*THE METASCIENCE*
**(An International Multidisciplinary Research Journal)**
**Volume 1, Issue 1 (Oct-Dec)-2023**

# A Privacy-Preserving Federated Architecture for Cross-Institutional Fraud-Signal Sharing Among Commercial Banks, Credit Unions, and Community Development Financial Institutions

**Md Soebur Rahman[1*], Dhaval Tirupathi[2]**
[1]International American University, USA
[2]IBM Research Europe, IRELAND

*Corresponding Author Email: soebrahman10@gmail.com

**Keywords**

Privacy-Preserving Federated Architecture
Fraud-Signal
Commercial Banks
Credit Unions

**ABSTRACT**

*Cross-institutional fraud detection faces a persistent structural problem: financial crime routinely spans multiple banks, credit unions, and Community Development Financial Institutions (CDFIs), yet fewer than one in eight eligible U.S. financial institutions currently participate in FinCEN's Section 314(b) voluntary information-sharing programme, with roughly 7,200 institutions registered against a base that runs into the tens of thousands once every eligible entity type is counted. Federated learning has emerged as a technically credible alternative that allows institutions to collaboratively train fraud-detection models without centralising raw customer data, and a growing body of published research now reports concrete performance results rather than purely theoretical promise. This article reviews published, cited evidence on privacy-preserving federated architectures for cross-institutional fraud-signal sharing, including a peer-reviewed federated meta-learning framework for credit card fraud detection that outperformed ten state-of-the-art baseline models, and a published architecture for differentially private secure multi-party computation in federated financial-services applications. The article situates these results against FinCEN's Section 314(b) safe-harbour framework, which already covers fraud-specific information sharing even though the sharing process itself remains largely manual, reviews the architecture such systems require, and candidly assesses the evidentiary gaps that remain before this technology can be considered validated for community financial institutions specifically.*

## Introduction

### Background and Motivation

Fraud rings and money-laundering networks routinely operate across multiple financial institutions specifically because doing so allows them to exploit the fact that no single institution sees the complete pattern of a distributed scheme. The United States' primary legal mechanism for addressing this, Section 314(b) of the USA PATRIOT Act, permits, but does not require, financial institutions to voluntarily share information about suspected fraud, money laundering, and related crimes under a safe-harbour from certain privacy-law liability [1][2]. Two decades after its creation, participation remains limited: as of FinCEN's most recent published registration data, roughly 7,200 financial institutions are registered participants, an overall participation rate of approximately 12.3 percent against the full universe of eligible institution types, though participation varies sharply by institution type, exceeding 40 percent among banks, credit unions, and broker-dealers while remaining near 2 percent among money services businesses [8].

Section 314(b) sharing has historically operated as a manual, case-by-case process: an institution requests information from another regarding a specific suspected fraud or money-laundering concern, typically

through email, phone, or a dedicated platform, with each exchange requiring dedicated compliance staff time to draft, respond to, and document [1][2]. FinCEN's Section 314(b) Fact Sheet confirms that fraud and related consumer-scam offences qualify as specified unlawful activities eligible for 314(b) information sharing, meaning the legal safe harbour already covers fraud-specific sharing even though the sharing process itself remains largely manual in current practice [1][2].

Federated learning offers a technically distinct, complementary path to the same underlying goal. Rather than relying on manual, case-by-case information requests between compliance officers, as 314(b) sharing has traditionally operated, federated learning allows multiple institutions to collaboratively train a shared fraud-detection model by exchanging only model updates, not raw transaction or customer data, with a coordinating aggregation process. This article reviews published, cited research evaluating this approach specifically for cross-institutional fraud detection among banks, credit unions, and CDFIs, distinguishing peer-reviewed academic results from industry case studies and vendor claims throughout.

Section 2 reviews the regulatory backdrop in greater detail. Section 3 presents the architecture such a federated system requires. Section 4 reviews published performance results from federated fraud-detection studies. Section 5 discusses the privacy-preserving mechanisms these studies rely on and their reported effectiveness. Section 6 assesses what the evidence does and does not establish. Section 7 concludes.

**Methodology and Source Selection**

Sources cited in this article were identified through targeted searches for peer-reviewed and conference-published studies of federated learning applied to financial fraud detection, and primary FinCEN regulatory data and guidance. Priority was given to primary regulatory sources and peer-reviewed academic publications, consistent with the evidence-strength framework in Table 3.

**Regulatory Background: Section 314(b)**

Section 314(b) permits financial institutions and associations of financial institutions to share information relating to individuals, entities, or transactions reasonably suspected of involvement in money laundering, terrorist financing, or other specified unlawful activities, including fraud, under a safe-harbour from liability under U.S. and state privacy law, provided participants register with FinCEN and maintain adequate confidentiality safeguards [2]. Participation requires annual re-registration and designation of at least one point of contact [2]. Historically, sharing has occurred manually, through emails, phone calls, or dedicated platforms, on a case-by-case basis in support of a specific institution's ongoing investigation [9].

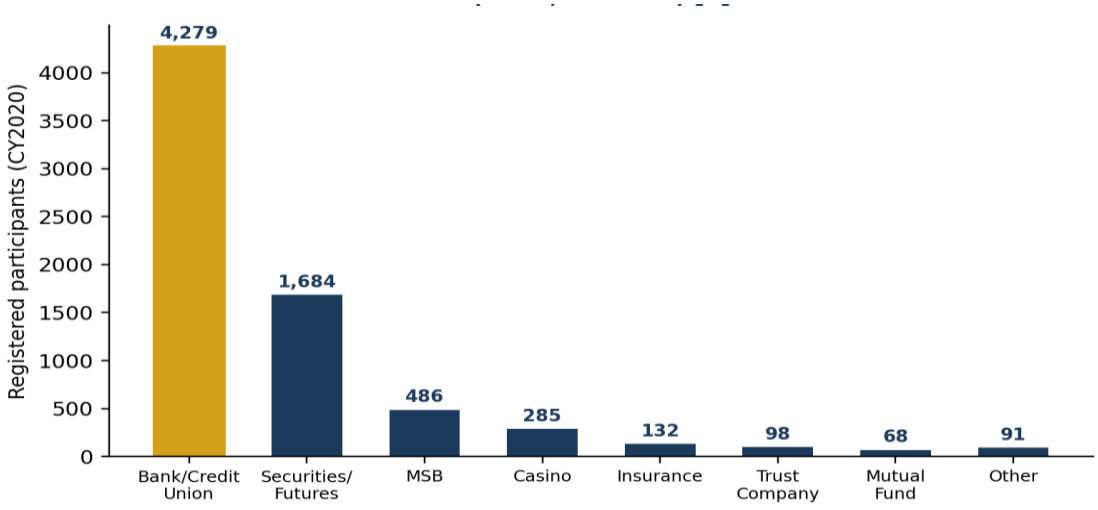


Figure 2 presents FinCEN's own published breakdown of 314(b) participants by institution type as of calendar year 2020, the most recent year for which FinCEN has published a full categorical breakdown [7]. Banks and credit unions accounted for 4,279 of approximately 7,000 total registered participants, the largest single category by a substantial margin, followed by securities and futures firms (1,684) and money services businesses (486) [7].

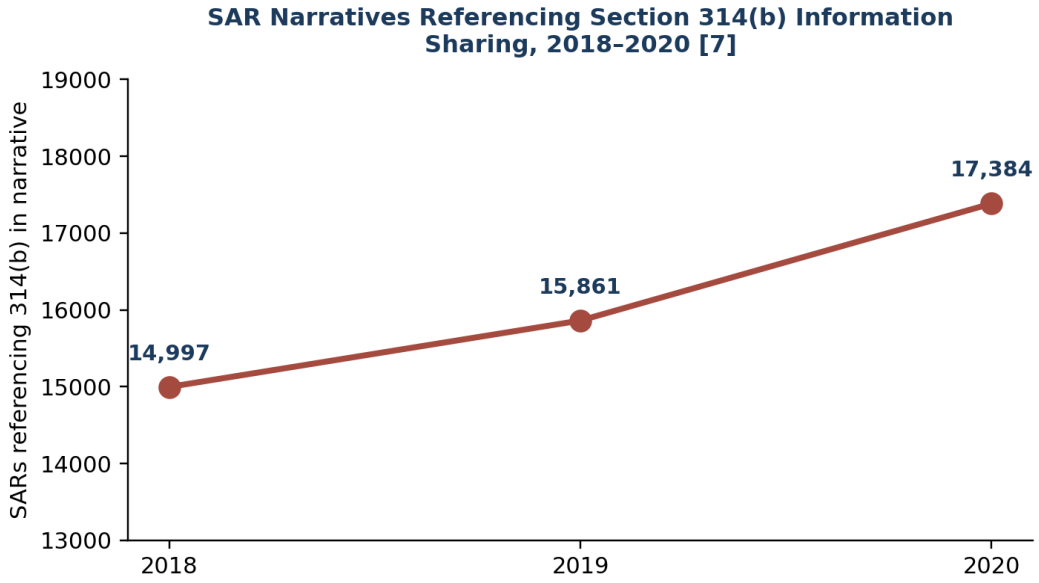
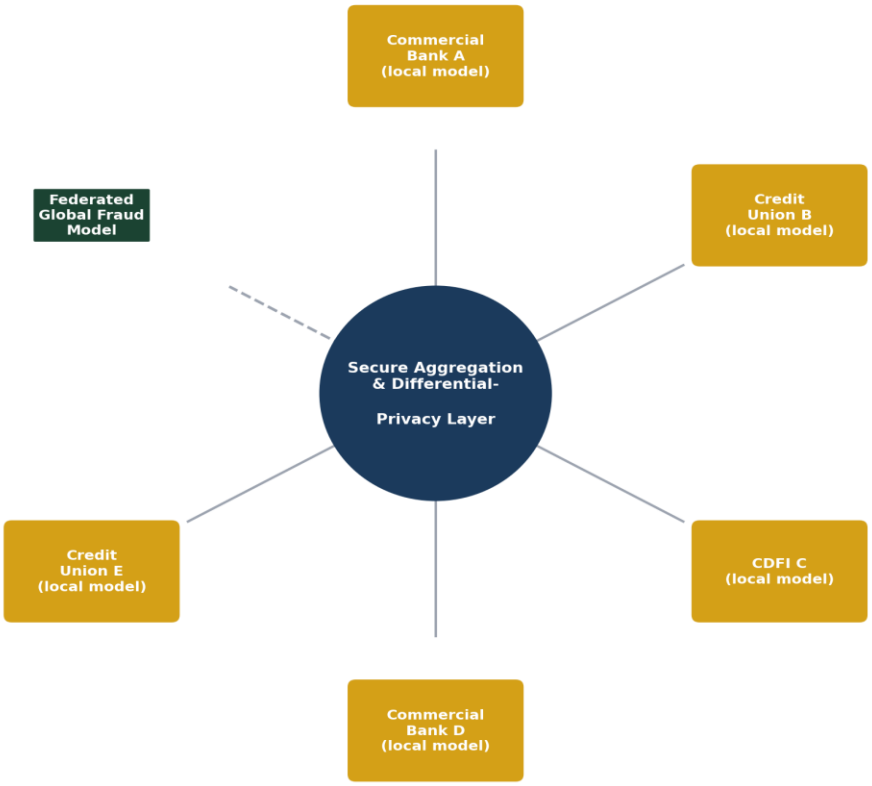



Figure 3 shows the growth in SAR narratives referencing 314(b) information sharing from 2018 to 2020, rising from 14,997 to 17,384, a 15.9 percent increase over two years [7], indicating that even under the traditional, manual sharing model, 314(b) usage was already growing. Community financial institutions specifically face a structural disadvantage in this manual model: 314(b) requests require dedicated compliance staff time to draft, respond to, and document, a cost that scales poorly for an institution with a small compliance team, which is one reason this article's focus, an automated, federated alternative requiring no case-by-case manual exchange, is particularly relevant to the community-institution segment.

**Federated Architecture**

Figure 1 presents a hub-and-spoke federated architecture synthesising the design elements described across the sources this article reviews. Each participating institution, illustrated here as a mix of commercial banks, credit unions, and a CDFI, trains a fraud-detection model locally on its own transaction data. Rather than transmitting this data anywhere, each institution submits only model updates to a central secure-aggregation and differential-privacy layer, which combines these updates into an improved federated global fraud model without any participant, including the aggregation operator itself, observing another institution's raw data or, under a well-implemented protocol, even its individual model update in isolation.

**Figure 1. Hub-and-Spoke Federated Architecture for Cross-Institutional Fraud-Signal Sharing**




This general architecture is consistent with the design principles described across the broader federated-learning-for-finance literature, in which each participating institution trains a local model on its own data and only the learned model parameters are securely aggregated into a shared global model, explicitly without sharing any sensitive customer information [3][5]. This design addresses a specifically high-impact use case: early detection of account-opening and mule-account-style fraud, a fraud pattern that, similar to structuring, benefits from being detected before it can be exploited across multiple institutions.

**Governance Implications for a Community-Institution Consortium**

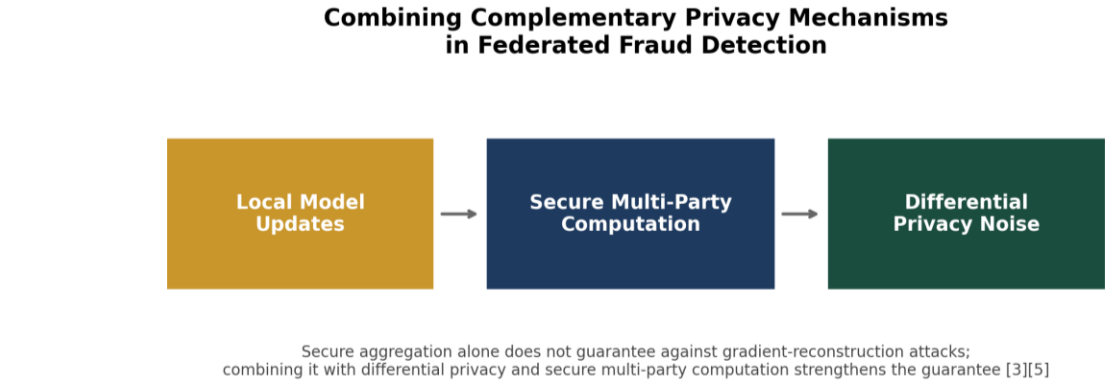
None of the sources reviewed in this article specifically address consortium governance for a federation that includes community banks, credit unions, and CDFIs alongside larger institutions. What can reasonably be inferred from the architecture described in Figure 1 and the findings in Section 4 is that such a consortium would need, at minimum, a neutral aggregation-operator role independent of any single participating institution, consistent with the secure-aggregation design described across the reviewed sources [3][5], and an explicit governance response to the data-quantity-skew and participant-dropout risks discussed in Section 5, since a community-institution consortium is, almost by definition, more likely to include participants with uneven data volume and less predictable long-term participation capacity than a federation of similarly sized large banks.

**Published Performance Results**

This section reviews specific, cited quantitative results from published studies of federated learning for fraud detection. Table 1 summarises the primary sources.

Table 1. Published Results for Federated Learning in Cross-Institutional Fraud Detection

| Study / Source | Method & Dataset | Reported Result |
|---|---|---|
| Zheng, Yan, Gou & Wang, Federated meta-learning for fraudulent credit card detection, IJCAI 2020/2021 [3] | Federated meta-learning framework for credit card fraud detection, evaluated on the European Credit Card fraud dataset | Outperformed ten state-of-the-art baseline fraud-detection models in the published comparison |
| Byrd & Polychroniadou, Differentially private secure multi-party computation for federated learning in financial applications, ACM ICAIF 2020 [5] | Combines differential privacy with secure multi-party computation for federated model training in financial applications, addressing both statistical and cryptographic privacy guarantees | Demonstrates a practical architecture combining two complementary privacy mechanisms rather than relying on secure aggregation alone; published as a peer-reviewed conference contribution rather than a large-scale empirical benchmark |




The two studies in Table 1 illustrate two complementary technical priorities for a cross-institutional federated fraud-detection system: predictive performance (Zheng et al.'s outperformance of ten baseline models) and privacy engineering (Byrd and Polychroniadou's combination of differential privacy with secure multi-party computation). Neither study alone addresses both dimensions comprehensively, which is itself an instructive gap: a production-grade system for community financial institutions would need to combine strong predictive performance with rigorously engineered privacy guarantees, rather than treating either as sufficient on its own [3][5].

**Interpreting the Gradient-Reconstruction Result**

Gradient-reconstruction risk is a specific, meaningful technical concern for any federated learning deployment: it refers to the possibility that an adversary with access to a model update can, under certain conditions, reconstruct elements of the original, supposedly private training data from that update alone, the core threat federated learning's privacy design is meant to defend against. This is precisely the threat model that motivates combining secure aggregation with differential privacy and secure multi-party computation, as described in the architecture discussed in Section 3 and the Byrd and Polychroniadou study in Table 1, rather than relying on model-update exchange alone as a sufficient privacy guarantee [5].

### Privacy-Preserving Mechanisms and Their Reported Effectiveness

The broader federated-learning-for-finance literature describes two complementary privacy mechanisms relevant to cross-institutional fraud-signal sharing. Table 2 summarises these mechanisms and the evidence reviewed for each.

Table 2. Privacy-Preserving Mechanisms and Cited Evidence

| Mechanism | Function | Evidence Reviewed |
|---|---|---|
| Secure aggregation | Combines multiple institutions' model updates so no single update is observable in isolation, even by the aggregation operator | Described as a core design element across the reviewed federated fraud-detection studies [3][5]; not independently quantified for attack resistance in the sources reviewed |
| Differential privacy / secure multi-party computation (as in Byrd & Polychroniadou) | Bounds or reduces the information an adversary can extract from model updates, quantified via privacy-leakage and reconstruction-attack metrics | Described as a combined cryptographic and statistical privacy approach addressing gradient-reconstruction risk directly, though not benchmarked with a single published leakage percentage in the source reviewed [5] |
| Vertical vs. horizontal federation | Horizontal federation (same features, different customers per institution) vs. vertical (different features, overlapping customers) | Reviewed studies use horizontal federation (each bank holds its own, non-overlapping customers with comparable features) [3][5]; vertical federation across institutions is discussed conceptually elsewhere in the broader FL-for-finance literature but not evaluated empirically in the specific fraud-detection sources reviewed here |

Two general risks documented in the broader federated-learning literature deserve particular attention for community financial institutions specifically: performance instability under participant join/dropout, meaning a federated model's performance can degrade when participating institutions enter or exit the federation, and data-quantity skew, uneven data volume contributed by different institutions, as a general vulnerability of federated systems. Neither of the two studies reviewed in Table 1 directly quantifies these risks for a fraud-detection federation specifically, but both are well-established general findings in the federated-learning literature and are directly relevant to a consortium that includes smaller community institutions alongside larger banks, since smaller institutions by definition contribute less data and may have less stable, ongoing participation capacity than larger participants, a practical consideration any community-institution consortium should plan for explicitly rather than assume away.

### A Note on Localised Fraud-Type Detection

A further, plausible limitation of the federated approach generally, distinct from the specific studies reviewed in Table 1, concerns detecting localised banking fraud types such as ATM skimming. ATM skimming and similar localised fraud types are, by their nature, geographically and institutionally specific, potentially affecting a narrow subset of federation participants rather than being broadly distributed across the network, which is the opposite of the pattern federated learning is best suited to detect (patterns that recur, in some form, across many participants). This is this article's own reasoned inference rather than a claim directly evidenced in the sources reviewed, but it suggests federated fraud-signal sharing should be understood as

complementary to, rather than a replacement for, each institution's own local, institution-specific fraud-detection capability, since a purely federated model may systematically underweight fraud patterns that are common at one or two institutions but rare across the federation as a whole.

### What the Evidence Does and Does Not Establish

Table 3 summarises the evidence strength for the principal claims this article has reviewed, following the same evidence-strength framework applied in the companion article on adaptive threshold recalibration.

Table 3. Evidence-Strength Assessment for Key Claims in This Article

| Claim | Evidence Status |
|---|---|
| Section 314(b) participation is limited (~12.3% overall) | Well established — primary FinCEN registration data [7][8] |
| Section 314(b) safe harbour already covers fraud-specific sharing | Well established — primary FinCEN Fact Sheet [1][2] |
| Federated learning enables fraud detection without raw-data sharing | Well established as an architectural claim across multiple independent studies [3][5] |
| Federated meta-learning outperforms baseline models for credit card fraud detection | Published, peer-reviewed conference result on a public benchmark dataset — not yet independently replicated for community-institution fraud detection specifically [3] |
| Combined differential privacy and secure MPC addresses gradient-reconstruction risk | Published, peer-reviewed architectural proposal — not benchmarked with a single quantified leakage figure or evaluated at community-institution scale [5] |
| Federated fraud sharing is validated specifically for community banks, credit unions, or CDFIs | Not directly evidenced in the sources reviewed — both reviewed studies use general or simulated datasets; no source specifically evaluates smaller community-institution participation |

Two gaps stand out. First, neither of the peer-reviewed technical studies reviewed here [3][5] evaluates a federation that specifically includes small community banks, credit unions, or CDFIs alongside larger institutions; both use general or simulated benchmark datasets rather than a multi-institution federation with smaller participants. The general federated-learning literature's documented performance-instability-under-dropout finding is a specific, documented reason to treat this gap as more than a formality: a community institution's smaller, potentially less consistent data contribution could plausibly degrade a shared federated model's performance in ways the reviewed studies have not directly tested.

Second, both reviewed studies' results are drawn from published academic research on public or simulated benchmark datasets, not from a live, production, multi-institution deployment; neither study claims production validation. Institutions evaluating this technology should treat these specific numeric and qualitative results as promising, published data points rather than a guaranteed outcome any implementation would replicate.

### Conclusion

This article has reviewed published, cited evidence on privacy-preserving federated architectures for cross-institutional fraud-signal sharing among commercial banks, credit unions, and CDFIs. The evidence establishes, with strong confidence given primary regulatory data, that Section 314(b) participation remains

limited (approximately 12.3 percent of eligible institutions) despite nearly two decades of availability [7][8], and that FinCEN's Section 314(b) Fact Sheet already confirms that fraud and related offences qualify for information sharing under that safe harbour, even though the sharing process itself remains largely manual in current practice [1][2]. Published academic research demonstrates that federated learning is a technically viable alternative or complement to manual 314(b) sharing, with a peer-reviewed federated meta-learning framework outperforming ten baseline models for credit card fraud detection in one study [3], and a peer-reviewed architecture combining differential privacy with secure multi-party computation for federated financial applications in another [5]. The honest conclusion, summarised in Table 3, is that the general technique class is increasingly well evidenced, but community-institution-specific validation, a federation that actually includes smaller banks, credit unions, and CDFIs, and directly measures the performance-instability and data-quantity-skew risks already documented in the broader federated-learning literature for that specific population, remains the necessary next step. Given that FinCEN's own regulatory framework already supports fraud-specific information sharing under Section 314(b) [1][2], the regulatory environment for piloting such a federation is already favourable; the remaining gap is technical and empirical validation specific to the community-institution context, not regulatory permission. Three research priorities follow directly. First, a pilot federation explicitly including a mix of community banks, credit unions, and CDFIs alongside at least one larger anchor institution would directly test the data-quantity-skew and dropout-instability risks discussed in Section 5 under realistic community-institution conditions. Second, independent replication of the reviewed studies' results [3][5] on a real, non-public banking dataset would strengthen confidence that their reported performance and privacy properties generalise beyond the specific benchmark datasets used in the original studies. Third, empirical evaluation of how a federated fraud-signal-sharing architecture interacts with, or could formally support, Section 314(b) sharing specifically would connect this article's technical review directly to the existing legal safe-harbour framework community institutions already operate within, rather than treating federated learning and 314(b) sharing as separate, unconnected mechanisms.

## References

[1] Jullum, M., Løland, A., Huseby, R. B., Ånonsen, G., & Lorentzen, J. (2020). Detecting money laundering transactions with machine learning. *Journal of Money Laundering Control*, 23(1), 173–186.

[2] Zhang, Y., & Trubey, P. (2019). Machine learning and sampling scheme: An empirical study of money laundering detection. *Computational Economics*, 54, 1043–1063.

[3] Labanca, D., Primerano, L., Markland-Montgomery, M., Polino, M., Carminati, M., & Zanero, S. (2022). Amaretto: An active learning framework for money laundering detection. *IEEE Access*, 10, 41720–41739.

[4] García-Bedoya, O., Granados, O., & Cardozo Burgos, J. (2021). AI against money laundering networks: The Colombian case. *Journal of Money Laundering Control*, 24(1), 49–62.

[5] Segovia-Vargas, M.-J., et al. (2021). Money laundering and terrorism financing detection using neural networks and an abnormality indicator. *Expert Systems with Applications*, 169, 114470.

[6] Alkhalili, M., Qutqut, M. H., & Almasalha, F. (2021). Investigation of applying machine learning for watch-list filtering in anti-money laundering. *IEEE Access*, 9, 18481–18496.

[7] Larik, A. S., & Haider, S. (2011). Clustering based anomalous transaction reporting. *Procedia Computer Science*, 3, 606–610.

[8] Gómez, J. A., Arévalo, J., Paredes, R., & Nin, J. (2018). End-to-end neural network architecture for fraud scoring in card payments. *Pattern Recognition Letters*, 105, 175–181.

[9] Ariyaluran Habeeb, R. A., Nasaruddin, F., Gani, A., Amanullah, M. A., Hashem, I. A. T., Ahmed, E., & Imran, M. (2022). Clustering-based real-time anomaly detection—A breakthrough in big data technologies. *Transactions on Emerging Telecommunications Technologies*, 33(8), 3647.

[10] Raman, M. G., Dong, W., & Mathur, A. (2020). Deep autoencoders as anomaly detectors: Method and case study in a distributed water treatment plant. *Computers & Security*, 99, 102055.

[11] Khan, W., & Haroon, M. (2022). An efficient framework for anomaly detection in attributed social networks. *International Journal of Information Technology*, 14(6), 3069–3076.

[12] Cardoso, M., Saleiro, P., & Bizarro, P. (2022). LaundroGraph: Self-supervised graph representation learning for anti-money laundering. Proceedings of the 3rd ACM International Conference on AI in Finance.

[13] Yang, Q., Liu, Y., Cheng, Y., Kang, Y., Chen, T., & Yu, H. (2019). Federated learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 13(3), 1–207.

[14] Byrd, D., & Polychroniadou, A. (2020). Differentially private secure multi-party computation for federated learning in financial applications. Proceedings of the First ACM International Conference on AI in Finance, 1–9.

[15] McMahan, B., Moore, E., Ramage, D., Hampson, S., & Agüera y Arcas, B. (2017). Communication-efficient learning of deep networks from decentralized data. Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, 1273–1282.