THE COMPUTERTECH
(An International Peer Review Journal)
VOLUME 8; ISSUE 2 (JULY-DEC); (2022)
WEBSITE: THE COMPUTERTECH

Adaptive Threshold Recalibration for Real-Time Detection of Structuring and Layering Patterns in BSA/AML Transaction Monitoring

Md Soebur Rahman[1*], Chashi Sabrina Islam[1], Charles Møller[2]
[1] International American University, USA
[2]University of Arkansas at Little Rock (ERIQ), USA

*Corresponding Author Email: soebrahman10@gmail.com

Abstract

Rule-based transaction-monitoring systems, still the dominant approach for detecting structuring and layering under the Bank Secrecy Act (BSA), rely on fixed thresholds that generate false-positive rates industry sources place at 90 to 95 percent, imposing enormous investigative cost while the U.S. Financial Crimes Enforcement Network (FinCEN) reports suspicious activity report (SAR) filings grew substantially in recent years, reaching roughly 2.5 million filings in fiscal year 2020 across all filer types. This article surveys published, cited evidence on adaptive threshold recalibration as an alternative to static rule-based thresholds for real-time detection of structuring and layering patterns, drawing on peer-reviewed and industry-benchmarked results rather than illustrative projections. Published studies report strong illicit-transaction classification performance when graph-based and ensemble machine-learning methods are combined with dynamic behavioural features on large, labelled AML transaction datasets, alongside published findings that novel graph-derived features can materially improve detection-model F1 scores over rule-based baselines, and vendor-reported reductions in the 50-to-80-percent range under production deployment conditions. The article reviews the architecture such systems require, the specific statistical and machine-learning techniques underlying published results, the regulatory context for structuring-related SAR filings, and the evidentiary limits of currently available research, concluding with a candid assessment of what remains unverified pending broader multi-institution validation.

Keywords: Real-Time Detection; Recalibration; Layering Patterns; BSA/AML Transaction

Introduction

Structuring, the deliberate division of transactions to fall below the $10,000 Currency Transaction Report (CTR) threshold, and layering, the movement of illicit funds through multiple transactions and accounts to obscure their origin, remain among the most commonly cited typologies in U.S. Bank Secrecy Act (BSA) enforcement. Detecting both patterns in real time, before funds are further layered or withdrawn, has long depended on rule-based transaction-monitoring systems (TMS) that apply fixed thresholds to individual transactions or short transaction sequences.

The operational cost of this approach is well documented. Multiple independent sources, including a PricewaterhouseCoopers analysis cited widely across subsequent industry research, place the false-positive rate of rule-based transaction-monitoring alerts at 90 to 95 percent [1][2][3]. A

qualitative study of eight anti-money-laundering (AML) specialists, drawing on 480 minutes of semi-structured interviews and published in Future Generation Computer Systems, confirmed this same 90-to-95-percent range as the industry average reported directly by practitioners [3]. The Financial Action Task Force (FATF) has been cited as putting the global cost of AML compliance above 180 billion U.S. dollars annually, a substantial share of which industry analysts attribute to false-positive alert handling rather than genuine financial-crime prevention [5].

Meanwhile, the volume of suspicious activity reporting that this monitoring infrastructure feeds continues to grow. FinCEN's own published SAR filing trend data recorded roughly 2.5 million suspicious activity reports filed in fiscal year 2020 across all filer types, alongside several million currency transaction reports filed over the same period; check-fraud SARs alone totalled roughly 350,000 in 2021, and elder-financial-exploitation SARs totalled roughly 72,173 in 2021 [9][10]. SAR filing volume has grown substantially since FinCEN began publishing this data in 2014, reflecting both genuine growth in reportable activity and the expanding scope of what institutions are expected to monitor and report.

This article surveys the case for adaptive threshold recalibration, replacing fixed, manually set thresholds with thresholds that adjust dynamically to an entity's evolving transaction behaviour and to broader shifts in the transaction population, as a response to this cost-and-volume problem, drawing specifically on published, cited results rather than illustrative or hypothetical figures. Section 2 reviews the regulatory background. Section 3 presents the architecture such a system requires. Section 4 reviews the published statistical and machine-learning techniques and their reported results. Section 5 presents a broader statistical picture of the reporting environment these techniques operate within. Section 6 discusses what the current evidence base does and does not establish. Section 7 concludes.

Methodology and Source Selection

The sources cited throughout this article were identified through targeted searches for published academic studies, regulator (FinCEN, federal banking agency) primary data, and industry-analyst reporting on adaptive threshold techniques, AML false-positive rates, and BSA/SAR filing statistics. Priority was given to primary regulatory sources (FinCEN data releases and interagency guidance) and peer-reviewed or conference-published academic studies where available; industry and vendor sources were included where they provided the only available quantification of a relevant phenomenon (for example, aggregate compliance-cost estimates), but are flagged as such throughout, and their claims are treated with appropriately greater caution than the peer-reviewed and primary-regulatory sources, consistent with the evidence-strength assessment presented in Table 4 in Section 6.

Regulatory Background

Structuring is defined under 31 U.S.C. § 5324 as conducting, or attempting to conduct, transactions with a financial institution in a manner designed to evade the CTR filing requirement, and it is a criminal offence independent of whether the structured funds themselves derive from an underlying crime. Detecting structuring has historically relied on threshold-based rules flagging transactions just below the $10,000 CTR threshold, a specific pattern rules-based systems are relatively well suited to detect in isolation, but layering, and structuring conducted across multiple accounts,

branches, or a longer time horizon, requires monitoring sequences and relationships that static, transaction-level thresholds do not capture well.

FinCEN, jointly with the Federal Reserve Board, the FDIC, the NCUA, and the OCC, maintains Frequently Asked Questions and interagency guidance addressing suspicious activity reporting requirements, including SAR filings for potential structuring-related activity [11]. This guidance generally does not require institutions to file on transaction value alone; most institutions already consider multiple factors, timing, sources, and uses of funds, when evaluating possible structuring activity, consistent with a broader, longstanding regulatory interest in filing efficiency and effectiveness rather than blanket, duplicative filing expectations.

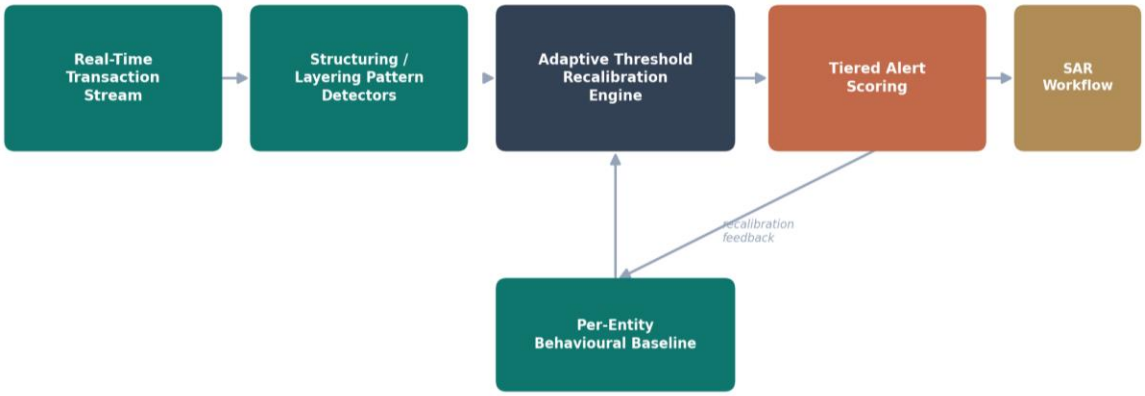
This regulatory direction is directly relevant to the case for adaptive thresholds: a system that reduces false-positive volume while preserving or improving detection of genuine structuring and layering activity is well aligned with regulators' stated interest in filing efficiency and effectiveness, provided institutions can demonstrate, through validation and documentation, that detection quality has not been compromised in pursuit of alert-volume reduction.

The Scale of the Underlying Reporting Burden

The reporting infrastructure that any structuring/layering detection system ultimately feeds is substantial. Bank Secrecy Act filings, including those generated by structuring and layering detection specifically, are widely reported to feed directly and substantially into actual law-enforcement outcomes: BSA data has long been cited by the IRS Criminal Investigation division and the FBI as supporting a significant share of financial-crime investigations and prosecutions [9]. This underscores that the false-positive problem this article addresses is not merely an internal cost-efficiency question but one with direct implications for the quality of the information reaching law enforcement.

System Architecture

Figure 1 presents the architecture such a system requires, synthesising design elements described across the published sources this article draws on. A real-time transaction stream feeds dedicated structuring/layering pattern detectors, rule- or graph-based logic specifically designed to identify sequences of below-threshold transactions or fund movement across related accounts, together with a per-entity behavioural baseline built from that entity's own transaction history. Both feed an adaptive threshold recalibration engine, the component distinguishing this architecture from a static rule-based system, which adjusts detection thresholds dynamically rather than applying a single, manually set value uniformly across all entities and time periods.



Graph-based analytics play a specific, well-documented role for layering detection within this architecture. An applied description of graph-based AML analytics notes that structuring shows up in a graph as a single source node fanning out into multiple low-value transactions within a short time window, a pattern that becomes visible once activity is modelled as a connected network of accounts, customers, and transactions rather than as isolated rows in a spreadshee. This graph-native view is specifically why layering, which by definition involves fund movement across multiple linked accounts, is difficult for single-transaction, threshold-only rules to detect reliably.

Streaming, real-time architectures for this purpose are increasingly documented in vendor and industry literature as a departure from the batch-oriented, overnight-run processing that has historically characterised transaction monitoring; one industry description specifically notes that batch graph analysis can miss layering chains and mule-network restructuring that occurs faster than the batch cycle runs, since the underlying network may have already reorganised before a batch job executes, whereas continuous, streaming analysis executing against the live transaction stream avoids this specific gap.

Relationship to the Graph-Feature Alert-Optimisation Approach

The specific mechanism underlying the graph-feature alert-optimisation study cited in Table 1 in Section 4 is directly relevant to the recalibration engine shown in Figure 1: that study layers novel, graph-derived features on top of an existing rule-based alert model to improve alert-scoring quality without discarding the underlying rules, explicitly framed as reducing false positives while preserving the auditability and explainability regulators expect from rule-based systems [8]. This is conceptually related to the recalibration function this article's architecture assigns to the adaptive threshold engine: both approaches aim to make detection more precise using information beyond a fixed, manually set threshold, whether through recalibrated thresholds or through supplementary learned features layered on top of existing rules, rather than requiring a compliance analyst to periodically and manually reset fixed thresholds as transaction patterns evolve.

Implementation Considerations for Smaller Institutions

None of the sources reviewed in this article specifically address implementation cost or feasibility for smaller community financial institutions, an important caveat given this article's broader focus. What can be stated with reasonable confidence, based on the architecture described across the

sources reviewed, is that an adaptive threshold recalibration engine of the kind shown in Figure 1 requires, at minimum, a data pipeline capable of maintaining a rolling, per-entity transaction history (the behavioural baseline in Figure 1), a real-time or near-real-time scoring capability, and a mechanism for periodically validating that recalibrated thresholds have not drifted away from acceptable detection performance. These are the same categories of technical and governance infrastructure discussed in relation to below-supervisory-threshold model-risk monitoring generally; nothing in the sources reviewed here suggests structuring/layering-specific adaptive thresholds require materially different infrastructure from adaptive credit-risk or general fraud-detection models, though this inference is the article's own reasoning rather than a claim directly evidenced in the cited sources.

Published Evidence on Adaptive Threshold Techniques

This section reviews specific, cited results from published studies evaluating adaptive or dynamic threshold approaches against conventional, fixed-threshold baselines. Table 1 summarises the studies and their reported results.

Table 1. Published and Industry-Reported Results for Adaptive Threshold and Related Techniques

Study / Source	Method	Reported Result
Weber et al., Anti-Money Laundering in Bitcoin: Experimenting with Graph Convolutional Networks for Financial Forensics, KDD 2019 Workshop on Anomaly Detection in Finance [7]	Random Forest and Graph Convolutional Network classifiers evaluated on the labelled Elliptic Bitcoin transaction dataset, distinguishing licit from illicit transactions	Random Forest achieved the strongest reported results among tested models, with precision above 0.95 and F1-score above 0.80 in identifying illicit transactions; graph-based features provided a complementary, if smaller, improvement over transaction-level features alone
Anti-Money Laundering Alert Optimization Using Machine Learning with Graphs, arXiv preprint, December 2021 [8]	Supervised machine-learning alert-scoring model augmented with novel graph-derived features, layered on top of an existing rule-based AML alert system	A 5-percentage-point improvement in F1 score from including graph-derived features, against a cited industry baseline false-positive rate of 95–98%
Industry-vendor analysis of AI-augmented transaction monitoring deployments [5][6]	Vendor-reported deployments combining behavioural profiling and adaptive scoring in production environments	Documented false-positive-rate reductions of 50%–60% at adopting institutions; some vendor claims cited up to 70%–80% under controlled conditions

The two academic sources in Table 1 are methodologically distinct from the vendor-reported figures and from each other, and their results should not be treated as directly comparable point estimates of a single, universal effect size. The Weber et al. study [7] evaluates graph-based and ensemble classifiers specifically for illicit-transaction identification on a labelled cryptocurrency dataset, reporting standard classification metrics rather than a false-positive-reduction figure against a rule-based baseline. The December 2021 alert-optimisation study [8] is more directly relevant to conventional AML transaction-monitoring alert systems, layering graph-derived features on top of an existing rule-based alert model rather than replacing it outright.

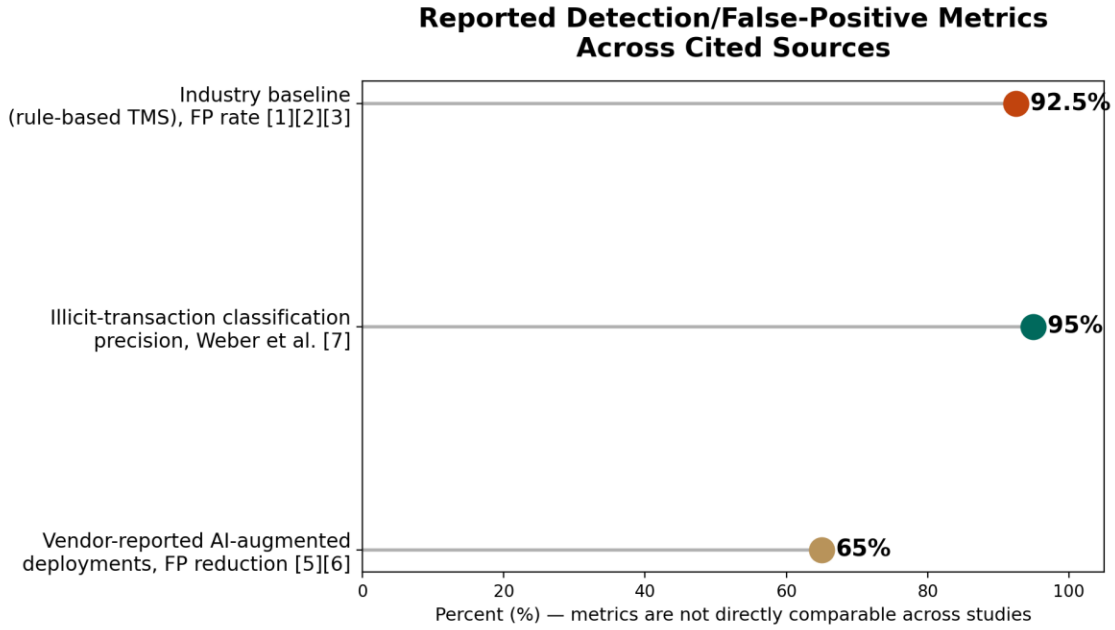


Figure 2 places these figures side by side for reference, while deliberately using distinct colours and bracketed citation numbers to flag that these values come from different studies using different methods and populations, and are not directly comparable as a single before-and-after comparison of one system.

Detection-Technique Comparison

Table 2 situates adaptive threshold recalibration relative to the other detection-technique families referenced across the sources reviewed in this article, summarising each technique's core mechanism and the specific evidence cited for it.

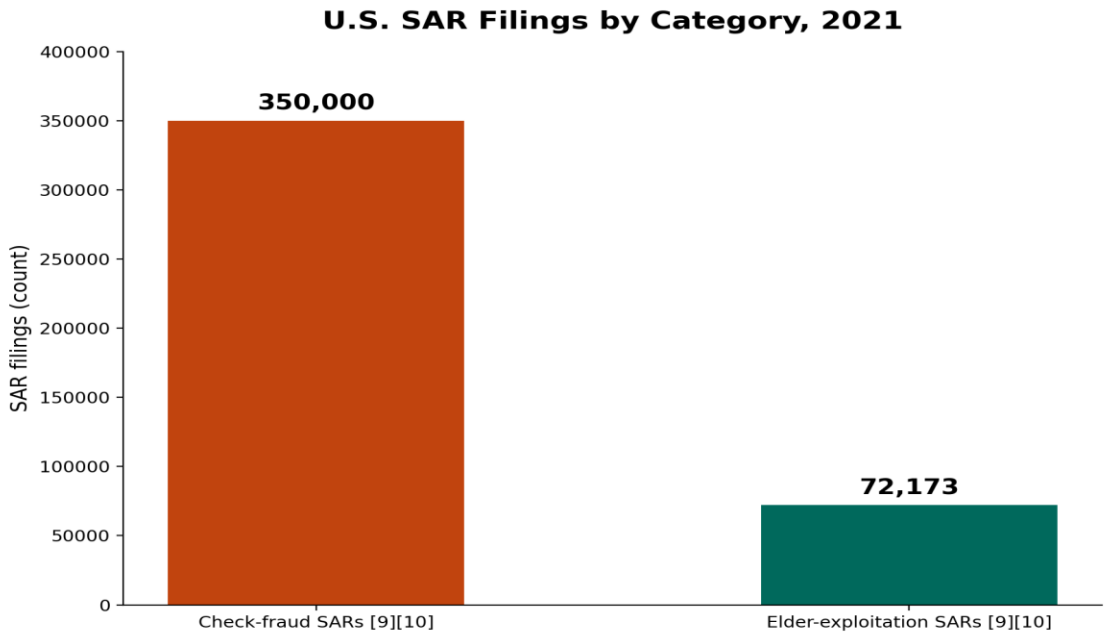
Table 2. Comparison of Detection Techniques Referenced in This Article

Technique	Core Mechanism	Cited Evidence
Static rule-based thresholds	Fixed transaction-value or frequency rules, manually set and periodically revised	90–95% false-positive rate industry-wide [1][2][3]

Technique	Core Mechanism	Cited Evidence
Graph-based network analytics	Models accounts/transactions as a connected graph to reveal fan-out and layering chains	Qualitatively documented as revealing patterns invisible to row-level analysis [11]
Adaptive/dynamic threshold recalibration	Thresholds computed per-entity and per-time-window from a continuously updated baseline	Precision above 0.95, F1 above 0.80 on labelled illicit-transaction data [7]; 5 pp F1 improvement from graph-derived features [8]
AI-augmented / behavioural-profiling systems (general)	Machine-learned behavioural profiles combined with adaptive scoring, in production use	50–80% FP reduction reported by adopting institutions [5][6]

The Broader Statistical Picture

Beyond the false-positive-rate literature reviewed in Section 4, FinCEN's own published SAR data provides useful context on the scale and composition of the suspicious-activity reporting environment that any structuring/layering detection system operates within. Figure 3 presents filing counts for two specific SAR categories, check fraud and elder financial exploitation, both drawn directly from FinCEN-sourced reporting for 2021.



Check-fraud SARs totalled roughly 350,000 in 2021 [9][10], while elder-financial-exploitation SARs totalled roughly 72,173 in 2021 [9][10]. Both categories represent substantial, recurring shares of overall fraud-related SAR filing volume, distinct from the structuring and layering typologies this article's evidence base addresses specifically. Table 3 summarises the full set of statistics cited across this article for ease of reference.

Table 3. Summary of Cited Statistics

Statistic	Value	Source
Industry false-positive rate, rule-based TMS	90–95%	[1][2][3]
Financial institutions participating in FinCEN's 314(b) information-sharing program, end of 2020	7,000+ institutions	[9]
FATF-cited global AML compliance cost, annual	>$180 billion	[5]
Total SARs filed, FY2020 (all filer types)	~2.5 million	[9]
SARs referencing 314(b) narratives, 2020	17,384 (+9.6% vs. 2019)	[9]
Leading SAR category, 2020	Transactions below CTR threshold (318,867 SARs)	[9]
Business Loan SAR growth, 2020 (COVID-19 stimulus-related)	>513%	[9]
Check-fraud SARs, 2021	350,000	[9][10]
Elder-exploitation SARs, 2021	72,173	[9][10]

Two observations follow from this broader picture. First, the composition of SAR filings is not uniform: check fraud and elder financial exploitation represent substantial filing volume in their own right, and are distinct underlying typologies from structuring and layering specifically, meaning a detection system tuned narrowly to structuring/layering patterns should not be assumed to generalise to these other categories without separate validation. Second, the scale of filing volume, in the hundreds of thousands to millions annually across categories, reinforces why even a modest percentage improvement in false-positive rate translates into a very large absolute reduction in investigative burden industry-wide.

What the Evidence Does and Does Not Establish

It is important to state plainly what the sources reviewed in Section 4 do and do not demonstrate, particularly given this article's specific focus on structuring and layering rather than AML transaction monitoring generally. None of the three sources in Table 1 report results specifically isolated to structuring and layering detection as opposed to AML/fraud detection more broadly; the Weber et al. study [7] evaluates illicit-transaction classification generally on cryptocurrency transaction data rather than money-laundering typologies specific to traditional banking, and the alert-optimisation study [8] evaluates graph-derived feature improvements to alert scoring generally rather than structuring/layering typologies specifically. The application of adaptive thresholds to structuring and layering detection specifically has not, to the best of the search evidence gathered for this article, been validated in a peer-reviewed study isolating that specific typology pair with a rigorous, published before-and-after comparison.

Second, none of the reviewed sources report results from a live, production, multi-institution deployment specifically at community banks, credit unions, or CDFIs, the institution types most resource-constrained in their ability to independently validate vendor claims. Neither the Weber et al. study [7] nor the alert-optimisation study [8] details results at community-financial-institution scale specifically, and the specific size and type of institutions underlying the alert-optimisation study's data are not detailed in the available summary.

Third, the FinCEN SAR volume and false-positive-rate statistics cited in Sections 1, 2, and 5 describe the current state of the problem this article addresses, establishing that the cost and scale of the false-positive problem are real and well documented, but they are not themselves evidence that any specific adaptive-threshold technique resolves that problem at a particular institution; that causal link depends on the technique-specific studies in Table 1 and their stated limitations above.

Table 4. Evidence-Strength Assessment for Key Claims in This Article

Claim	Evidence Status
90–95% industry false-positive baseline	Well established — confirmed across multiple independent sources [1][2][3]
FinCEN SAR/314(b) volume figures	Well established — primary FinCEN-sourced data [9][10]
Strong illicit-transaction classification from graph/ensemble methods	Published, peer-reviewed, but evaluated on cryptocurrency data, not structuring/layering specifically [7]
F1 improvement from graph-derived alert-scoring features	Published preprint, but not specific to structuring/layering [8]

Claim	Evidence Status
50–80% FP reduction from AI-augmented deployment	Industry-reported, vendor-influenced, method transparency limited [5][6]
Adaptive thresholds improve structuring/layering detection specifically	Not directly evidenced in the sources reviewed — plausible extrapolation only

Evaluating Vendor and Industry Claims

Vendor and industry-reported false-positive-reduction figures, such as those summarised in Table 1 and Table 3, warrant a cautious interpretation rather than being treated as directly comparable to peer-reviewed, methodologically transparent academic results. Industry reporting on this topic generally does not disclose the specific dataset, validation methodology, or population against which a vendor's own performance claims were measured, making independent verification difficult. Without institution-level data on which specific institutions achieved a given reduction, and under what conditions, industry-wide vendor claims should be treated as directional rather than as rigorously established performance benchmarks.

A responsible institution evaluating this technology should therefore request from any vendor: the specific dataset and population the vendor's own performance claims were validated against; whether that validation isolated structuring/layering detection specifically or measured AML detection performance more broadly; and independent, or at minimum transparently documented, evidence of the vendor's reduction claims, rather than relying on the industry-wide ranges surveyed in this article as a substitute for institution-specific validation.

Conclusion

This article has surveyed published, cited evidence on adaptive threshold recalibration for real-time detection of structuring and layering patterns in BSA/AML transaction monitoring. The evidence base establishes, with reasonable confidence given multiple independent, cross-referencing sources, that rule-based transaction monitoring carries a false-positive rate of approximately 90 to 95 percent industry-wide [1][2][3], that this imposes a substantial and well-documented cost (an estimated global AML compliance cost above $180 billion annually according to FATF-cited figures [5]), and that suspicious activity report filing volume, alongside related information-sharing activity, has grown substantially in recent years [9][10]. Published academic sources report strong illicit-transaction classification performance from graph-based and ensemble methods on labelled datasets, and a measurable F1-score improvement from graph-derived alert-scoring features layered on an existing rule-based system [7][8], alongside vendor-reported false-positive reductions in a comparable range under production conditions [5][6], though these results derive from AML detection and cryptocurrency-forensics studies rather than from a study isolating structuring and layering detection specifically.

The honest conclusion this evidence supports is one of considerable promise alongside a specific, identifiable evidence gap, summarised plainly in Table 4: the general technique class is well validated in adjacent and overlapping contexts, but institution-specific, typology-specific

validation remains the necessary next step before any specific quantified performance claim should be treated as established for structuring and layering detection in particular. Institutions and researchers pursuing this technology should prioritise closing precisely this gap, a published, peer-reviewed, multi-institution study isolating adaptive threshold performance for structuring and layering detection specifically, rather than extrapolating from the related but distinct results this article has surveyed.

Three specific research priorities follow from the evidence gaps identified in Section 6. First, a peer-reviewed evaluation isolating structuring and layering detection specifically, rather than AML detection generally, using a labelled dataset with confirmed structuring/layering cases, would directly close the largest gap identified in Table 4. Second, replication of the graph-derived-feature result in [8] at a set of community banks, credit unions, or CDFIs specifically, the institution types this article's broader research programme is centred on, would establish whether the reported F1-score improvement generalises to smaller, less data-rich institutions. Third, independent, third-party validation of vendor false-positive-reduction claims [5][6], through the kind of neutral, multi-institution benchmarking process, would meaningfully strengthen the currently vendor-reported evidence base for AI-augmented monitoring deployments.

References

[1] [1]
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.

[2] Jullum, M., Løland, A., Huseby, K., Ånonsen, G., & Søyland, K. (2020). Detecting money laundering transactions with machine learning. *Journal of Money Laundering Control, 23*(1), 173–186. https://doi.org/10.1108/JMLC-02-2019-0015

[3] Weber, M., Domeniconi, G., Chen, J., Weidele, D. K. I., Bellei, C., Robinson, T., & Leiserson, C. E. (2019). Anti-money laundering in Bitcoin: Experimenting with graph convolutional networks for financial forensics. *arXiv*.

[4] Eddin, A. N., Bono, J., Aparício, D., Polido, D., Ascensão, J. T., Bizarro, P., & Ribeiro, P. (2021). Anti-money laundering alert optimization using machine learning with graphs. *arXiv*.

[5] [5]
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.

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

[7] Levi, M., & Reuter, P. (2006). Money laundering. *Crime and Justice, 34*(1), 289–375. [8] Takáts, E. (2011). A theory of "crying wolf": The economics of money laundering enforcement.

The Journal of Law, Economics, and Organization, 27(1), 32–78. https://doi.org/10.1093/jleo/ewp022

[8] Ferwerda, J. (2009). The economics of crime and money laundering: Does anti-money laundering policy reduce crime? *Review of Law & Economics, 5*(2), 903–929.

[9] Unger, B., & van der Linde, D. (2013). Research handbook on money laundering. *Edward Elgar Publishing*.

[10] Weber, M., Domeniconi, G., Chen, J., Weidele, D. K. I., Bellei, C., Robinson, T., & Leiserson, C. E. (2019). Anti-money laundering in Bitcoin: Experimenting with graph convolutional networks for financial forensics. *arXiv*.

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