Targets differences among banks' data using meta-learning; boundary set by the datasets and baselines used.
Suzumura et al. (2019)	Federated graph learning for financial-crimes detection (workshop-stage preprint)	Extends FL to relational financial data across institutions; early-stage, not a full production evaluation.
Study	Method / scope	Main contribution and boundary
Awosika et al. (2024)	Explainable AI combined with FL in financial fraud detection	Highlights transparency and privacy together; scope and evidence narrower than the foundational FL literature.
Dal Pozzolo et al. (2018)	Centralised, realistic card-fraud learning strategy (non-	Models imbalance, drift and delayed feedback; does not address multi-institution privacy.

	federated)	
--	------------	--

Table 2. Enabling techniques and their boundaries for fraud detection

Theme	Representative studies	Boundary for fraud use
Aggregation	McMahan et al. (2017) (FedAvg); Li et al. (2020b) (FedProx); Karimireddy et al. (2020) (SCAFFOLD); Wang et al. (2020)	Evaluated mostly on non-fraud benchmarks; suitability for fraud data untested in this set.
Non-IID data	Zhao et al. (2018); Hsieh et al. (2020)	Skew analysed on standard tasks; real inter-bank heterogeneity differs.
Class imbalance	Dal Pozzolo et al. (2015); Chawla et al. (2002); Carcillo et al. (2021)	Centralised; interaction with federated aggregation not studied here.
Differential privacy	Abadi et al. (2016); McMahan et al. (2018); Geyer et al. (2017); Wei et al. (2020)	Favourable results often assume many clients; cross-silo effect to be tested.
Secure aggregation	Bonawitz et al. (2017)	Hides individual updates; no protection against malicious updates by itself.
Poisoning / robustness	Blanchard et al. (2017); Yin et al. (2018); Bhagoji et al. (2019); Bagdasaryan et al. (2020)	Not evaluated on fraud data in this set; assumptions on honest participants.
Comm. overhead	Konečný et al. (2016); Sattler et al. (2020)	Cross-silo links may be less constrained than cross-device ones.
Concept drift	Gama et al. (2014); Lu et al. (2019); Dal Pozzolo et al. (2018)	Centralised treatments; federated drift coordination open.

A first theme is the difference between collaboration and accuracy gains: the fraud-specific federated studies indicate that sharing learning across institutions is feasible, but their evidence is bounded by experimental choices (Yang et al., 2019b; Zheng et al., 2020). A second theme is the layering of protections. Differential privacy, secure aggregation, and robust aggregation address different threats, and none is sufficient alone (Bonawitz et al., 2017; Wei et al., 2020; Blanchard et al., 2017).

A third theme is heterogeneity: non-IID data and imbalance interact, and the interaction is most

acute when institutions have different fraud rates (Zhao et al., 2018; Dal Pozzolo et al., 2015). A fourth theme is time: fraud patterns change and labels arrive late (Dal Pozzolo et al., 2018), yet federated studies in the selected set give this limited attention. Finally, deployment questions, including explainability and governance, remain largely outside experimental papers (Awosika et al., 2024; Kairouz et al., 2021).

The methods also determine what each study can establish. Algorithmic papers demonstrate behaviour under defined experimental conditions; theoretical analyses give guarantees under assumptions; surveys map open problems; and attack papers show what is possible for an adversary. No single type of study covers all dimensions of federated fraud detection.

V. RESEARCH CHALLENGES AND IDENTIFIED GAPS

The following are potential gaps suggested by the selected papers. They should not be interpreted as proof that no other research has addressed these topics.

Realistic data and benchmarks

Publicly available fraud datasets are generally single-source and often anonymised, which limits how faithfully a multi-institution federation can be simulated by partitioning one dataset. Studies could document how institutional partitions are created and whether they reflect realistic differences in fraud rates, customer mix, and feature schemas.

Joint treatment of non-IID data and class imbalance

Non-IID data (Zhao et al., 2018; Hsieh et al., 2020) and class imbalance (Dal Pozzolo et al., 2015) have been studied largely in isolation. Further research could examine how local rebalancing, calibration, and aggregation interact when each institution has a different and very small fraud rate.

Privacy-utility trade-offs in small federations

Differential-privacy results reported with many participants (McMahan et al., 2018; Geyer et al., 2017) may not carry over to a consortium of a few institutions. Evaluations should report the privacy budget, the number of participants, and fraud-specific metrics such as precision at a fixed alert volume.

Robustness against malicious or compromised participants

Poisoning and backdoor studies (Bhagoji et al., 2019; Bagdasaryan et al., 2020) were not designed around fraud data, and the tension between secure aggregation and update inspection is noted in the attack literature (Bagdasaryan et al., 2020). Fraud-specific threat models, including participants who wish to hide particular fraud patterns, are an open area.

Concept drift and delayed labels in federations

Drift and verification latency are central to card fraud (Dal Pozzolo et al., 2018; Gama et al., 2014), but coordinating drift detection and retraining across institutions with different label delays is not addressed in the selected fraud-specific federated studies.

Evaluation, reproducibility, and fair comparison

Federated fraud studies use different datasets, partitions, and metrics, which makes comparison difficult. Studies should document model versions, data preparation, partitioning, hyperparameters, privacy settings, and evaluation metrics suitable for rare events.

Deployment, explainability, and governance

Limited evidence exists on production federated fraud systems. Open questions include explainability for investigators and regulators (Awosika et al., 2024), schema alignment, auditability, incentives for participation, and legal arrangements for collaborative model training. Legal and regulatory conclusions should be drawn with reference to the applicable law.

VI. PROPOSED RESEARCH DIRECTIONS

The following directions are proposals derived from the gaps above. They are not results of this review, and none has been tested here.

Benchmark protocols for federated fraud detection: define shared guidelines for building multi-institution partitions, reporting imbalance, and choosing rare-event metrics so that studies can be compared.

Imbalance-aware federated training: evaluate local resampling, cost-sensitive learning, and probability calibration together with different aggregation rules under differing local fraud rates.

Personalised and meta-learning approaches: extend meta-learning and personalisation ideas (Zheng et al., 2020) to compare global, personalised, and locally trained models under institutional heterogeneity.

Drift-aware federated learning: design and test coordinated drift detection, retraining schedules, and handling of delayed feedback (Dal Pozzolo et al., 2018; Lu et al., 2019) across institutions.

Layered privacy and security evaluation: study differential privacy, secure aggregation, and robust aggregation together, reporting their combined effect on detection quality, cost, and attack resistance.

Fraud-specific adversarial testing: develop threat models for participants who attempt to hide fraud patterns and test robust aggregation under realistic non-IID institutional data.

Federated graph and relational learning: evaluate cross-institution graph-based detection (Suzumura et al., 2019) against tabular federated models.

Explainable and auditable federated models: investigate explanation methods and audit records that satisfy investigators and regulators (Awosika et al., 2024).

Deployment studies and governance: document pilots with partner institutions, covering latency, integration with investigation workflows, incentive design, and legal agreements.

Future studies should distinguish measured findings from conceptual claims and avoid assuming that results from one dataset or model apply universally. They should also report trade-offs: stronger privacy may reduce detection quality, robust aggregation may add cost, and personalisation may increase complexity.

VII.CONCLUSION AND FUTURE SCOPE

This review considered selected literature on federated learning for fraud detection. Together, the works suggest that FL is a plausible approach to collaborative fraud detection without centralising raw transaction data, but that it does not eliminate the core difficulties of fraud detection and adds challenges of heterogeneity, privacy, security, communication, and deployment. The fraud-specific federated studies reviewed (Yang et al., 2019b; Zheng et al., 2020; Suzumura et al., 2019; Awosika et al., 2024) are fewer than the foundational studies on which they depend, and many enabling results come from non-fraud settings.

The review is limited by its selected set of publications and is not an exhaustive account of the field. The studies use different methods, datasets, and settings, and their findings should not be treated as one uniform experimental result. A broader future review should search multiple scholarly databases, document search terms and screening decisions, and include additional work on vertical FL, anti-money-laundering, insurance and securities fraud, regulatory perspectives, and industry deployments.

In conclusion, federated learning is best treated as a privacy-oriented tool within a wider fraud-detection programme that also requires careful evaluation, layered security, drift management, and human oversight. Future research should focus on realistic benchmarks, joint handling of

imbalance and heterogeneity, robust and privacy-preserving training, drift adaptation, and evidence from real deployments.

REFERENCES

1. Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS '16)* (pp. 308–318). ACM. <https://doi.org/10.1145/2976749.2978318>
2. Abdallah, A., Maarof, M. A., & Zainal, A. (2016). Fraud detection system: A survey. *Journal of Network and Computer Applications*, 68, 90–113. <https://doi.org/10.1016/j.jnca.2016.04.007>
3. Awosika, T., Shukla, R. M., & Pranggono, B. (2024). Transparency and privacy: The role of explainable AI and federated learning in financial fraud detection. *IEEE Access*, 12, 64551–64560. <https://doi.org/10.1109/ACCESS.2024.3394528>
4. Bagdasaryan, E., Veit, A., Hua, Y., Estrin, D., & Shmatikov, V. (2020). How to backdoor federated learning. In *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics (AISTATS) (PMLR Vol. 108, pp. 2938–2948)*. <https://proceedings.mlr.press/v108/bagdasaryan20a.html>
5. Bhagoji, A. N., Chakraborty, S., Mittal, P., & Calo, S. (2019). Analyzing federated learning through an adversarial lens. In *Proceedings of the 36th International Conference on Machine Learning (ICML) (PMLR Vol. 97, pp. 634–643)*. <https://proceedings.mlr.press/v97/bhagoji19a.html>
6. Blanchard, P., El Mhamdi, E. M., Guerraoui, R., & Stainer, J. (2017). Machine learning with adversaries: Byzantine tolerant gradient descent. In *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)*. <https://papers.nips.cc/paper/2017/hash/f4b9ec30ad9f68f89b29639786cb62ef-Abstract.html>

7. Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., Ramage, D., Segal, A., & Seth, K. (2017). Practical secure aggregation for privacy-preserving machine learning. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS '17)* (pp. 1175–1191). ACM. <https://doi.org/10.1145/3133956.3133982>
8. Bonawitz, K., Eichner, H., Grieskamp, W., Huba, D., Ingerman, A., Ivanov, V., Kiddon, C., Konečný, J., Mazzocchi, S., McMahan, H. B., Van Overveldt, T., Petrou, D., Ramage, D., & Roselander, J. (2019). Towards federated learning at scale: System design. In *Proceedings of Machine Learning and Systems (MLSys 2019)*. <https://arxiv.org/abs/1902.01046>
9. 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. <https://doi.org/10.1016/j.ins.2019.05.042>
10. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321–357. <https://doi.org/10.1613/jair.953>
11. Dal Pozzolo, A., Caelen, O., Johnson, R. A., & Bontempi, G. (2015). Calibrating probability with undersampling for unbalanced classification. In *2015 IEEE Symposium Series on Computational Intelligence (SSCI)* (pp. 159–166). IEEE. <https://doi.org/10.1109/SSCI.2015.33>
12. Dal Pozzolo, A., Boracchi, G., Caelen, O., Alippi, C., & Bontempi, G. (2018). Credit card fraud detection: A realistic modeling and a novel learning strategy. *IEEE Transactions on Neural Networks and Learning Systems*, 29(8), 3784–3797. <https://doi.org/10.1109/TNNLS.2017.2736643>
13. Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3–4), 211–407. <https://doi.org/10.1561/04000000042>
14. Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM Computing Surveys*, 46(4), Article 44. <https://doi.org/10.1145/2523813>
15. Geyer, R. C., Klein, T., & Nabi, M. (2017). Differentially private federated learning: A client level perspective. *arXiv preprint arXiv:1712.07557*. <https://arxiv.org/abs/1712.07557>
16. Hilal, W., Gadsden, S. A., & Yawney, J. (2022). Financial fraud: A review of anomaly detection techniques and recent advances. *Expert Systems with Applications*, 193, Article 116429. <https://doi.org/10.1016/j.eswa.2021.116429>
17. Hsieh, K., Phanishayee, A., Mutlu, O., & Gibbons, P. (2020). The non-IID data quagmire of decentralized machine learning. In *Proceedings of the 37th International Conference on Machine Learning (ICML)* (PMLR Vol. 119, pp. 4387–4398). <https://proceedings.mlr.press/v119/hsieh20a.html>
18. Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., et al. (2021). Advances and open problems in federated learning. *Foundations and Trends in Machine Learning*, 14(1–2), 1–210. <https://doi.org/10.1561/22000000083>
19. Karimireddy, S. P., Kale, S., Mohri, M., Reddi, S., Stich, S., & Suresh, A. T. (2020). SCAFFOLD: Stochastic controlled averaging for federated learning. In *Proceedings of the 37th International Conference on Machine Learning (ICML)* (PMLR Vol. 119, pp. 5132–5143). <https://proceedings.mlr.press/v119/karimireddy20a.html>
20. Konečný, J., McMahan, H. B., Yu, F. X., Richtárik, P., Suresh, A. T., & Bacon, D. (2016). Federated learning: Strategies for improving communication efficiency. *arXiv preprint arXiv:1610.05492*. <https://arxiv.org/abs/1610.05492>
21. Li, Q., Wen, Z., Wu, Z., Hu, S., Wang, N., Li, Y., Liu, X., & He, B. (2023). A survey on federated

- learning systems: Vision, hype and reality for data privacy and protection. *IEEE Transactions on Knowledge and Data Engineering*, 35(4), 3347–3366. <https://doi.org/10.1109/TKDE.2021.3124599>
23. Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020a). Federated learning: Challenges, methods, and future directions. *IEEE Signal Processing Magazine*, 37(3), 50–60. <https://doi.org/10.1109/MSP.2020.2975749>
24. Li, T., Sahu, A. K., Zaheer, M., Sanjabi, M., Talwalkar, A., & Smith, V. (2020b). Federated optimization in heterogeneous networks. In *Proceedings of Machine Learning and Systems (MLSys 2020)*, 2, 429–450. <https://arxiv.org/abs/1812.06127>
25. Lu, J., Liu, A., Dong, F., Gu, F., Gama, J., & Zhang, G. (2019). Learning under concept drift: A review. *IEEE Transactions on Knowledge and Data Engineering*, 31(12), 2346–2363. <https://doi.org/10.1109/TKDE.2018.2876857>
26. McMahan, B., Moore, E., Ramage, D., Hampson, S., & Agüera y Arcas, B. (2017). Communication-efficient learning of deep networks from decentralized data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS) (PMLR Vol. 54, pp. 1273–1282)*. <https://proceedings.mlr.press/v54/mcmahan17a.html>
27. McMahan, H. B., Ramage, D., Talwar, K., & Zhang, L. (2018). Learning differentially private recurrent language models. In *International Conference on Learning Representations (ICLR 2018)*. <https://arxiv.org/abs/1710.06963>
28. Nasr, M., Shokri, R., & Houmansadr, A. (2019). Comprehensive privacy analysis of deep learning: Passive and active white-box inference attacks against centralized and federated learning. In *2019 IEEE Symposium on Security and Privacy (SP) (pp. 739–753)*. IEEE. <https://doi.org/10.1109/SP.2019.00065>
29. 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. <http://doi.org/10.1016/j.dss.2010.08.006>
30. Sattler, F., Wiedemann, S., Müller, K.-R., & Samek, W. (2020). Robust and communication-efficient federated learning from non-i.i.d. data. *IEEE Transactions on Neural Networks and Learning Systems*, 31(9), 3400–3413. <https://doi.org/10.1109/TNNLS.2019.2944481>
31. Suzumura, T., Zhou, Y., Baracaldo, N., et al. (2019). Towards federated graph learning for collaborative financial crimes detection. *arXiv preprint arXiv:1909.12946*. <https://arxiv.org/abs/1909.12946>
32. Wang, J., Liu, Q., Liang, H., Joshi, G., & Poor, H. V. (2020). Tackling the objective inconsistency problem in heterogeneous federated optimization. In *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*. <https://arxiv.org/abs/2007.07481>
33. Wei, K., Li, J., Ding, M., Ma, C., Yang, H. H., Farokhi, F., Jin, S., Quek, T. Q. S., & Poor, H. V. (2020). Federated learning with differential privacy: Algorithms and performance analysis. *IEEE Transactions on Information Forensics and Security*, 15, 3454–3469. <https://doi.org/10.1109/TIFS.2020.2988575>
34. West, J., & Bhattacharya, M. (2016). Intelligent financial fraud detection: A comprehensive review. *Computers & Security*, 57, 47–66. <https://doi.org/10.1016/j.cose.2015.09.005>
35. Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019a). Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology*, 10(2), Article 12. <https://doi.org/10.1145/3298981>
36. Yang, W., Zhang, Y., Ye, K., Li, L., & Xu, C.-Z. (2019b). FFD: A federated learning based method for credit card fraud detection. In *Big Data – BigData 2019 (Lecture Notes in Computer Science, Vol. 11514, pp. 18–32)*. Springer. https://doi.org/10.1007/978-3-030-23551-2_2
37. Yin, D., Chen, Y., Ramchandran, K., & Bartlett, P. (2018). Byzantine-robust distributed

- learning: Towards optimal statistical rates. In Proceedings of the 35th International Conference on Machine Learning (ICML) (PMLR Vol. 80, pp. 5650–5659). <https://proceedings.mlr.press/v80/yin18a.html>
39. Zhao, Y., Li, M., Lai, L., Suda, N., Civin, D., & Chandra, V. (2018). Federated learning with non-IID data. arXiv preprint arXiv:1806.00582. <https://arxiv.org/abs/1806.00582>
 40. Zheng, W., Yan, L., Gou, C., & Wang, F.-Y. (2020). Federated meta-learning for fraudulent credit card detection. In Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-20) (pp. 4654–4660). <https://doi.org/10.24963/ijcai.2020/642>
 41. Zhu, L., Liu, Z., & Han, S. (2019). Deep leakage from gradients. In Advances in Neural Information Processing Systems 32 (NeurIPS 2019). <https://arxiv.org/abs/1906.08935>