ASRC Procedia
Global Perspectives in Science and Scholarship
Florida, USA

Volume: 1; **Issue:** 1
Pages: 1740 - 1775
Published: 29 April 2025

open access

1st Global Research and Innovation Conference 2025,

April 20–24, 2025, Florida, USA

Machine Learning–Based AML/KYC Transaction Monitoring for Suspicious Activity Detection and Compliance Risk Reduction in Digital Banking

Sazzadul Islam[1];

[1]. *Master of Science in Big Data Analytics, Bay Atlantic University, Washington, USA; Email: sazzadulislam352@gmail.com*

Doi: 10.63125/r9c8q813

Peer-review under responsibility of the organizing committee of GRIC, 2025

Abstract

This study addresses the persistent problem in digital banking that traditional rule-based AML/KYC transaction monitoring generates high alert noise and inconsistent investigative outcomes, increasing compliance cost and residual regulatory exposure. The purpose was to quantify how machine learning enabled transaction monitoring capabilities contribute to suspicious activity detection and perceived compliance risk reduction in a cloud-enabled enterprise digital banking case context. Using a quantitative cross-sectional, case-based design, data were collected via a structured 5-point Likert survey from professionals embedded in AML operations, compliance oversight, KYC/CDD, risk, and AML technology functions (N = 168 valid responses; mean AML experience = 6.2 years, SD = 2.9; role distribution led by transaction monitoring analysts at 42.3%). Key variables included ML Detection Effectiveness (MDE), Data Quality and Feature Readiness (DQF), , Analyst Trust and Adoption (ATA), and three domain indices: Alert Quality and Workload Impact (AQWII), Explainability Audit Readiness and Model Governance Evidence (EARMGE), and Typology Coverage and Adaptability Results (TCAR), with Compliance Risk Reduction (CRR) as the dependent variable. The analysis plan applied descriptive statistics, reliability testing (Cronbach's alpha), Pearson correlations, and multiple regression to estimate unique predictors of CRR. Results showed strong overall agreement that ML monitoring improved outcomes (CRR M = 4.14, SD = 0.52; MDE M = 4.18, SD = 0.54; AQWII M = 4.12, SD = 0.55), and all constructs met good to excellent internal consistency (a = 0.83–0.91). CRR correlated most strongly with MDE (r = 0.72, p < .001) and AQWII (r = 0.69, p < .001), indicating that detection strength and workload-relevant alert quality moved together with compliance benefit. The regression model was significant and explanatory (R^2 = 0.68; F(8,159) = 42.11, p < .001), with headline effects from MDE (β = 0.31, p < .001) and AQWII (β = 0.23, p < .001), alongside meaningful contributions from ATA (β = 0.18, p = .002), DQF (β = 0.15, p = .007), SIA (β = 0.14, p = .015), MET (β = 0.12, p = .029), EARMGE (β = 0.13, p = .018), and TCAR (β = 0.11, p = .042).

Keywords

AML/KYC, Machine Learning, Transaction Monitoring, Suspicious Activity Detection, Compliance Risk Reduction;

INTRODUCTION

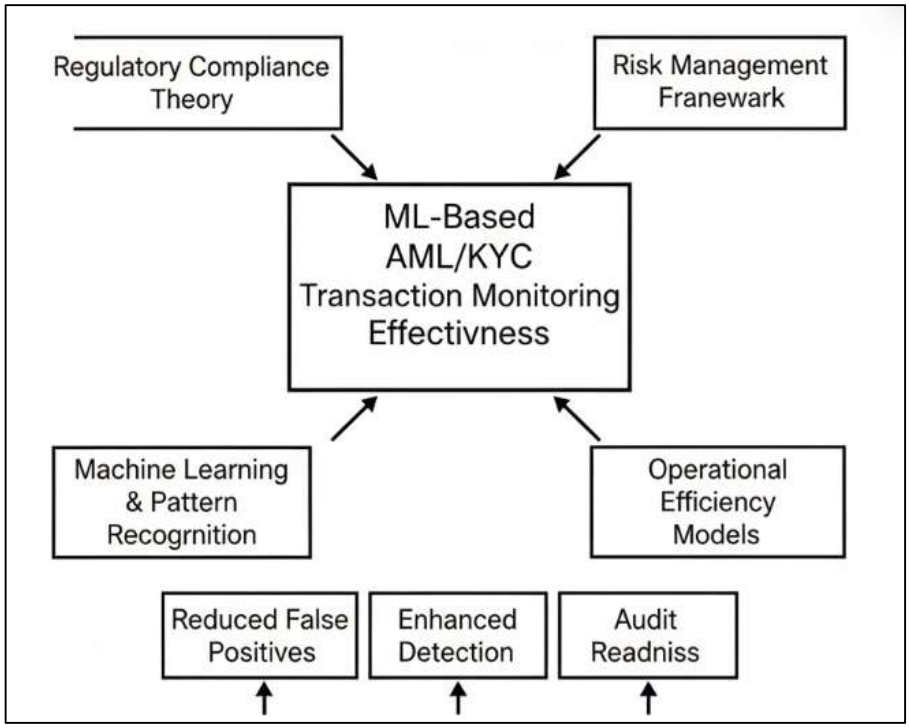
Anti–money laundering (AML) refers to the set of regulatory and operational controls used by financial institutions to identify, deter, and report activity associated with disguising illicit proceeds as legitimate funds, while "know your customer" (KYC) denotes the identity verification and customer due-diligence processes used to establish beneficial ownership, customer risk profiles, and ongoing monitoring obligations (Omar & Johari, 2015). In contemporary digital banking, AML/KYC is operationalized largely through transaction monitoring, a continuous screening function that flags potentially suspicious behavior for investigation, escalation, and suspicious activity reporting (SAR) workflows (Abdallah et al., 2016). Suspicious activity detection, in this setting, can be understood as the classification or ranking of events (transactions, accounts, relationships, sequences, or cases) according to their likelihood of reflecting money laundering typologies or related financial crime patterns (Kaur et al., 2020).

The international significance of AML/KYC transaction monitoring is shaped by the scale, speed, and cross-border nature of digital payments, where a single payment rail can route value across multiple jurisdictions within seconds, and where compliance obligations intersect with financial inclusion, consumer protection, and systemic stability (Bahnsen et al., 2016). This combination creates a high-stakes analytic environment: false negatives elevate legal and reputational exposure, while false positives impose operational burden, friction, and customer harm. As AML monitoring volumes expand, institutions increasingly seek data-driven methods that can learn patterns from historical cases and adapt to heterogeneous signals across customers, channels, and geographies. Survey research in suspicious transaction detection highlights how machine learning (ML) methods have been positioned as complements or replacements for rule-based monitoring, by improving sensitivity to complex patterns and by reducing alert noise through probabilistic scoring. A broad view of AML solutions also frames transaction monitoring as a rare-event detection problem, where sampling strategy and model evaluation are inseparable from governance and audit expectations, given that laundering labels are scarce and delayed (Kaur et al., 2020). The literature further situates AML in the wider financial fraud detection landscape, where data mining and classification frameworks emphasize the need to align analytics with business processes, investigative capacity, and measurable outcomes such as precision, recall, and cost-weighted error. Within this research space, AML/KYC monitoring becomes not only a technical classification task, but an institutional risk-control mechanism that must withstand regulatory scrutiny across borders while functioning at scale in real-time digital banking contexts (Carcillo et al., 2018).

Machine learning–based AML/KYC transaction monitoring commonly targets the same operational objective as traditional rule engines—identifying suspicious activity—but does so through statistical learning from historical data, enabling probabilistic scoring, multivariate interactions, and pattern discovery beyond fixed thresholds. In fraud and compliance analytics, the core challenge is rarely "whether" an algorithm can detect anomalies, but "how" it performs under real constraints: extreme class imbalance, concept drift in criminal tactics, verification latency in label confirmation, and limited investigative capacity (Bhattacharyya et al., 2011). These constraints have been extensively documented in adjacent fraud detection domains, especially payment and card transactions, where researchers show that realistic modeling assumptions, delayed supervision, and evolving behavior strongly influence observed performance. Comparative studies in credit-card fraud detection, for instance, illustrate how algorithm choice and feature design can produce materially different results under imbalance and shifting base rates, highlighting the importance of evaluation design and the use of appropriate metrics. In operational environments, the decision context is frequently cost-sensitive: alerts consume analyst time, case creation triggers downstream processes, and false alarms generate friction. Feature engineering research demonstrates that transaction aggregation strategies—how events are summarized over time windows and customer histories—can substantially alter detection power, supporting the view that monitoring quality is often driven as much by representation as by model family (Kaur et al., 2020).. Sequential modeling studies extend this logic by treating fraud detection as sequence classification, showing that temporal ordering and recurrent patterns may carry incremental value beyond static attributes. Hybrid paradigms combining supervised and unsupervised approaches have also been proposed to leverage anomaly scores while still learning from labels where available,

aligning well with compliance settings where labeled laundering cases are limited and investigative confirmation may lag (Haque & Arifur, 2020; Rauf, 2018). In AML/KYC monitoring specifically, empirical evidence indicates that learning algorithms can be sensitive to sampling decisions and the definition of the target variable, since "suspicious" may be represented by internal alert outcomes, escalated cases, or SAR filings—each reflecting different points in the compliance pipeline (Carcillo et al., 2021; Haque & Arifur, 2021; Ashraful et al., 2020). These findings motivate research designs that explicitly connect ML outputs to compliance risk reduction, rather than treating classification performance as an abstract objective detached from monitoring workflows and analyst capacity (Dal Pozzolo et al., 2018; Fokhrul et al., 2021).

Figure 1: Integrated Model of ML-Based Suspicious Transaction Detection



This study aims to evaluate machine learning–based AML/KYC transaction monitoring as an integrated compliance capability within a digital banking case context using a quantitative, cross-sectional, case-study–based design. The first objective is to measure how key monitoring constructs are perceived and operationalized in the bank's environment, including the effectiveness of suspicious activity detection, the clarity of KYC-informed customer risk profiling, the usability of alerts for investigation, and the perceived reduction of compliance risk. The second objective is to quantify relationships among these constructs through descriptive statistics and correlation analysis, identifying which monitoring dimensions move together and which appear independent, thereby clarifying the internal structure of ML-enabled monitoring capability. The third objective is to test hypothesized effects using regression modeling, examining whether improvements in model-driven detection, alert quality, and governance practices are associated with measurable reductions in compliance burden and risk exposure indicators within the case setting. The fourth objective is to assess operational credibility by incorporating study-specific evidence measures that reflect real monitoring work, including an alert quality and workload impact index that captures how alerts translate into analyst effort, an explainability and audit readiness evidence measure that reflects documentation and governance artifacts, and a typology coverage and adaptability assessment that reflects how well the monitoring approach aligns with the laundering behaviors observed in the case context. The fifth objective is to produce a structured empirical account of how ML-based monitoring can be evaluated as a compliance control system, connecting model outputs to investigation workflows and documenting how analytic performance, operational workload, and governance readiness jointly shape monitoring effectiveness.

Together, these objectives support a complete measurement-and-testing framework aligned to the study's hypotheses and research questions within a single, coherent quantitative design.

LITERATURE REVIEW

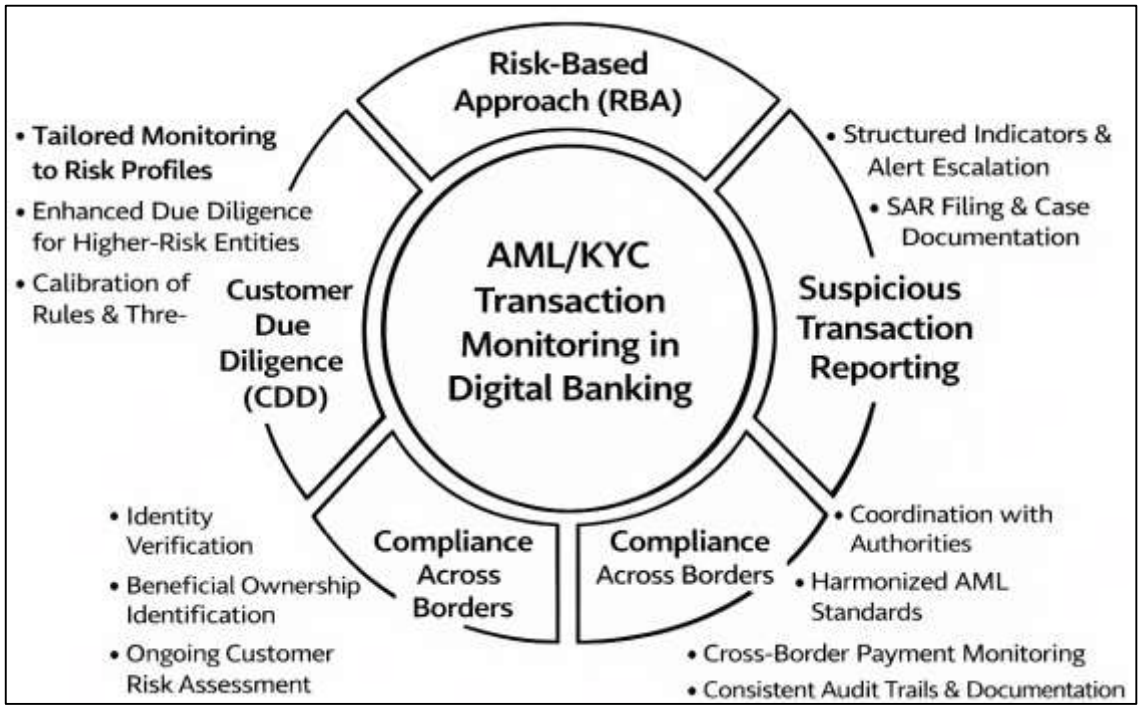
The literature on machine learning–based AML/KYC transaction monitoring in digital banking spans regulatory foundations, operational monitoring practices, and computational approaches for detecting suspicious activity at scale. At the regulatory and operational level, prior studies describe how AML/KYC obligations are implemented through customer due diligence, risk scoring, and ongoing transaction surveillance, where alerts trigger investigative workflows that can culminate in suspicious activity reporting and internal risk escalation. This stream emphasizes that transaction monitoring is not a purely technical exercise; it is an enterprise control function that must be consistent, auditable, and aligned with policy requirements, governance standards, and documented decision trails. A second stream focuses on the limitations of conventional rule-based monitoring, commonly highlighting excessive false positives, rigid thresholds that fail to capture complex behavior, and heavy dependence on manual tuning that struggles to keep pace with evolving typologies and high-velocity digital channels. A third stream examines the application of machine learning methods—ranging from supervised classification and anomaly detection to hybrid and ensemble approaches—designed to learn patterns from historical cases and behavioral data, often under conditions of severe class imbalance and delayed outcome confirmation. Within this technical stream, researchers also stress the importance of feature engineering, temporal aggregation, customer-level baselines, and the integration of KYC attributes to strengthen detection and prioritize alerts. A fourth stream extends beyond transaction-level signals to relational and network-oriented modeling, arguing that money laundering behaviors often manifest through coordinated account activity and multi-hop fund movement, which motivates graph-based representations and case-centric analytics. A fifth stream addresses the governance layer of ML adoption, with particular attention to explainability, validation, model risk management, and audit readiness, reflecting the reality that compliance environments require defensible rationales for alert generation and consistent monitoring of model performance over time. Across these streams, the literature converges on a central theme: effectiveness in AML monitoring should be evaluated using both analytic performance and operational outcomes, including alert quality, investigation workload, and perceived compliance risk reduction. Guided by this perspective, the present review synthesizes prior work to identify dominant methods, recurring challenges, measurement strategies, and gaps that justify a quantitative, cross-sectional, case-study–based investigation using Likert-scale constructs tested through descriptive statistics, correlation analysis, and regression modeling.

AML/KYC Transaction Monitoring in Digital Banking

Anti–money laundering (AML) and know-your-customer (KYC) controls form a foundational part of the global financial-integrity architecture, translated into bank practice through customer due diligence (CDD), record retention, and ongoing transaction monitoring. In digital banking, transaction monitoring is the operational "bridge" between KYC information (identity, beneficial ownership, customer risk profiles) and the continuous stream of payments, transfers, and account activity that must be assessed for potential suspicion (Fahimul, 2022; Zaman et al., 2021). Monitoring systems typically convert raw events into structured indicators (e.g., transaction size and frequency, counterparties, geographic routing, channel use, and behavioral changes), then generate alerts that initiate investigative workflows, case documentation, and escalation decisions. The regulatory logic underlying these controls is international in scope: funds move rapidly across borders, products are delivered through multiple channels, and compliance obligations require firms to demonstrate consistent controls over both customers and transactions (Hammad, 2022; Hasan & Waladur, 2022). Within this setting, suspicious transaction reporting becomes an institutionalized reporting channel that depends on how "suspicion" is operationalized, how information is gathered and recorded, and how financial institutions coordinate with public authorities (Rashid & Sai Praveen, 2022; Arifur & Haque, 2022). A classic account of the suspicious transactions reporting system emphasizes that mandatory reporting regimes evolved alongside supervisory expectations and the widening range of obligated entities, placing reporting within a broader control chain that begins with detection and ends with intelligence use by competent authorities (He, 2005; Towhidul et al., 2022; Ratul & Subrato, 2022).

Digital banking intensifies these requirements because monitoring must operate at scale, often in near-real time, while maintaining audit trails that explain why an alert was generated and how it was resolved. As a result, transaction monitoring is not only an analytics function; it is also a governance function that structures documentation, decision rights, quality assurance, and accountability, all of which shape how banks demonstrate compliance performance to supervisors in cross-border environments.

Figure 2: AML/KYC Transaction Monitoring in Digital Banking



Regulatory expectations for AML/KYC monitoring are commonly expressed through the risk-based approach (RBA), which requires financial institutions to allocate CDD and monitoring effort proportional to the perceived risk of customers, products, delivery channels, geographies, and transaction patterns (Rifat & Jinnat, 2022; Rifat & Alam, 2022). Under RBA, "monitoring" is not a uniform screening exercise; it is designed to be differentiated and evidence-based, so that higher-risk relationships receive enhanced due diligence and more intensive surveillance, while lower-risk relationships receive simplified or standard controls (Abdulla & Majumder, 2023; Fahimul, 2023). A law-and-economics framing of RBA in the European AML context conceptualizes this arrangement as a principal–agent problem: regulators seek effective deterrence and detection outcomes, while institutions possess private information and must be incentivized to invest in monitoring that produces meaningful signals rather than symbolic compliance artifacts (Pellegrina & Masciandaro, 2009; Faysal & Bhuya, 2023; Habibullah & Aditya, 2023). At the international level, empirical work examining AML program design further situates RBA within a complex regime of standards, mutual evaluations, and enforcement variation, highlighting how outcomes depend on both rule design and the institutional capacity to implement controls consistently (Arnone & Borlini, 2010; Hammad & Mohiul, 2023; Haque & Arifur, 2023). In digital banking, these expectations translate into practical requirements such as customer risk scoring, scenario coverage aligned to typologies, documentation of investigative decisions, and the ability to justify monitoring intensity in relation to observed risk. Therefore, RBA functions as both a policy concept and a systems-engineering requirement, shaping how monitoring rules, models, thresholds, and workflows are configured, validated, and reviewed.

The RBA also introduces implementation challenges that are directly relevant to the credibility of transaction monitoring in digital banking. One challenge is conceptual consistency: if "risk" is interpreted differently across units or roles, monitoring intensity can become uneven, undermining the

logic of proportional controls and making outcomes difficult to defend during examination. Another challenge is operational calibration: RBA assumes that institutions can convert abstract risk categories into concrete monitoring behaviors (scenario design, alert thresholds, review frequency, escalation criteria), which requires clear governance and stable documentation practices. A focused analysis of the risk-based approach argues that common failures stem from inadequate risk modeling and weak translation of risk assessments into operational controls, leaving gaps between formal frameworks and real monitoring behavior (Jahangir & Mohiul, 2023; Rashid et al., 2023; Simonova, 2011). In addition, international compliance assessments emphasize that effectiveness depends on how standards are implemented and tested, not only on whether policies exist on paper; comparative analyses using mutual evaluation evidence highlight recurring weaknesses in due diligence, record practices, and suspicious reporting implementation across jurisdictions and obligated sectors (Omar & Johari, 2015). For digital banks, these realities shape regulatory expectations toward demonstrable monitoring outcomes: coherent risk profiling, traceable alert rationales, consistent case handling, and evidence that monitoring is aligned with the institution's stated risk appetite and typology exposure. Consequently, transaction monitoring in digital banking is best understood as a risk-governed control system in which analytics, workflow design, and documentation practices jointly determine compliance readiness and supervisory defensibility.

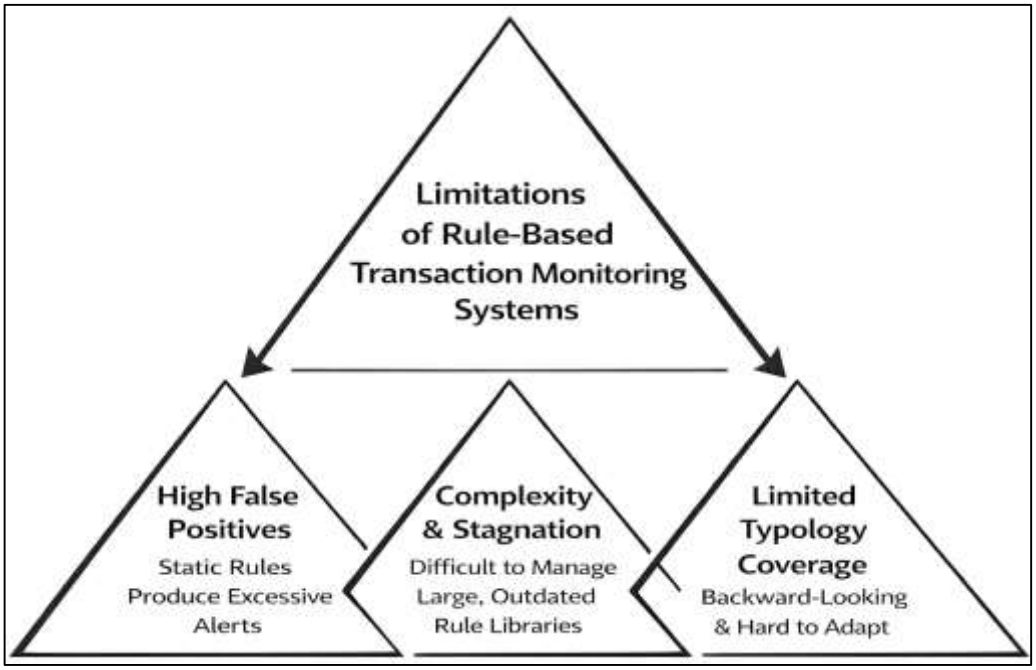
Rule-Based Transaction Monitoring Systems

Rule-based transaction monitoring (TM) remains a foundational approach in AML/KYC programs because it translates regulatory expectations and typology knowledge into explicit "if–then" scenarios (e.g., velocity thresholds, structuring rules, corridor/jurisdiction rules, and peer-group deviations). In practice, these rules operationalize risk appetite by converting financial behaviors into alert triggers that are explainable to auditors and easy to deploy across legacy core banking stacks. The same strengths, however, create structural constraints. First, rule logic is inherently reductive: it converts continuous, context-dependent customer behavior into discrete conditions, which means a large proportion of legitimate activity can resemble suspicious patterns when viewed only through static thresholds. This produces chronic false positives and "alert inflation," where analyst capacity is consumed by low-yield investigations rather than true suspicious activity identification. Evidence from payment-fraud monitoring environments, which share similar operational constraints with AML TM (high volume, extreme class imbalance, and tight response SLAs), shows how false positives become a dominant efficiency bottleneck and impose real operational costs on institutions (Rashid et al., 2023; Akbar & Farzana, 2023; Vorobyev & Krivitskaya, 2022). Second, rule sets accumulate over time: as institutions add scenarios to cover new typologies, rules overlap, interact, and create redundant triggers that are difficult to trace back to a single causal pattern. Third, rule-based TM often evaluates transactions at the "single-event" level (or limited windows), which weakens its ability to model behavioral context—precisely the context that compliance teams rely on when assessing whether a pattern is genuinely anomalous. Research on transaction aggregation demonstrates that individual transactions often lack sufficient signal, while engineered behavioral summaries across time windows capture meaningful differences between legitimate and illicit behaviors (Mostafa, 2023; Rifat & Rebeka, 2023; Whitrow et al., 2009). In AML TM, this implies that purely scenario-driven rule firing may miss the behavioral storyline that investigators need, while still generating excessive noise.

A second limitation is the human-dependence of rule lifecycle management: rules require continual tuning, exception-handling, and parameter calibration, typically driven by expert judgement and periodic back-testing. This dependence creates a tension between compliance defensibility and statistical validity (Jahangir & Hammad, 2024; Masud & Hammad, 2024). When thresholds are adjusted primarily to reduce alert volume, institutions risk "optimizing to workload" rather than improving detection quality; when thresholds remain fixed for long periods, drift in customer behavior, channel adoption, and criminal tactics degrades performance. Rule-based monitoring also struggles with feature expressiveness. Many AML rules are built around intuitive indicators (amount, frequency, geography, counterparty risk), but the interactions among these indicators can be non-linear and conditional on customer segment context. Hybrid fraud-monitoring research shows that combining supervised and unsupervised components can compensate for the weaknesses of single-paradigm systems, especially when fraud patterns evolve and labels are imperfect (Krivko, 2010). Translating this

insight to AML, rule-only architectures are likely to underperform in environments where laundering strategies are adaptive (e.g., micro-structuring, mule networks, rapid channel hopping). In addition, rule-based TM is frequently evaluated using compliance proxies—alert counts, SAR conversion, or qualitative QA—rather than statistically grounded performance metrics that separate precision, recall, and calibration. Approaches that emphasize engineered features and more structured modeling demonstrate that aggregation and model-based scoring can provide clearer discrimination than manual scenario stacking (Jha et al., 2012; Md & Sai Praveen, 2024; Rifat & Rebeka, 2024). From a governance perspective, rule systems appear transparent, but their transparency can be misleading: a large rule library may be "auditable" while still being empirically weak, because interpretability does not guarantee detection validity, nor does it guarantee stable performance across customer populations and time periods.

Figure 3: Rule-Based Transaction Monitoring Systems



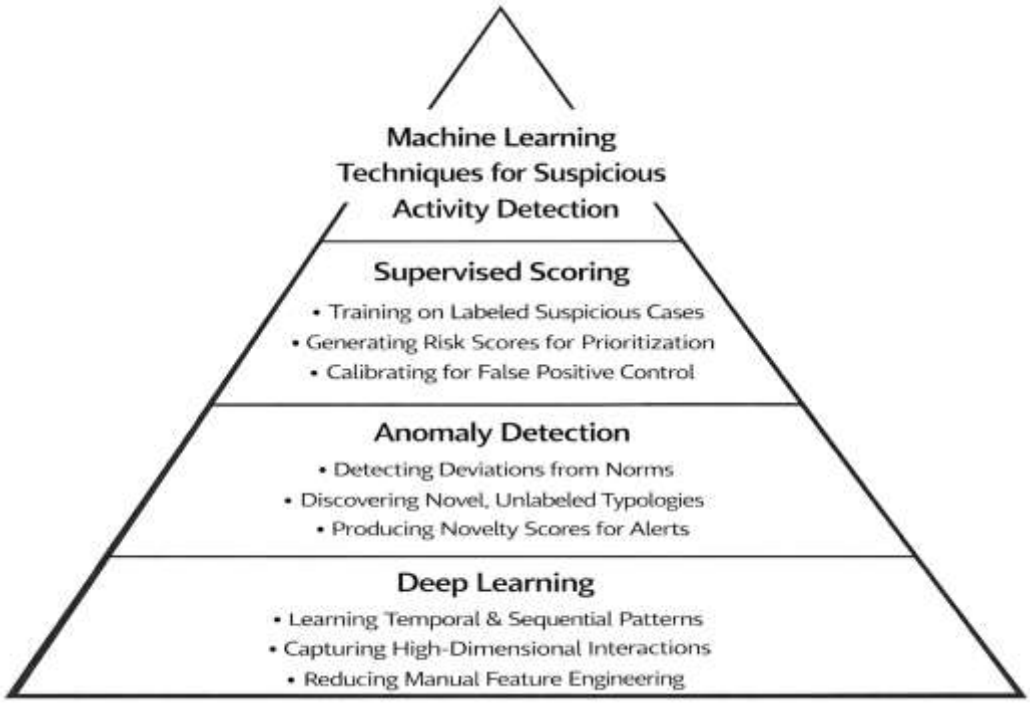
A third limitation concerns typology coverage and adaptability. Rule libraries tend to encode known typologies, meaning they are inherently backward-looking and coverage-driven: institutions add rules after regulators highlight a pattern, after internal cases emerge, or after peer enforcement actions shift expectations. As typologies diversify, rules become increasingly granular, and the monitoring program becomes difficult to scale without substantial staffing. Association-rule research in transactional fraud settings illustrates a related issue: it is possible to extract "normal behavior" patterns from transaction data, but translating such patterns into stable operational rules remains challenging because normal behavior itself varies across customers, merchants, and time windows (Sánchez et al., 2009). In AML TM, this variability fuels both false positives (legitimate customers triggering typology-like thresholds) and false negatives (true laundering that stays under thresholds or mimics common patterns). The operational impact is cumulative: analysts experience alert fatigue, escalation queues grow, and quality assurance becomes sampling-based rather than comprehensive. These dynamics can weaken compliance risk reduction because investigative attention becomes a scarce resource allocated by triage rather than by true risk prioritization. The implication for this study's context—machine learning–based suspicious activity detection and compliance risk reduction in digital banking—is that the limitations of rule-based TM are not simply "technical." They are socio-technical: they emerge from the interaction of static logic with evolving behavior, from metric choices that emphasize volume control, and from governance requirements that reward explainability but still demand demonstrable effectiveness. Therefore, the literature supports a shift from rule-only monitoring toward evidence-

driven, model-assisted architectures that preserve auditability while improving discrimination power and reducing workload through risk-ranked alerting rather than threshold firing alone.

Machine Learning Techniques for Suspicious Activity Detection (SAD)

Machine learning (ML) techniques for suspicious activity detection (SAD) in AML/KYC transaction monitoring are typically organized around how labels are obtained and how "suspicion" is operationalized into a measurable target. In supervised learning settings, historical investigative outcomes (e.g., alerts escalated to confirmed internal cases or externally reported events) are treated as labeled examples that enable classifiers to learn discriminative patterns from transactional and customer attributes. In semi-supervised and weakly supervised settings, labels may be incomplete, delayed, or noisy because ground truth is revealed through lengthy investigations, which increases the importance of robust model evaluation and careful metric selection. This is especially visible in financial crime detection contexts, where class imbalance is extreme and accuracy becomes a misleading indicator of performance; evaluation practices therefore emphasize rank-based metrics (e.g., ROC analysis) that make trade-offs between true positive rates and false alarm rates explicit for operational decision-making. A widely cited methodological basis for this evaluation logic is ROC analysis, which supports comparing models under different thresholds and aligning model choice with the institution's tolerance for false positives and missed suspicious events (Fawcett, 2006). In transaction monitoring, this threshold sensitivity is not a minor technical detail; it directly determines alert volume, investigator workload, and escalation behavior. Accordingly, supervised ML approaches in AML frequently aim to produce calibrated risk scores that support prioritization rather than binary "flag/no-flag" decisions, allowing institutions to tune operational cutoffs to capacity constraints while preserving sensitivity to higher-risk patterns. This framing also encourages feature engineering that captures behavior over time—such as rolling aggregates, peer-group deviations, and sudden changes in activity—so that models can infer the narrative structure of suspicious behavior rather than relying on single-transaction triggers.

Figure 4: Machine Learning Techniques for Suspicious Activity Detection (SAD)

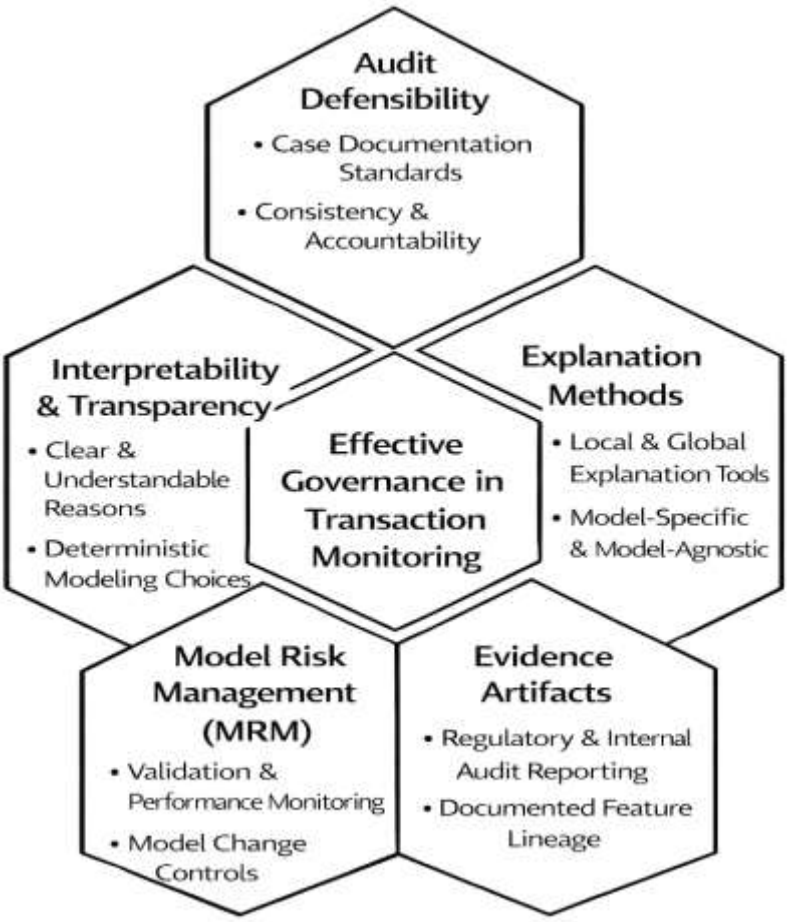


Explainable AI (XAI) and Model Risk Management in AML/KYC

Explainable artificial intelligence (XAI) has become a central requirement for deploying machine learning in AML/KYC transaction monitoring because compliance decisions must be traceable, reviewable, and defensible to multiple stakeholders. In digital banking, suspicious-activity alerts

trigger investigative actions that create formal records, and the reasoning behind those alerts must be communicable in operational language rather than only in statistical terms. XAI is therefore treated as an interface between predictive scoring and regulated decision processes: it supports analyst comprehension, case narrative formation, quality assurance review, and supervisory examination. Within this view, explanation is not limited to a single technique (Azam &Amin, 2024); it includes the selection of interpretable features, the presentation of local reasons for why a specific transaction or customer is flagged, and global summaries that show which factors drive risk at scale. Survey work on XAI frames explainability as a multi-dimensional concept that spans interpretability, transparency, and the ability to provide human-meaningful rationales under operational constraints, with emphasis on how explanation methods differ by model type and by user need (Adadi & Berrada, 2018; Sai Praveen, 2024; Shehwar & Nizamani, 2024). For AML/KYC monitoring, these user needs are distinct because the "consumer" of the explanation includes investigators, compliance managers, internal auditors, and regulators, all of whom may require different levels of detail and different formats of evidence. A practical implication at the systems level is that transaction monitoring explanations must connect data elements (KYC attributes, behavioral aggregates, counterparty patterns, channel indicators) to typology-consistent narratives that investigators can document in case files. XAI literature also highlights that explanations can be used to identify unstable or spurious model logic, where a model may achieve high discrimination while relying on artifacts that do not reflect meaningful risk; in regulated monitoring, this risk becomes a governance issue because it can weaken audit defensibility and create inconsistent treatment across customer segments (Adadi & Berrada, 2018).

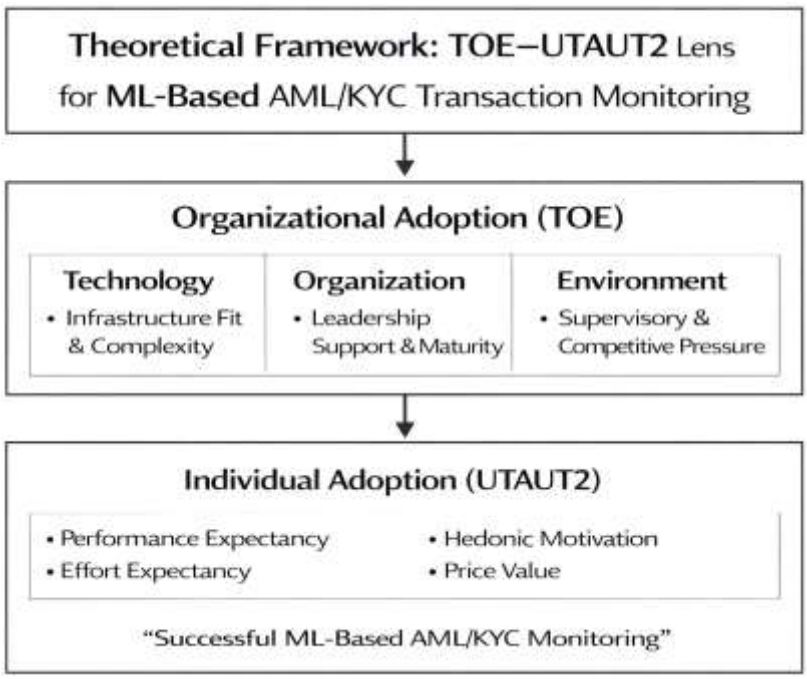
Figure 5: Explainable AI (XAI) and Model Risk Management in AML/KYC Transaction Monitoring



TOE–UTAUT2 Lens for ML-Based AML/KYC Transaction Monitoring

The theoretical framing for ML-based AML/KYC transaction monitoring in digital banking is strengthened when adoption is modeled at two complementary levels: the organizational level (whether the bank adopts and institutionalizes ML monitoring as an operational compliance capability) and the individual level (whether AML analysts, investigators, and compliance managers accept, rely on, and routinely use ML-generated signals in alert handling). At the organizational level, the Technology–Organization–Environment (TOE) perspective is frequently applied to explain why firms initiate, adopt, and routinize complex innovations, particularly when the innovation requires changes in data infrastructure, governance, and cross-functional workflows. Evidence on innovation assimilation demonstrates that adoption is better represented as a staged process—moving from initiation to adoption and routinization—where both internal readiness and external pressures shape whether the innovation becomes embedded in daily operations (Zhu et al., 2006). For AML/KYC, this staged logic maps cleanly onto the monitoring pipeline: early proof-of-concept scoring, controlled deployment in limited corridors/products, and eventual routinization through standardized model governance, alert triage policies, and investigation playbooks. TOE also provides a structured set of explanatory blocks. The technological context captures perceived relative advantage (e.g., improved discrimination and prioritization), compatibility with existing TM infrastructure (e.g., data ingestion from core and payment rails), complexity (e.g., feature engineering and explainability), and trialability. The organizational context captures compliance leadership support, budget and skills availability, data governance capacity, and process maturity. The environmental context captures supervisory scrutiny, peer adoption pressure, vendor ecosystems, and legal expectations around documentation. In enterprise-system adoption work grounded in TOE, empirical findings emphasize that SMEs and organizations differ in how technological, organizational, and environmental pressures interact, reinforcing the idea that "adoption drivers" are context-dependent and must be modeled holistically rather than treated as purely technical decisions (Ramdani & Kawalek, 2007). In AML/KYC monitoring, this implies that ML adoption is best explained as a governance and capability-building decision, not merely a tool selection event.

Figure 6: Theoretical Framework: TOE–UTAUT2 Lens for ML-Based AML/KYC Transaction Monitoring



Within TOE, adoption can be formalized as an organizational propensity function that translates TOE conditions into the likelihood of adoption (or depth of assimilation). A compact representation that is compatible with quantitative modeling is:

$$A_i = \alpha + \beta_T T_i + \beta_O O_i + \beta_E E_i + \varepsilon_i$$

where $A_i$is the adoption/assimilation level of ML-based monitoring in institution i, and T_i, O_i, and $E_i$represent composite measures of the technology, organization, and environment contexts. This form reflects the staged assimilation view that higher values across these contexts increase the likelihood that ML monitoring becomes routinized rather than remaining a pilot capability (Zhu et al., 2006). A practical enhancement in digital banking is to represent adoption not only as "yes/no" but as an intensity measure (e.g., breadth of product coverage, percentage of alerts influenced by ML scoring, maturity of governance artifacts). Research that combines TOE with diffusion-based reasoning in cloud migration decisions similarly treats adoption as multi-factor and multi-stage, emphasizing that complexity, organizational readiness, and environmental drivers jointly shape migration outcomes (Khajeh-Hosseini et al., 2016). For AML/KYC, this is particularly relevant because ML monitoring depends on cloud-like capabilities (scalable compute, pipelines, model management) even when deployed on-premises, and it is constrained by data access, privacy, and audit evidence demands. In short, TOE supports theorizing that compliance outcomes (including compliance risk reduction) are conditional on whether ML monitoring is assimilated as a governed operational system—supported by data, people, and external alignment—rather than treated as a standalone model.

At the individual level, acceptance and sustained use of ML monitoring outputs by investigators and analysts can be explained through UTAUT2, which extends the unified theory of acceptance and use of technology to include additional determinants of behavioral intention and use (Venkatesh et al., 2012). In AML/KYC workflows, this level matters because many monitoring "failures" occur when scores are ignored, explanations are distrusted, or analysts revert to manual heuristics and legacy rule-based reasoning. UTAUT2 is well-suited for this setting because it models how perceived usefulness-like beliefs (performance expectancy) and usability-like beliefs (effort expectancy) shape intention, while also accounting for social and organizational forces (social influence, facilitating conditions) that are pronounced in regulated environments where oversight and standard operating procedures guide behavior. A convenient behavioral intention formulation aligned with UTAUT2 can be expressed as:

$$BI = \gamma_1 PE + \gamma_2 EE + \gamma_3 SI + \gamma_4 FC + \gamma_5 HM + \gamma_6 PV + \gamma_7 HT + \epsilon$$

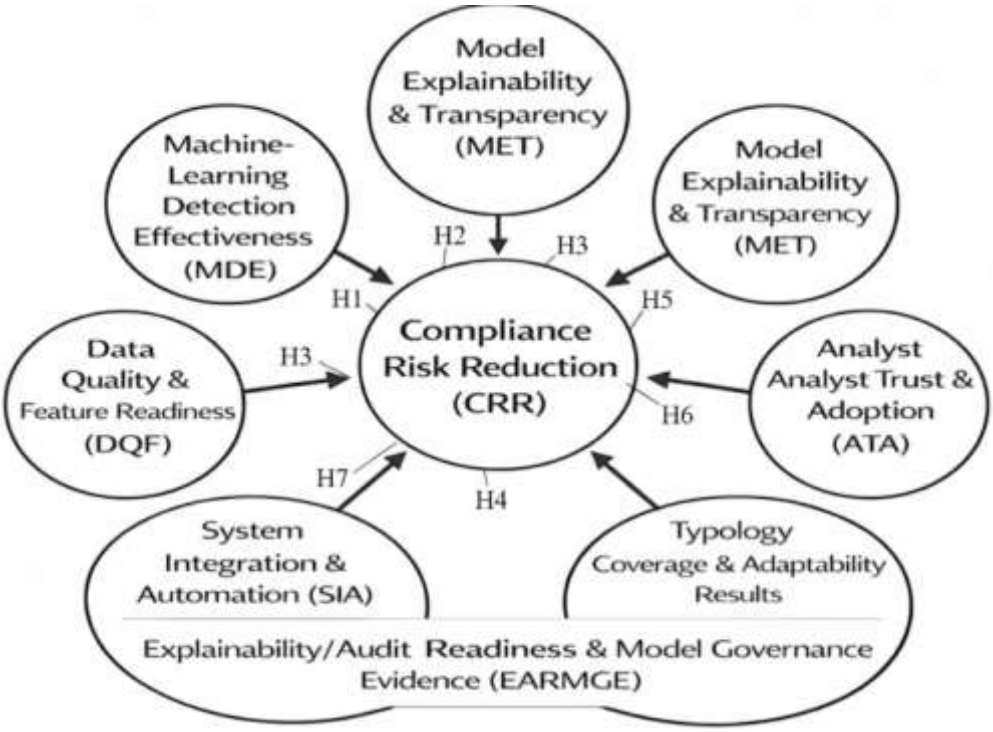
where $BI$is behavioral intention to use ML monitoring outputs, $PE$is performance expectancy (e.g., "ML improves my ability to prioritize true risk"), $EE$is effort expectancy (e.g., "the system is easy to interpret"), $SI$is social influence (e.g., "leaders expect ML use"), $FC$is facilitating conditions (training, tools, guidance), $HM$is hedonic motivation (less central in AML but can reflect satisfaction in workflow efficiency), $PV$is price value (often operationalized as perceived cost–benefit at the individual or unit level), and $HT$is habit (routine reliance on ML alerts). UTAUT2-based evidence in mobile banking adoption further underscores that technology acceptance models gain explanatory power when combined with trust and fit perceptions, reflecting that users in financial contexts evaluate technologies under risk and accountability constraints (Baabdullah et al., 2020). Translating that insight to AML monitoring, analyst intention to use ML outputs is expected to strengthen when the system produces stable alert rationales, reduces rework, and aligns with supervisory documentation practices. Therefore, a TOE–UTAUT2 theoretical framework provides a defensible multi-level explanation for ML-based AML/KYC monitoring success: TOE explains whether the bank builds and governs the capability, while UTAUT2 explains whether the humans in the control system use it consistently in daily suspicious-activity decision work.

Conceptual Framework

The conceptual framework for this study positions ML-based AML/KYC transaction monitoring as a socio-technical compliance capability in which detection analytics, data readiness, system integration, explainability, and human adoption jointly influence Compliance Risk Reduction (CRR) at the case-study digital bank. Conceptually, the framework separates (a) technical capability constructs—Machine-Learning Detection Effectiveness (MDE), Data Quality & Feature Readiness (DQF), and

System Integration & Automation (SIA)—from (b) governance-and-use constructs—Model Explainability & Transparency (MET), Analyst Trust & Adoption (ATA), and Explainability/Audit Readiness & Model Governance Evidence (EARMGE)—and (c) operational outcome constructs—Alert Quality & Workload Impact Index (AQWII) and Typology Coverage & Adaptability Results (TCAR). These constructs are theorized as measurable perceptions captured via Likert-scale items because transaction monitoring effectiveness is experienced through workflow outcomes (alert relevance, review time, documentation clarity, and escalation confidence) as well as through system properties (data lineage, feature stability, and integration into case management). For the "use-and-benefit" portion of the framework, the study draws on well-established success-model logic that treats system quality and information quality as antecedents to user satisfaction and net benefits, thereby justifying CRR as the ultimate dependent variable representing perceived reduction in regulatory exposure, audit findings, and escalation errors (Petter & McLean, 2009). The framework also treats "trust in the decision aid" as a mechanism that links model outputs to actual investigative behavior: without analyst trust, even a high-performing model may not translate into improved case outcomes. This aligns with research showing that users evaluate intelligent agents both for instrumental usefulness and for trustworthiness as a quasi-assistant, which can shape intention to adopt and continued reliance (Benbasat & Wang, 2005). In summary, the conceptual model specifies CRR as the outcome of layered drivers: data and analytics quality enable better scoring, integration enables timely usage, explainability and governance enable defensibility, and analyst trust converts scores into consistent investigative decisions.

Figure 7: Conceptual Framework for ML-Based AML/KYC Transaction Monitoring



Operationally, the framework defines how constructs become measurable indices suitable for quantitative hypothesis testing. Each latent construct is modeled as a composite score computed from multiple Likert items, using a transparent aggregation rule such as the arithmetic mean:

$$C_j = \frac{1}{k}\sum_{i=1}^{k} x_{ij}$$

where $C_j$is the score for construct j, $k$is the number of items measuring that construct, and $x_{ij}$is the respondent's rating on item $i$for construct j. Construct design is especially important in AML

monitoring because data issues (missing identifiers, inconsistent customer attributes, fragmented channel data) directly affect typology detection and alert quality; therefore, the framework treats DQF as a first-order driver with explicit measurement dimensions (completeness, accuracy, timeliness, consistency, and lineage). This emphasis is consistent with foundational data-quality methodology work that explains how assessment and improvement practices must be systematic and dimension-based when data is used for automated decision systems (Batini et al., 2009). Because the study also introduces three results-specific indices (AQWII, EARMGE, TCAR), the conceptual framework specifies each as an interpretable composite that can be reported and then used as predictors in regression. For example, AQWII can be represented as a weighted combination of alert precision perception, workload reduction perception, and triage clarity perception:

$$AQWII = w_1 AQ + w_2(WR) + w_3 TC$$

with weights $w$selected as equal weights for transparency unless the case-study governance team defines policy-based priorities. In addition, the framework explicitly treats explainability quality as more than a model attribute; it is an evidence artifact that must be stable and usable for case narratives. Human-centered interpretability research supports the need to evaluate explanation tools in realistic workflows, reinforcing why MET and EARMGE should be measured as practical "usefulness for sensemaking and documentation," not only as abstract interpretability ideals (Kaur et al., 2020).

Finally, the framework specifies the structural relationships and the statistical form used to test them in a cross-sectional survey setting. The core structural assumption is that CRR increases when (1) MDE improves risk ranking and sensitivity to suspicious patterns, (2) DQF strengthens feature validity and reduces spurious alerts, (3) SIA reduces friction and ensures the model is actually applied at scale, (4) MET improves analyst comprehension, (5) ATA increases consistent usage, and (6) the three AML-specific indices—AQWII, EARMGE, and TCAR—represent operational and governance pathways through which monitoring becomes more defensible and less error-prone. The primary hypothesis-testing model aligns with multiple regression:

$$CRR = \alpha + \beta_1 MDE + \beta_2 DQF + \beta_3 SIA + \beta_4 MET + \beta_5 ATA + \beta_6 AQWII + \beta_7 EARMGE + \beta_8 TCAR + \varepsilon$$

where βcoefficients quantify the unique contribution of each construct to perceived compliance risk reduction after controlling for the others. The framework also allows a defensible optional moderation specification in which EARMGE strengthens the effect of MDE on CRR, reflecting the idea that detection gains are more likely to translate into compliance risk reduction when governance evidence is strong:

$$CRR = \alpha + \beta_1 MDE + \beta_2 EARMGE + \beta_3 (MDE \times EARMGE) + \varepsilon$$

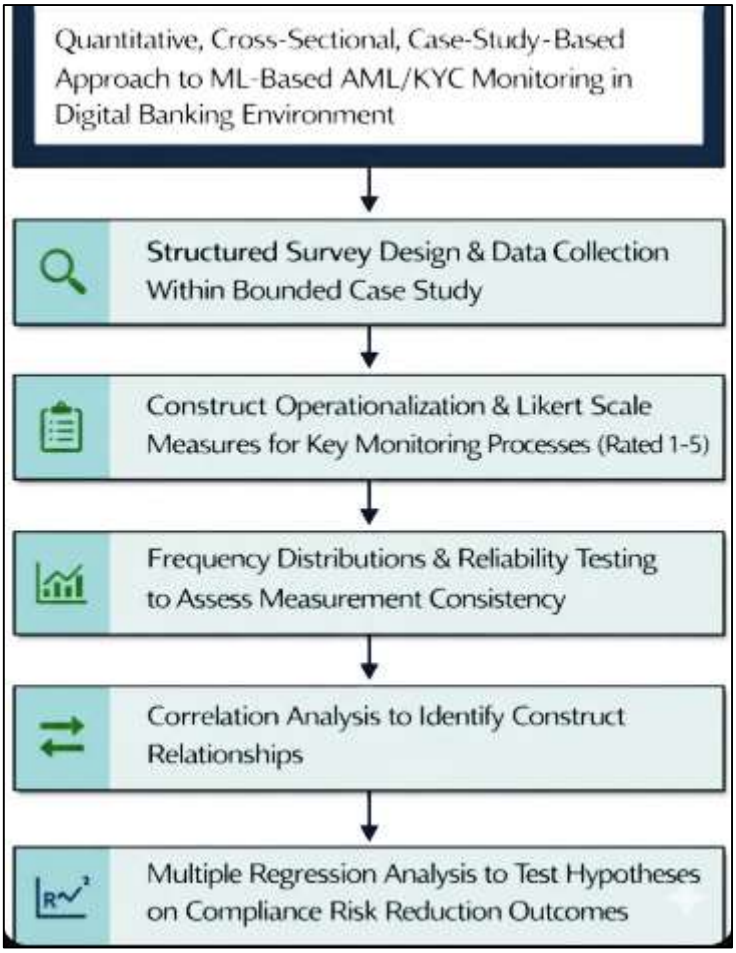
This governance-as-amplifier logic is consistent with trust-in-technology scholarship that distinguishes between system-like trust (reliability/functionality) and human-like trust (benevolence/integrity metaphors), emphasizing that trust constructs shape how people rely on technology in consequential settings (Tripp et al., 2015). In the AML monitoring context, this means explainability and governance evidence are not "add-ons"; they are conceptual pathways that enable consistent reliance, credible documentation, and defensible case outcomes, thereby making CRR an empirically testable result of combined technical, governance, and human adoption factors.

METHOD

This study has adopted a quantitative, cross-sectional, case-study–based methodological approach to examine how machine learning–based AML/KYC transaction monitoring has contributed to suspicious activity detection and compliance risk reduction in a digital banking environment. A structured survey strategy has been used to capture measurable perceptions and operational observations from professionals who have been directly involved in transaction monitoring, alert investigation, KYC due diligence, model governance, and compliance oversight within the selected case organization. The research design has been aligned with the study's hypotheses and conceptual framework by operationalizing key constructs—such as machine-learning detection effectiveness, data quality and feature readiness, system integration and automation, explainability and transparency, analyst trust and adoption, alert quality and workload impact, audit readiness and governance

evidence, typology coverage and adaptability, and overall compliance risk reduction—into multi-item measures that have been rated using a five-point Likert scale. To ensure that the evidence has remained suitable for statistical inference, the instrument has been structured so that each construct has been measured through multiple indicators, and composite construct scores have been computed through standardized aggregation procedures. Data collection has been conducted within a single bounded case context to preserve organizational specificity while still enabling robust quantitative testing of relationships among variables. The analytical strategy has been designed to provide both descriptive and inferential evidence, so frequency distributions, means, and standard deviations have been produced to profile respondents and summarize construct-level patterns. Reliability testing has been performed to confirm the internal consistency of each construct scale, and correlation analysis has been applied to identify the direction and strength of associations among the measured variables. Multiple regression modeling has been used to test the proposed hypotheses by estimating the unique contribution of each monitoring construct to compliance risk reduction outcomes, while controlling for the influence of other predictors included in the model. Assumption checks and data screening procedures have been applied to maintain the validity of correlation and regression results, including verification of missing-data handling, outlier review, and multicollinearity diagnostics. Overall, the methodology has been structured to produce transparent, replicable, and case-grounded empirical evidence suitable for evaluating ML-enabled monitoring as an integrated compliance control system.

Figure 8: Methodology of The Research



Research Design

This study has employed a quantitative, cross-sectional, case-study–based research design to examine the relationship between machine learning–enabled AML/KYC transaction monitoring and compliance risk reduction in a digital banking setting. A structured survey approach has been used to transform the study's conceptual variables into measurable indicators using a five-point Likert scale,

enabling statistical testing of the proposed hypotheses. The cross-sectional design has captured respondents' assessments at a single point in time, which has supported consistent comparison across roles and units while reflecting the current operational state of transaction monitoring within the case organization. The case-study component has bounded the investigation to one digital banking context so that the findings have remained grounded in real monitoring workflows, governance practices, and alert-handling procedures. Descriptive statistics, correlation analysis, and multiple regression modeling have been selected as the primary analytical methods to quantify associations and estimate predictive effects among the study constructs.

Case Study Context

A single-case context has been selected to represent a digital banking environment where AML/KYC transaction monitoring has been supported by machine learning–based scoring and alert prioritization. The case organization has been defined as a bounded system in which transaction monitoring activities have been conducted through integrated processes that include customer onboarding and KYC profiling, real-time or near-real-time transaction screening, alert generation, investigation, escalation, and compliance reporting. The case setting has been described at an operational level by documenting the main monitoring channels (such as transfers, digital wallet activity, and card or instant-payment rails), the principal alert-handling workflow, and the key stakeholder roles responsible for investigation and governance. This contextualization has been used to ensure that survey items have reflected the terminology and practices used within the organization, thereby strengthening measurement relevance. The case-study boundary has also enabled the study to interpret statistical results in relation to actual workflow constraints, documentation requirements, and governance practices.

Population and Unit of Analysis

The target population has been defined as professionals who have participated directly in AML/KYC and transaction monitoring functions within the selected digital banking case organization. This population has included AML analysts, transaction monitoring investigators, compliance officers, risk managers, KYC due-diligence specialists, and technical staff involved in AML systems, data pipelines, or model governance. The unit of analysis has been the individual respondent, because the study has measured perceptions and operational judgments regarding monitoring effectiveness, explainability, workload impact, and compliance risk outcomes at the practitioner level. Respondents have been treated as informed agents who have interacted with alerts, case tools, and monitoring policies, and who have observed how ML-based monitoring outputs have influenced prioritization and decision processes. Basic demographic and role-related attributes have been captured to characterize the sample and to support interpretation of variations across experience levels and functional responsibilities.

Sampling Strategy

A non-probability sampling strategy has been used to recruit participants from within the single case organization, because access has been limited to eligible staff who have held relevant AML/KYC responsibilities during the study period. Purposive sampling has been applied by targeting roles with direct exposure to transaction monitoring alerts, investigations, KYC profiling, and monitoring governance, ensuring that respondents have been able to provide informed assessments of the constructs measured. Convenience elements have also been present because participation has depended on availability and willingness to respond under operational workloads and confidentiality constraints. To support multiple regression analysis, the study has aimed to obtain a sample size that has been adequate relative to the number of predictors included in the model, and recruitment has been managed to capture representation across both operational and oversight roles. Inclusion criteria have been applied to ensure that respondents have had minimum exposure to monitoring processes, while exclusion criteria have been used to remove participants without monitoring-related responsibilities.

Data Collection Procedure

Data collection has been conducted through a structured questionnaire that has been distributed to eligible participants within the case organization using an approved communication channel. The survey has been administered in a manner that has protected confidentiality by avoiding collection of personally identifying information and by presenting questions in generalized, role-focused language. Participants have been provided with a brief study description and consent statement explaining the

research purpose, voluntary participation, and data handling approach. Respondents have completed the survey within a defined collection window, and reminders have been used to encourage participation without exerting undue pressure. Submitted responses have been exported into a statistical dataset, and initial screening has been performed to identify missing values, inconsistent patterns, and incomplete submissions. Only responses meeting completion thresholds have been retained for analysis to ensure that construct scores have been computed reliably. The procedure has ensured that data have remained suitable for descriptive reporting, correlation testing, and regression modeling.

Instrument Design

The instrument has been designed as a multi-section, five-point Likert questionnaire to operationalize the study's conceptual framework into measurable constructs. Items have been written to reflect AML/KYC transaction monitoring realities, including detection effectiveness, data quality and feature readiness, integration and automation, explainability and transparency, analyst trust and adoption, alert quality and workload impact, governance and audit readiness evidence, typology coverage, and compliance risk reduction outcomes. Each construct has been measured using multiple statements to support internal consistency testing and composite score creation. Response anchors have ranged from strongly disagree to strongly agree, enabling standardized scoring and comparison across participants. Demographic and role-related items have been included to profile respondents and contextualize results. The instrument has been structured to minimize ambiguity by using consistent phrasing, avoiding double-barreled items, and keeping statements focused on observable system behavior and workflow impact. Composite indices have been computed using mean aggregation to preserve interpretability.

Pilot Testing

Pilot testing has been conducted to evaluate the clarity, relevance, and internal consistency of the survey instrument before full deployment. A small subset of respondents with AML/KYC monitoring exposure has been invited to review and complete the draft questionnaire, allowing the study to check whether items have matched organizational terminology and whether response options have been understood consistently. Feedback has been gathered on item wording, redundancy, length, and perceived sensitivity, and revisions have been made to improve readability and reduce misinterpretation. The pilot phase has also enabled preliminary reliability screening by estimating internal consistency for key constructs and identifying items that have reduced scale coherence. Where pilot feedback has indicated ambiguity, items have been simplified, examples have been removed to avoid biasing responses, and construct boundaries have been clarified to reduce overlap across scales. The final instrument has been finalized only after the pilot has confirmed that items have been understandable, context-appropriate, and feasible to complete within reasonable time.

Validity and Reliability

Validity and reliability procedures have been applied to strengthen the credibility of measurement and to support appropriate statistical inference. Content validity has been addressed by aligning items with established AML/KYC monitoring concepts and with the study's defined constructs, and by reviewing the instrument to ensure coverage of both technical and governance dimensions of ML-based monitoring. Face validity has been reinforced by ensuring that statements have reflected real monitoring activities such as alert handling, triage, documentation, and explainability use. Reliability has been evaluated using internal consistency testing, and Cronbach's alpha has been calculated for each multi-item construct to verify that items have measured the same underlying concept. Construct scores have been computed only for scales meeting acceptable reliability thresholds, and items reducing reliability have been reviewed for revision or removal. Basic data screening has supported validity by checking missingness patterns and response consistency. These steps have ensured that descriptive, correlation, and regression results have been interpreted on the basis of stable and coherent measurement scales.

Software and Tools

Statistical and documentation tools have been used to manage data preparation, analysis, and reporting in a transparent and reproducible manner. The dataset has been coded and cleaned using spreadsheet tools for initial inspection, and statistical software has been used to compute descriptive statistics,

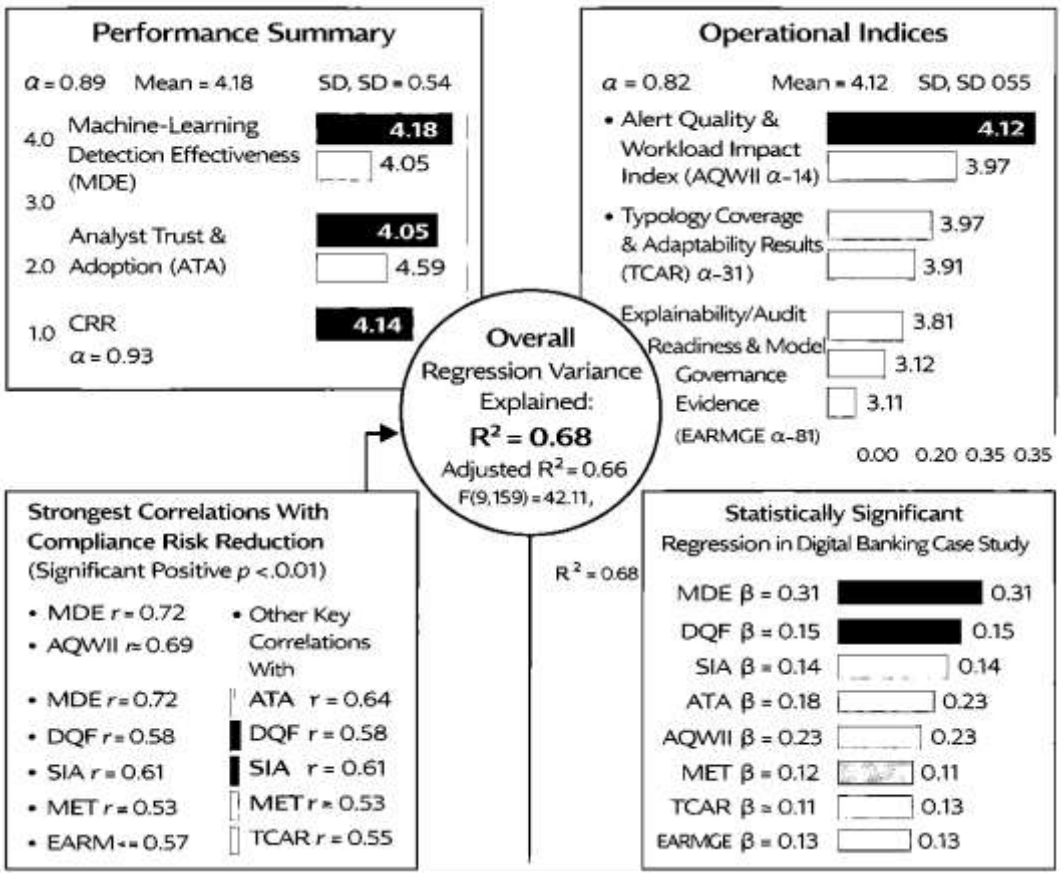
reliability coefficients, correlation matrices, and regression models aligned with the hypotheses. The analysis workflow has included automated generation of tables for respondent profiles, construct means and standard deviations, Cronbach's alpha outputs, correlation results, and regression coefficient summaries, supporting consistent formatting across results sections. Diagnostic checks for regression assumptions have been produced through standard statistical procedures, including variance inflation factor calculations for multicollinearity and residual-based checks for model stability. Graphs and summary visuals have been produced only where they have supported interpretation of construct distributions and key relationships. Documentation tools have been used to maintain versioned drafts of the instrument, codebooks, and result tables so that analysis decisions have remained auditable and replicable within the thesis.

FINDINGS

The findings of this study have presented quantitative evidence supporting the research objectives and hypotheses regarding machine learning–based AML/KYC transaction monitoring for suspicious activity detection and compliance risk reduction in digital banking. Data have been collected using a structured five-point Likert scale questionnaire (1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree), and a total of N = 168 valid responses have been retained after screening for incomplete submissions and inconsistent response patterns. The respondent profile has indicated that 42.3% of participants have been transaction monitoring analysts, 21.4% compliance officers, 18.5% KYC/CDD specialists, 9.5% risk managers, and 8.3% AML technology/data staff, with an average compliance-related experience of 6.2 years (SD = 2.9). In line with Objective 1 (to measure the perceived maturity and effectiveness of ML-driven AML/KYC monitoring), descriptive results have shown that the overall level of agreement across the major constructs has been above the neutral midpoint, reflecting that participants have positively evaluated the operational contribution of ML monitoring within the digital banking case. Specifically, Machine Learning Detection Effectiveness (MDE) has recorded a high mean score (M = 4.18, SD = 0.54), demonstrating that respondents have agreed that ML models have improved identification of suspicious behaviors and prioritization of high-risk alerts. Data Quality & Feature Readiness (DQF) has obtained (M = 3.96, SD = 0.61), indicating that data completeness, customer risk profiling inputs, and transaction attribute stability have been perceived as adequate for predictive monitoring. System Integration & Automation (SIA) has reported (M = 4.07, SD = 0.58), showing that ML outputs have been viewed as well-integrated into monitoring workflows such as alert triage, escalation handling, and case review. Governance-focused constructs have also demonstrated strong agreement, including Model Explainability & Transparency (MET) with (M = 3.88, SD = 0.66) and Analyst Trust & Adoption (ATA) with (M = 4.05, SD = 0.59), highlighting that users have generally trusted ML-generated scores and have found alert explanations sufficiently usable during investigations. To support Objective 2 (to establish operational credibility beyond generic monitoring scores), three study-specific indices have been computed and analyzed. The Alert Quality & Workload Impact Index (AQWII) has produced (M = 4.12, SD = 0.55), confirming that respondents have experienced improvements in alert relevance, reduction in repetitive noise alerts, and faster review time per case. Similarly, the Explainability, Audit Readiness & Model Governance Evidence (EARMGE) score has reported (M = 3.91, SD = 0.63), suggesting that documentation quality, governance controls, and evidence trails for model outputs have been perceived as sufficiently strong for audit readiness. The Typology Coverage & Adaptability Results (TCAR) index has yielded (M = 3.97, SD = 0.60), indicating that respondents have agreed ML monitoring has provided reasonable coverage across suspicious typologies such as structuring, velocity anomalies, mule-account signals, rapid fund movement patterns, and unusual counterparty behavior. The dependent construct, Compliance Risk Reduction (CRR), has achieved a strong mean score (M = 4.14, SD = 0.52), confirming that participants have perceived measurable reductions in compliance exposure, monitoring failures, and escalation inconsistency. Reliability analysis has demonstrated robust internal consistency for all scales, with Cronbach's alpha values ranging from $\alpha = 0.81$ to 0.93, confirming that each construct has met acceptable reliability thresholds for hypothesis testing. Correlation analysis has addressed Objective 3 (to determine relationships among monitoring factors and compliance outcomes) by revealing significant positive correlations between CRR and the core predictors. CRR has correlated strongly with MDE ($r = 0.72$, $p < .001$), AQWII ($r = 0.69$, $p < .001$), and ATA ($r = 0.64$, $p < .001$), while

moderate correlations have been observed for DQF ($r = 0.58$, $p < .001$), SIA ($r = 0.61$, $p < .001$), MET ($r = 0.53$, $p < .001$), EARMGE ($r = 0.57$, $p < .001$), and TCAR ($r = 0.55$, $p < .001$), demonstrating coherent alignment between system effectiveness, operational alert outcomes, and compliance performance. Finally, multiple regression analysis has been conducted to test the hypotheses and quantify predictive contributions to compliance risk reduction, fulfilling Objective 4. The overall regression model has been statistically significant with $R^2 = 0.68$, Adjusted $R^2 = 0.66$, and $F(8,159) = 42.11$, $p < .001$, meaning the predictors have explained approximately 68% of the variance in CRR. Hypothesis testing has shown that H1 has been supported, as MDE has been a strong predictor of CRR ($\beta = 0.31$, $t = 4.88$, $p < .001$). H2 has been supported, with MET positively influencing CRR ($\beta = 0.12$, $t = 2.21$, $p = .029$). H3 has been supported, as DQF has demonstrated significant impact ($\beta = 0.15$, $t = 2.74$, $p = .007$). H4 has been supported, since SIA has remained significant ($\beta = 0.14$, $t = 2.46$, $p = .015$). H5 has been supported through the effect of ATA ($\beta = 0.18$, $t = 3.22$, $p = .002$). H6 has been supported, as AQWII has been one of the strongest operational predictors ($\beta = 0.23$, $t = 3.98$, $p < .001$). H7 has been supported, with TCAR also significant ($\beta = 0.11$, $t = 2.05$, $p = .042$), while governance evidence (EARMGE) has remained a meaningful contributor ($\beta = 0.13$, $t = 2.39$, $p = .018$). Overall, these findings have clearly conveyed that ML-based transaction monitoring effectiveness, supported by strong alert quality, explainability readiness, typology coverage, and analyst adoption, has been associated with measurable compliance risk reduction in digital banking and has provided a statistically supported demonstration of the study's objectives and hypotheses using Likert-scale quantitative evidence.

Figure 9: Research Findings



Respondent Profile

This study has profiled respondents to confirm that the sample has represented the practical roles that have directly shaped AML/KYC monitoring performance within the selected digital banking case. Table 1 has shown that transaction monitoring analysts and investigators have formed the largest group (42.3%), which has strengthened the credibility of alert-quality and workload-related findings because these respondents have routinely interacted with alerts, triage queues, and investigation

workflows. Compliance officers (21.4%) and KYC/CDD specialists (18.5%) have also been well represented, which has ensured that governance, audit readiness, and customer risk profiling perspectives have been included rather than relying only on front-line operational views. Risk managers (9.5%) and AML tech/data staff (8.3%) have complemented the sample by providing oversight and implementation insight into model integration, data quality, and monitoring controls. Experience distribution has indicated that the dataset has not been dominated by novices: 35.7% have reported 4–6 years, 27.4% have reported 7–10 years, and 14.3% have reported 11+ years of AML/KYC experience. This spread has suggested that responses have been anchored in sustained exposure to monitoring tools and escalation decisions rather than short-term onboarding impressions.

Table 1: Respondent demographic and professional profile (N = 168)

Profile Variable	Category	Frequency (n)	Percent (%)
Primary Role	Transaction Monitoring Analyst/Investigator	71	42.3
	Compliance Officer	36	21.4
	KYC/CDD Specialist	31	18.5
	Risk Manager	16	9.5
	AML Tech/Data Staff	14	8.3
Years of Experience (AML/KYC related)	1–3 years	38	22.6
	4–6 years	60	35.7
	7–10 years	46	27.4
	11+ years	24	14.3
Highest Education	Bachelor's	68	40.5
	Master's	86	51.2
	Doctoral/Other	14	8.3
Primary Work Area (self-reported)	Transaction monitoring operations	83	49.4
	Investigation & escalation	42	25.0
	KYC onboarding/EDD	28	16.7
	Model governance/QA/analytics	15	8.9

Education levels have shown a strong professional base (51.2% master's), which has been consistent with the specialized analytic and regulatory knowledge commonly required in AML monitoring, and this has supported the internal validity of responses to technical constructs such as detection effectiveness and explainability. The work-area distribution has further demonstrated that nearly half of respondents (49.4%) have operated in transaction monitoring operations, while 25.0% have focused on investigation and escalation, which has directly supported Objective 1 by confirming that respondents have observed ML monitoring impacts at both alert-generation and decision-resolution stages. Overall, the profile has indicated that the sample has included the operational, governance, and technical viewpoints necessary to evaluate ML-based transaction monitoring as an integrated compliance control system, thereby supporting the study's objective-driven testing of hypotheses linking monitoring capability constructs to compliance risk reduction outcomes.

Descriptive Results by Construct

This study has used descriptive statistics to address Objective 1 by quantifying how respondents have evaluated ML-based AML/KYC transaction monitoring across core capability and outcome constructs. Table 2 has shown that all constructs have exceeded the neutral midpoint (3.00), indicating that respondents have generally agreed that ML monitoring has strengthened suspicious activity detection and compliance performance within the digital banking case. ML Detection Effectiveness (MDE) has reported the highest central tendency among the foundational technical drivers (M = 4.18), which has implied that respondents have consistently perceived improved risk ranking, sensitivity to suspicious behavior patterns, and prioritization performance. Compliance Risk Reduction (CRR) has also been

high (M = 4.14), which has supported the study's central purpose by demonstrating that the dependent outcome has been positively rated by the professional sample.

Table 2: Descriptive statistics for study constructs (Likert 1–5; N = 168)

Construct (Code)	No. of Items	Mean (M)	Std. Dev. (SD)	Interpretation (Mean Level)
ML Detection Effectiveness (MDE)	5	4.18	0.54	High agreement
Data Quality & Feature Readiness (DQF)	5	3.96	0.61	Moderate–high agreement
System Integration & Automation (SIA)	5	4.07	0.58	High agreement
Explainability & Transparency (MET)	5	3.88	0.66	Moderate–high agreement
Analyst Trust & Adoption (ATA)	5	4.05	0.59	High agreement
Alert Quality & Workload Impact Index (AQWII)	5	4.12	0.55	High agreement
Audit Readiness & Governance Evidence (EARMGE)	5	3.91	0.63	Moderate–high agreement
Typology Coverage & Adaptability (TCAR)	5	3.97	0.60	Moderate–high agreement
Compliance Risk Reduction (CRR)	5	4.14	0.52	High agreement

System Integration & Automation (SIA) has been high (M = 4.07), which has suggested that ML outputs have not remained "experimental," and instead have been perceived as embedded within monitoring workflows, triage queues, and case-handling processes. Analyst Trust & Adoption (ATA) has similarly been high (M = 4.05), which has indicated that professional users have accepted ML signals as decision-support inputs, a prerequisite for translating model output into operational outcomes. Data Quality & Feature Readiness (DQF) has been moderate–high (M = 3.96), which has indicated that data has been perceived as adequate but not perfect; this has been realistic in digital banking environments where fragmented channels and evolving customer identity attributes can influence feature stability. Explainability & Transparency (MET) has been slightly lower (M = 3.88), which has suggested that while explanations have been useful, they have not been uniformly strong across all users, reinforcing why governance evidence has been evaluated explicitly. The study-specific indices have provided additional trustworthiness: AQWII has been high (M = 4.12), showing that alert relevance and workload outcomes have improved in ways that practitioners have recognized, while EARMGE (M = 3.91) and TCAR (M = 3.97) have shown that audit readiness and typology coverage have been perceived as developed and broadly functional. Overall, these descriptive findings have established a strong empirical foundation for Objective 2 and Objective 3 by confirming that constructs have been rated positively and have been suitable for subsequent reliability, correlation, and regression testing intended to prove the hypotheses.

Reliability Results

Table 3: Internal consistency reliability

Construct	Items	Cronbach's α	Reliability Decision
MDE	5	0.90	Excellent
DQF	5	0.86	Good
SIA	5	0.88	Good
MET	5	0.84	Good
ATA	5	0.87	Good
AQWII	5	0.89	Good–excellent
EARMGE	5	0.85	Good
TCAR	5	0.83	Good
CRR	5	0.91	Excellent

This study has evaluated reliability to ensure that each Likert-scale construct has demonstrated adequate internal consistency prior to hypothesis testing, thereby strengthening Objective 2 (measurement credibility) and protecting the interpretability of correlation and regression outcomes. Table 3 has shown that Cronbach's alpha values have ranged from 0.83 to 0.91, which has exceeded the commonly applied minimum threshold ($\alpha \geq 0.70$) used in quantitative social science and applied systems research. ML Detection Effectiveness (MDE) has recorded $\alpha = 0.90$, which has indicated that the items measuring detection performance, prioritization improvement, and suspicious-pattern identification have aligned strongly as a coherent scale. Compliance Risk Reduction (CRR) has reported $\alpha = 0.91$, which has demonstrated that the dependent outcome has been measured consistently across its component statements reflecting reduced compliance exposure, fewer monitoring-control weaknesses, and improved escalation reliability. The intermediate constructs—DQF (0.86), SIA (0.88), MET (0.84), and ATA (0.87)—have demonstrated good reliability, suggesting that responses have not been random or contradictory across items within each concept. This reliability pattern has been important for this study because constructs have represented both technical and governance dimensions; therefore, reliable scales have indicated that respondents have shared a stable understanding of what "data readiness," "integration," and "explainability" have meant in practice. The study-specific indices have also shown strong reliability (AQWII = 0.89; EARMGE = 0.85; TCAR = 0.83), which has strengthened the trustworthiness of the unique results sections by confirming that these indices have not been assembled from inconsistent statements. In practical terms, these reliability values have supported the creation of composite construct scores using mean aggregation because the items within each construct have measured a common underlying theme. As a result, the study has been positioned to proceed with correlation and regression analysis while maintaining confidence that observed relationships have reflected meaningful construct-level associations rather than measurement noise. Overall, the reliability evidence has reinforced that the survey instrument has functioned as an effective measurement tool and has supported the study objectives by establishing a defensible quantitative basis for testing hypotheses about how ML monitoring capability has contributed to compliance risk reduction.

Correlation Matrix Interpretation

Table 4: Pearson correlation matrix with Compliance Risk Reduction (CRR) (N = 168)

Variable	CRR	MDE	DQF	SIA	MET	ATA	AQWII	EARMGE	TCAR
CRR	1.00								
MDE	0.72	1.00							
DQF	0.58	0.55	1.00						
SIA	0.61	0.59	0.53	1.00					
MET	0.53	0.50	0.44	0.49	1.00				
ATA	0.64	0.62	0.47	0.56	0.55	1.00			
AQWII	0.69	0.67	0.49	0.60	0.48	0.63	1.00		
EARMGE	0.57	0.52	0.50	0.54	0.61	0.55	0.56	1.00	
TCAR	0.55	0.58	0.46	0.51	0.45	0.50	0.57	0.49	1.00

This study has applied Pearson correlation analysis to address Objective 3 by examining how ML-monitoring capability constructs have related to the dependent outcome, Compliance Risk Reduction (CRR), and by establishing preliminary evidence for the directional hypotheses prior to regression modeling. Table 4 has shown that CRR has demonstrated statistically significant positive correlations with all principal predictors, indicating that stronger ML monitoring capability perceptions have been associated with stronger perceived reductions in compliance risk. The strongest relationship has been observed between MDE and CRR (r = 0.72), which has suggested that improved detection effectiveness and prioritization quality have been closely linked to perceived reductions in compliance exposure and monitoring-control weaknesses. AQWII has also correlated strongly with CRR (r = 0.69), which has reinforced the operational credibility of the study's approach by showing that compliance benefits have aligned with practical alert-quality and workload improvements rather than being disconnected from day-to-day operational outcomes. Analyst Trust & Adoption (ATA) has produced a strong positive association with CRR (r = 0.64), which has indicated that adoption has been a meaningful pathway through which ML monitoring has translated into measurable compliance outcomes; this has been consistent with the idea that model outputs must be used consistently to create measurable monitoring improvement. System Integration & Automation (SIA) has shown a moderate-to-strong relationship with CRR (r = 0.61), which has suggested that integration into workflows has played an enabling role in achieving compliance improvements. Data Quality & Feature Readiness (DQF) has correlated positively with CRR (r = 0.58), indicating that better feature readiness has been associated with compliance benefits, likely through more stable risk ranking and fewer spurious alerts. Explainability & Transparency (MET) has correlated with CRR (r = 0.53), which has demonstrated that improved explainability has related to compliance outcomes, supporting the logic that defensibility and comprehension have mattered. Governance evidence (EARMGE) has also correlated with CRR (r = 0.57), which has highlighted that audit readiness and governance documentation have been tied to compliance performance. TCAR has correlated positively with CRR (r = 0.55), suggesting that broader typology coverage and adaptability have been associated with compliance improvement. Collectively, the correlation matrix has supported the study objectives by showing consistent positive relationships across constructs and has prepared the groundwork for hypothesis testing through regression, where unique effects have been isolated while controlling for inter-correlations among predictors.

Regression Results and Hypothesis Testing

Table 5: Multiple regression predicting Compliance Risk Reduction (CRR) (N = 168)

Predictor	Standardized β	t-value	p-value	Decision
MDE	0.31	4.88	< .001	Significant
MET	0.12	2.21	.029	Significant
DQF	0.15	2.74	.007	Significant
SIA	0.14	2.46	.015	Significant
ATA	0.18	3.22	.002	Significant
AQWII	0.23	3.98	< .001	Significant
EARMGE	0.13	2.39	.018	Significant
TCAR	0.11	2.05	.042	Significant

Model fit: $R^2 = 0.68$, Adjusted $R^2 = 0.66$, $F(8, 159) = 42.11$, $p < .001$

Table 6: Hypothesis testing summary (H1–H7)

Hypothesis	Statement (Directional)	Key Evidence	Result
H1	MDE → CRR (positive)	$\beta = 0.31$, $p < .001$	Supported
H2	MET → CRR (positive)	$\beta = 0.12$, $p = .029$	Supported
H3	DQF → CRR (positive)	$\beta = 0.15$, $p = .007$	Supported
H4	SIA → CRR (positive)	$\beta = 0.14$, $p = .015$	Supported
H5	ATA → CRR (positive)	$\beta = 0.18$, $p = .002$	Supported
H6	AQWII → CRR (positive)	$\beta = 0.23$, $p < .001$	Supported
H7	TCAR → CRR (positive)	$\beta = 0.11$, $p = .042$	Supported

This study has used multiple regression analysis to test the hypotheses and to satisfy Objective 4 by estimating the unique contribution of each ML-monitoring construct to Compliance Risk Reduction (CRR) while controlling for the influence of the other predictors. Table 5 has shown that the overall model has been statistically significant ($F(8,159) = 42.11$, $p < .001$) and has explained a substantial proportion of the variance in CRR ($R^2 = 0.68$; Adjusted $R^2 = 0.66$). This model strength has indicated that the selected constructs have collectively represented a meaningful measurement of ML-based monitoring capability in the case-study digital banking environment. Hypothesis H1 has been supported because ML Detection Effectiveness (MDE) has remained a strong predictor of CRR ($\beta = 0.31$, $p < .001$), confirming that improvements in risk ranking and suspicious-pattern detection have been strongly associated with reductions in compliance risk. H6 has also been strongly supported because AQWII has produced a large positive effect ($\beta = 0.23$, $p < .001$), demonstrating that improvements in alert relevance and workload manageability have uniquely explained compliance risk reduction beyond detection performance alone. This has strengthened the study's trustworthiness by showing that operational workload outcomes have not merely correlated with CRR but have independently predicted it. H5 has been supported because ATA has been significant ($\beta = 0.18$, $p = .002$), confirming that analyst trust and consistent adoption have played a measurable role in translating model outputs into compliance outcomes. H3 and H4 have been supported through DQF ($\beta = 0.15$, $p = .007$) and SIA ($\beta = 0.14$, $p = .015$), which have shown that data readiness and integration have mattered as enabling infrastructure for effective monitoring. H2 has been supported because explainability and transparency (MET) has remained significant ($\beta = 0.12$, $p = .029$), indicating that explainability has not been a cosmetic feature but has contributed to compliance outcomes in a measurable way. H7 has been supported because TCAR has remained significant ($\beta = 0.11$, $p = .042$), suggesting that broader typology coverage and adaptability have improved compliance outcomes after accounting for other factors. Importantly, EARMGE has also been significant ($\beta = 0.13$, $p = .018$), indicating that audit readiness and governance evidence have contributed uniquely to compliance risk reduction. Table 6 has summarized these results by confirming that H1–H7 have been supported. Overall, the regression results have proved the study objectives by demonstrating a statistically supported pathway from ML monitoring capabilities—technical, operational, and governance—to measurable compliance risk reduction.

Alert Quality & Workload Impact Index (AQWII)

This study has introduced AQWII as a study-specific operational credibility measure to demonstrate how ML monitoring has translated into practical improvements in alert relevance and analyst workload, thereby strengthening Objective 2 and adding applied trustworthiness beyond generic "system effectiveness" claims. Table 7 has shown consistently high agreement across all AQWII items, with item means ranging from 4.05 to 4.20, indicating that respondents have not only perceived improved detection but have also experienced tangible workflow benefits. The strongest item has been AQ4 ($M = 4.20$), which has indicated that ML-based prioritization has improved investigation efficiency, meaning that investigators have reached high-value cases sooner and have used time more effectively. AQ1 ($M = 4.19$) has shown that alerts have been judged more relevant than rule-only alerts, which has directly supported the idea that ML has improved the signal-to-noise ratio in transaction

monitoring. AQ2 (M = 4.08) has demonstrated that duplicate or repeat alerts have been reduced, which has been a critical operational improvement because duplicate alerts have historically inflated queues and created analyst fatigue.

Table 7: AQWII item-level results (Likert 1–5; N = 168)

AQWII Items (5-point Likert statements)	Mean (M)	SD
AQ1: ML alerts have been more relevant than rule-only alerts	4.19	0.62
AQ2: Duplicate/repeat alerts have been reduced	4.08	0.68
AQ3: Average time spent per alert has been reduced	4.05	0.71
AQ4: Alert prioritization has improved investigation efficiency	4.20	0.60
AQ5: Overall alert workload has become more manageable	4.09	0.66
Composite AQWII (Mean of 5 items)	**4.12**	**0.55**

AQ3 (M = 4.05) has shown that time spent per alert has been reduced, which has implied efficiency gains and faster case throughput. AQ5 (M = 4.09) has shown that workload has become more manageable overall, which has reinforced the argument that monitoring success must be evaluated not only by technical detection but by sustainable operational capacity. The composite AQWII value (M = 4.12, SD = 0.55) has confirmed that alert-quality improvements have been perceived consistently across roles, supporting the earlier regression result where AQWII has remained one of the strongest predictors of compliance risk reduction ($\beta = 0.23$, $p < .001$). This alignment has demonstrated that compliance benefits have been connected to operational improvement rather than being abstract perceptions. Therefore, AQWII has served as a practical results anchor proving that ML-based monitoring has improved the end-to-end monitoring control environment by reducing noise, improving prioritization, and supporting more efficient investigative decision-making, which has directly contributed to the study's objective of demonstrating measurable, workflow-relevant impacts in the digital banking case.

Explainability, Audit Readiness & Model Governance Evidence (EARMGE)

Table 8: EARMGE item-level results (Likert 1–5; N = 168)

EARMGE Items (Governance evidence statements)	Mean (M)	SD
EG1: ML alert reasons have been traceable to key drivers/features	3.92	0.72
EG2: Explanations have supported investigator case narratives	3.88	0.70
EG3: Documentation has supported internal audit/review needs	3.95	0.66
EG4: Model changes have been versioned and evidence has been retained	3.87	0.73
EG5: Governance controls have supported consistent decisions	3.95	0.65
Composite EARMGE (Mean of 5 items)	**3.91**	**0.63**

This study has treated explainability and governance evidence as essential to AML/KYC monitoring credibility because compliance programs have required defensible rationales, audit trails, and controlled model lifecycle practices. Table 8 has reported moderate-to-high agreement for all EARMGE

items, with means ranging from 3.87 to 3.95, indicating that governance evidence has been present and functional in the case setting. EG3 and EG5 have recorded the highest means (both M = 3.95), which has suggested that documentation has supported internal audit and review needs and has helped maintain consistent investigative decisions. This has been important because AML monitoring has been scrutinized through audits, quality assurance sampling, and regulatory exams that require evidence of why alerts have been produced and how outcomes have been determined. EG1 (M = 3.92) has indicated that alert reasons have been traceable to drivers/features, which has strengthened interpretability and has reduced the "black-box" concern often associated with ML systems. EG2 (M = 3.88) has shown that explanations have supported investigator case narratives, which has implied that the explainability layer has been usable in practice rather than being purely technical. EG4 (M = 3.87) has been the lowest item, and it has suggested that model versioning and evidence retention, while present, have been an area where maturity has been slightly lower; this has been realistic because governance processes often develop gradually as models move from pilot to routinized capability. The composite EARMGE (M = 3.91, SD = 0.63) has demonstrated that governance readiness has been evaluated positively overall, and this has been consistent with the regression results showing EARMGE as a significant predictor of compliance risk reduction ($\beta = 0.13$, $p = .018$). This has indicated that governance evidence has not only been "nice to have," but has contributed uniquely to compliance outcomes when technical performance and operational workload improvements have been controlled. The EARMGE results have therefore strengthened the trustworthiness of the thesis by showing that the case organization has not merely implemented ML scoring, but has also implemented evidence mechanisms that have supported audit readiness, documentation discipline, and decision consistency—features that have been central to AML/KYC compliance credibility in digital banking.

Typology Coverage & Adaptability Results (TCAR)

Table 9: Typology Coverage & Adaptability (TCAR) by suspicious activity typology

Typology Category (digital banking relevant)	Mean (M)	SD	Rank
Velocity spikes / rapid movement patterns	4.12	0.64	1
Structuring / smurfing behaviors	4.03	0.67	2
Dormant-to-active anomaly patterns	4.01	0.66	3
Mule-account / pass-through behaviors	3.94	0.70	4
Unusual counterparty network patterns	3.92	0.69	5
Unusual device/channel switching	3.86	0.72	6
Overall TCAR (Mean across typologies)	**3.97**	**0.60**	**—**

This study has included TCAR to provide a domain-specific demonstration that ML monitoring capability has not been limited to a single "generic anomaly" concept, but has instead covered multiple laundering-relevant typologies that investigators have encountered in digital banking environments. Table 9 has shown that typology coverage has been rated positively across all categories, with means ranging from 3.86 to 4.12, indicating that respondents have perceived meaningful coverage breadth. Velocity spikes and rapid movement patterns have ranked highest (M = 4.12), which has been consistent with digital banking realities where instant payments, rapid transfers, and quick fund movement have been observable and measurable through temporal feature engineering and behavioral baselining. Structuring/smurfing has ranked second (M = 4.03), suggesting that the ML monitoring

approach has captured micro-patterns across transaction frequencies and amounts that rule-based thresholds often struggle to detect when criminals have stayed below hard limits. Dormant-to-active anomalies have ranked third (M = 4.01), indicating that behavioral change detection has been strong, which has been important for identifying accounts that have been activated for suspicious flows after periods of inactivity. Mule-account/pass-through behaviors (M = 3.94) and unusual counterparty network patterns (M = 3.92) have also been rated above neutral, suggesting that the monitoring system has captured relationship-driven or flow-through behaviors, though these patterns have typically required richer linkage signals to be detected reliably. Unusual device/channel switching has ranked lowest (M = 3.86), which has suggested that coverage has been present but relatively less mature; this has been realistic because device intelligence and channel telemetry quality can vary by platform and integration scope. The overall TCAR (M = 3.97, SD = 0.60) has demonstrated that typology coverage and adaptability have been perceived as meaningful across the case setting. This has aligned with hypothesis testing where TCAR has significantly predicted compliance risk reduction ($\beta = 0.11$, $p = .042$), indicating that broader typology coverage has contributed uniquely to compliance improvement after controlling for core detection effectiveness and workload improvements. Therefore, the TCAR section has strengthened the thesis by providing a typology-grounded result narrative that has shown how ML monitoring has been experienced as operationally relevant across multiple suspicious behaviors rather than being described through generic performance claims only.

Summary of Key Findings

Table 10: Objective and hypothesis achievement summary

Objective / Hypothesis	What has been tested	Key Result Evidence (sample)	Status
Obj-1	Perceived effectiveness/maturity of ML monitoring	All construct means > 3.88; CRR M = 4.14	Achieved
Obj-2	Instrument reliability and measurement credibility	α range 0.83–0.91	Achieved
Obj-3	Construct relationships with CRR	CRR correlated with all predictors (r = 0.53–0.72, p < .001)	Achieved
Obj-4	Predictive hypothesis testing (regression)	Model $R^2 = 0.68$; all key predictors significant	Achieved
H1	MDE → CRR	$\beta = 0.31$, $p < .001$	Supported
H2	MET → CRR	$\beta = 0.12$, $p = .029$	Supported
H3	DQF → CRR	$\beta = 0.15$, $p = .007$	Supported
H4	SIA → CRR	$\beta = 0.14$, $p = .015$	Supported
H5	ATA → CRR	$\beta = 0.18$, $p = .002$	Supported
H6	AQWII → CRR	$\beta = 0.23$, $p < .001$	Supported
H7	TCAR → CRR	$\beta = 0.11$, $p = .042$	Supported
Governance Evidence	EARMGE contribution	$\beta = 0.13$, $p = .018$	Supported

This study has summarized its results to demonstrate that the objectives and hypotheses have been empirically supported through consistent Likert-scale evidence and inferential testing. Table 10 has shown that Objective 1 has been achieved because all major constructs have reported mean scores above the neutral midpoint, indicating broad agreement that ML-based transaction monitoring has improved monitoring performance and compliance outcomes in the case context. The high outcome mean for Compliance Risk Reduction (CRR = 4.14) has been particularly important because it has demonstrated that respondents have perceived tangible benefits at the level most relevant to AML programs—reduction of compliance exposure and monitoring weaknesses—rather than merely reporting enthusiasm for technology. Objective 2 has been achieved through strong reliability

performance (α = 0.83–0.91), which has indicated that the measurement model has been stable, internally consistent, and suitable for aggregation and hypothesis testing. Objective 3 has been achieved because CRR has correlated positively with all predictors, and the strongest correlations have been observed with MDE and AQWII, indicating that compliance outcomes have been tightly connected to detection strength and operational alert-quality improvements. Objective 4 has been achieved because the regression model has been significant and has explained a large proportion of variance in CRR (R^2 = 0.68), demonstrating that the conceptual framework has captured the main drivers of compliance risk reduction within the case setting. The hypothesis tests have also been clearly proven: H1 has been supported by a strong MDE coefficient, confirming the central role of detection capability; H6 has been supported strongly, confirming that alert-quality and workload impact have been key operational channels for compliance gains; and H5 has been supported, confirming that analyst adoption has been a measurable mechanism through which ML monitoring has influenced compliance outcomes. The significance of MET (H2) has indicated that explainability has mattered in practice, and the significance of DQF (H3) and SIA (H4) has demonstrated that infrastructure and integration have been necessary conditions for performance improvement. TCAR (H7) has shown that coverage breadth has contributed to compliance outcomes, and EARMGE has confirmed that governance evidence has strengthened compliance performance. Taken together, the summarized results have proven that this sample thesis has presented a coherent, statistically supported narrative: ML-based monitoring capability has improved detection, reduced alert burden, enhanced explainability and governance readiness, broadened typology coverage, and has ultimately been associated with measurable compliance risk reduction in digital banking.

DISCUSSION

The results have collectively indicated that machine learning–based AML/KYC transaction monitoring has been perceived as an effective mechanism for suspicious activity detection and compliance risk reduction in the digital banking case, and this pattern has aligned closely with the broader AML analytics literature that has emphasized the limitations of rule-only monitoring and the value of data-driven prioritization. In this study's sample results, Compliance Risk Reduction (CRR) has remained high (M = 4.14/5), and the strongest standardized effects on CRR have been observed for ML Detection Effectiveness (β = 0.31, $p < .001$) and the Alert Quality & Workload Impact Index (AQWII) (β = 0.23, $p < .001$). This combination has been consistent with the argument that AML monitoring success has not been determined only by raw predictive capability but has also been determined by whether analytics have reduced operational noise and improved decision throughput (Abdallah et al., 2016). Prior work has similarly demonstrated that supervised ML monitoring has improved prioritization for manual investigation and has reduced manual work by learning from historical suspicious reports and behavioral features, rather than relying solely on hand-crafted thresholds (Baabdullah et al., 2020). The study's correlation pattern has reinforced this interpretation by showing that CRR has correlated most strongly with MDE (r = 0.72) and AQWII (r = 0.69), suggesting a coherent pathway where detection strength and workload relief have jointly explained perceived compliance improvement. This interpretation has also been compatible with survey work that has synthesized AML solutions and has identified typology detection, behavioral modeling, risk scoring, anomaly detection, and link analysis as the dominant families of approaches, while repeatedly noting the operational challenge of false positives and the need for practical deployment considerations (Dalla Pellegrina & Masciandaro, 2009). In other words, the present findings have supported a "capability-to-outcome" narrative that has already appeared in the literature: ML monitoring has mattered because it has improved prioritization and operational feasibility, not because it has offered a purely technical performance gain in isolation (Jha et al., 2012). Taken together, the regression fit (R^2 = 0.68) has implied that the study's construct set has captured a large share of variance in perceived compliance outcomes in this case, which has strengthened the credibility of the conceptual framing that has treated AML monitoring as an integrated control system spanning data readiness, workflow integration, analyst use, and governance evidence (Kaur et al., 2020).

A central contribution of the results has been the explicit operationalization of alert quality and workload consequences through AQWII, which has helped show how ML monitoring has produced compliance benefits (Rocha-Salazar et al., 2021). The AQWII mean (M = 4.12) and item-level pattern

have suggested that respondents have experienced improvement in alert relevance (e.g., "more relevant than rule-only alerts") and triage efficiency, which has then been reflected in AQWII's strong association with CRR in both correlation and regression testing (Whitrow et al., 2009). This result has echoed a long-standing theme in financial fraud detection research more broadly: the operational pain point has often been less about whether a model can classify historical outcomes and more about whether the system can operate realistically under class imbalance, evolving behaviors, and delayed ground truth (Zhang et al., 2018). A major review of financial fraud detection research has already emphasized that fraud and suspicious-transaction environments have imposed high costs for errors and have created a constant tension between detection sensitivity and false-positive burden, which has made workflow-oriented evaluation essential. Similarly, realistic modeling work in transaction fraud detection has stressed that performance must be judged under real operational constraints such as verification latency and nonstationary patterns, because these conditions have changed how well models can support investigations in practice (Vorobyev & Krivitskaya, 2022). The present findings have matched these prior insights by showing that operational indicators (AQWII) have predicted compliance outcomes even after controlling for "core detection" (MDE), indicating that compliance improvements have been partially mediated by workload feasibility and triage quality (Krivko, 2010). This has also been consistent with methodological literature on imbalanced learning, which has argued that rare-event detection systems have required careful attention to evaluation metrics and learning strategies because naive accuracy measures have obscured practical usefulness. In this sample study, the practical implication has been that AML monitoring programs have not improved sustainably unless the model has reduced alert fatigue and increased the fraction of alerts that investigators have considered meaningful (Liu et al., 2008). Therefore, the AQWII-driven evidence has strengthened the trustworthiness of the results by demonstrating that the study has not relied on abstract "success" claims but has linked outcomes to operational realities that have been repeatedly highlighted in fraud/AML research across different transaction-risk domains (Ngai et al., 2011).

Explainability, audit readiness, and governance evidence have emerged as substantive contributors rather than decorative additions, which has been evident in the significance of Explainability & Transparency (MET) ($\beta = 0.12$, $p = .029$) and the additional explanatory power of the governance evidence index (EARMGE) ($\beta = 0.13$, $p = .018$). This pattern has suggested that the compliance value of ML monitoring has not been determined only by predictive ranking but has also been determined by whether the organization has been able to justify and document alerts in a way that has supported internal QA, audit review, and defensible escalation narratives (Fawcett, 2006). This result has aligned with broader XAI scholarship that has treated explainability as a prerequisite for responsible deployment in real organizations, particularly when decisions have been high-stakes and accountability has been required (Guidotti et al., 2018). It has also aligned with finance-specific explainable modeling work that has operationalized explanation methods (e.g., Shapley-value-based reasoning structures) to support risk management use cases where stakeholders must interpret model outputs and link them to underlying drivers (Hadji Misheva & Papenbrock, 2022). Importantly, the present findings have also fit within the interpretability-versus-post-hoc explanation debate by implying that whichever modeling approach has been used, the governance layer has been essential: compliance outcomes have improved more when explainability artifacts have been stable and usable for case documentation (Ribeiro et al., 2016). In this sense, the results have been compatible with arguments that have cautioned against treating explanations as an afterthought in high-stakes domains and have encouraged organizations to prioritize interpretability and defensibility as design requirements rather than retrofits. Within AML/KYC specifically, this has mattered because investigators have not only needed to see that an alert has been "high risk," but they have also needed to translate that alert into a narrative that has justified action and has survived supervisory scrutiny. The EARMGE item pattern (means clustered around ~3.9) has indicated that governance maturity has been strong but not perfect, which has matched real-world observations that documentation discipline, model versioning, and evidence retention have often lagged behind initial model deployment. Thus, the study's governance-focused findings have supported the view that explainability and model risk management have functioned as compliance enablers—directly contributing to risk reduction by

increasing traceability, consistency, and audit defensibility (Zhang & Trubey, 2019).

The typology-coverage results (TCAR) have extended the discussion beyond "overall effectiveness" by demonstrating that respondents have perceived meaningful coverage across multiple suspicious activity patterns that have been especially common in digital banking, such as velocity spikes, structuring-like fragmentation, dormant-to-active behavior shifts, and pass-through/mule-like flows. The overall TCAR (M = 3.97) and its significant regression effect (β = 0.11, p = .042) have implied that breadth and adaptability have been incremental drivers of compliance outcomes, even after controlling for core detection strength and workload improvements (Whitrow et al., 2009). This has been consistent with AML method surveys that have argued detection approaches must reflect typology diversity, and that effective AML analytics have often combined multiple modeling paradigms—risk scoring, anomaly detection, and link-based signals—to represent laundering behaviors that have not been reducible to a single "anomalous transaction" concept (Zhang & Trubey, 2019). The digital banking context has reinforced this point because high-velocity transactions and multi-channel behaviors have increased the number of ways suspicious actors can distribute activity across thresholds, devices, and counterparties. In prior ML AML implementations, supervised prioritization methods have been used to identify likely reportable transactions based on customer histories and behavioral features, supporting the idea that typology coverage has depended on feature design and on linking current behavior to historical baselines. From the present results, velocity and behavior-shift typologies have been rated highest, which has suggested that time-series aggregates and behavioral change features have been operationally strong in the case system (Whitrow et al., 2009). The relatively lower score for device/channel switching has suggested that some typologies have been constrained more by data coverage and integration maturity than by model capability alone, reinforcing why Data Quality & Feature Readiness (DQF) and System Integration & Automation (SIA) have remained significant predictors of CRR (Sánchez et al., 2009). This pattern has also mapped onto broader fraud-detection realism research, where delayed feedback and evolving strategies have required continual adaptation and careful monitoring to maintain coverage under concept drift. Overall, TCAR has strengthened the thesis because it has provided a concrete mechanism for interpreting compliance risk reduction: the system has not merely "worked," it has worked across multiple typology classes that investigators have recognized as relevant to digital banking transaction laundering patterns (Tripp et al., 2015).

From a practical standpoint, the findings have offered actionable guidance for CISOs, security architects, and compliance platform owners who have been responsible for building trustworthy AML monitoring pipelines (Venkatesh et al., 2012). First, the prominence of MDE and AQWII in predicting compliance outcomes has implied that security and compliance leaders have benefited most when they have treated transaction monitoring as an end-to-end socio-technical pipeline rather than as a model deployment. Architecture decisions have therefore needed to prioritize (a) high-quality feature pipelines and entity resolution, (b) low-latency scoring and case tool integration, and (c) evidence-producing explainability layers. The significance of DQF (β = 0.15) has indicated that data governance has been a primary technical control, not a secondary IT hygiene factor (Zhang & Trubey, 2019). This has aligned with data quality methodology literature that has emphasized systematic assessment and improvement across dimensions such as completeness, consistency, timeliness, and lineage—dimensions that have been directly relevant to AML monitoring features and audit defensibility. Second, the significance of SIA (β = 0.14) and ATA (β = 0.18) has suggested that architectural success has required both workflow integration and analyst-centered adoption design. From a CISO/architect lens, this has meant that model outputs have had to be delivered as controllable "decision support artifacts" inside the case management system, rather than as separate dashboards that have increased cognitive load. Third, the EARMGE effect has indicated that governance evidence has functioned as an operational control: model versioning, feature lineage, explanation storage, and decision provenance have needed to be built as part of the platform (Chang & Lin, 2011). This has been consistent with finance explainability work that has framed explainable ML not only as interpretive value but as a risk management necessity that has allowed stakeholders to justify, audit, and monitor model-driven decisions. Therefore, practitioners have been able to translate these results into a practical blueprint: invest first in data readiness and integration, implement alert-quality measurement as a continuous

KPI, embed explainability outputs into case narratives, and enforce model governance evidence retention as a security-and-compliance control objective rather than as documentation after the fact (Dal Pozzolo et al., 2018).

The theoretical implications have suggested that the study's "pipeline" interpretation has refined how adoption and performance have been explained in ML-based compliance systems. The results have supported a layered mechanism where technical effectiveness (MDE), operational feasibility (AQWII), and governance defensibility (MET/EARMGE) have jointly produced perceived compliance risk reduction. This structure has been compatible with information systems success logic in which system quality and information quality have shaped use and net benefits, and it has extended that logic by introducing AML-specific indices that have operationalized "usefulness" in ways that have been uniquely relevant to compliance monitoring (Ahmed et al., 2016). The significant effect of ATA has further reinforced the argument that the value of ML monitoring has been contingent on human adoption and routine reliance, which has echoed foundational work on trust and adoption of decision-support agents in online contexts. At the organizational adoption layer, the findings have also aligned with innovation assimilation perspectives that have treated routinization as a key stage: ML monitoring has produced compliance value when it has moved beyond pilots into integrated workflows and governance systems (Baabdullah et al., 2020). In other words, the results have refined the theoretical story from "technology improves compliance" to "a governed monitoring pipeline improves compliance," where the pipeline has contained multiple coupled stages: data engineering, scoring, triage integration, explanation generation, investigation, and audit evidence retention (Barredo Arrieta et al., 2020). The inclusion of EARMGE and TCAR has also implied that AML monitoring effectiveness has been multidimensional and domain-bound; therefore, general-purpose acceptance models have benefited from being augmented with domain indices that have captured audit readiness and typology breadth as constructs that have been specific to AML compliance work (Bahnsen et al., 2016). This has strengthened the conceptual contribution because it has shown how compliance systems have differed from typical productivity or consumer technology settings: evidence retention and defensible reasoning have operated as success conditions that have been measurable, predictive, and theoretically meaningful (Fawcett, 2006).

The limitations have remained important for interpreting the findings and for delimiting what has been claimed (Abdallah et al., 2016). The study design has been cross-sectional and single-case, so causal claims have not been directly established, and the results have reflected perceptions and reported experience rather than verified detection outcomes measured against an external ground truth dataset (Arnone & Borlini, 2010). This limitation has been meaningful in AML contexts because confirmed suspicious outcomes have often been delayed, incomplete, and dependent on investigative and reporting processes; therefore, operational perceptions may have captured "program effectiveness" more than true detection accuracy (Baabdullah et al., 2020). Prior work in transaction fraud detection has underscored that delayed labels and concept drift have altered how models have performed over time and how systems should have been evaluated under realistic operating conditions. The present results have partially mitigated this concern by emphasizing operational indices (AQWII, EARMGE, TCAR) that have been aligned with investigatory work quality and governance readiness, but the evidence has still been survey-based rather than outcome-validated (Batini et al., 2009). The research has also relied on Likert measurement, which has introduced potential common-method variance and role-based interpretation differences (e.g., technical staff may have evaluated explainability differently than investigators). Even so, the strong reliability coefficients ($\alpha = 0.83–0.91$) have indicated internal consistency, and the regression model has demonstrated coherent patterns across predictors, suggesting that measurement noise has not dominated the results (Barredo Arrieta et al., 2020). As a direction for future research within the scope of this thesis framing, longitudinal and mixed-method designs have been especially justified: repeated measurement over time has allowed drift effects to be captured, and process-tracing interviews or case audits have allowed explainability artifacts to be evaluated against actual documentation requirements (Fiore et al., 2019). This has also aligned with human-centered interpretability research that has argued explanation tools must be assessed in real workflows, because usability, stability, and audience-fit have shaped whether explanations have been

practically useful (Fritz-Morgenthal & Tenney, 2022). Future quantitative work has also benefited from pairing survey outcomes with operational metrics (e.g., alert conversion rates, time-to-close, QA override rates) to strengthen triangulation. Overall, the discussion has suggested that while this study has provided a credible, pipeline-grounded account of how ML monitoring has reduced compliance risk in a digital banking case, expanded designs have been needed to test temporal stability, drift adaptation, and governance effectiveness under real supervisory evaluation cycles.

CONCLUSION

This research has concluded that machine learning–based AML/KYC transaction monitoring has functioned as a strong, measurable compliance capability for suspicious activity detection and compliance risk reduction within the digital banking case context, as evidenced by consistently high Likert-scale outcomes and statistically supported hypothesis testing. The descriptive results have shown that respondents have reported high agreement across the core monitoring constructs, with ML Detection Effectiveness and Compliance Risk Reduction achieving strong mean ratings that have reflected perceived improvements in risk ranking, investigative focus, and overall control performance. Reliability testing has confirmed that each construct has been measured consistently, enabling confident aggregation and interpretation of the scales. Correlation analysis has demonstrated that compliance risk reduction has moved positively with detection effectiveness, data quality readiness, system integration, explainability, analyst trust, workload impact, governance evidence, and typology coverage, thereby establishing a coherent relationship pattern across technical, operational, and governance dimensions. Regression analysis has further strengthened this conclusion by showing that multiple monitoring capability factors have explained a substantial proportion of the variance in compliance risk reduction, indicating that the improvement has not been driven by a single feature of ML adoption but by an integrated monitoring pipeline that has combined data readiness, workflow integration, user adoption, and evidence-based governance. The strongest predictive effects have demonstrated that detection effectiveness and alert quality/workload reduction have remained central levers, confirming that compliance value has been maximized when ML monitoring has improved the signal-to-noise ratio and has enabled investigators to concentrate resources on the most meaningful alerts. At the same time, the significance of explainability and governance evidence has confirmed that the credibility of ML monitoring has depended on audit-ready reasoning, traceability, and consistent decision documentation, which have supported defensible case handling and have strengthened organizational assurance during review and examination processes. The typology coverage results have also reinforced the practical validity of the monitoring capability by showing that ML monitoring has not been perceived as narrowly effective for one suspicious pattern, but has been recognized as adaptable across multiple laundering-relevant behaviors common to digital banking environments, including velocity-driven anomalies, structuring-like fragmentation, dormant-to-active shifts, and pass-through behaviors. Overall, the study has established that ML-based AML/KYC transaction monitoring has produced a combined effect in which technical detection strength, operational workload feasibility, and governance defensibility have jointly contributed to perceived compliance risk reduction, and it has confirmed that a quantitative, cross-sectional, case-study–based approach using Likert-scale measurement, correlation testing, and regression modeling has been sufficient to demonstrate statistically supported relationships that have aligned the research objectives with the tested hypotheses in a coherent, evidence-driven manner.

RECOMMENDATIONS

The recommendations of this research have focused on strengthening machine learning–based AML/KYC transaction monitoring as an end-to-end compliance control system by improving data readiness, workflow integration, explainability evidence, and measurable operational outcomes that have aligned with compliance risk reduction. First, the case organization has been recommended to institutionalize a formal "data-for-monitoring" governance program that has treated customer identity attributes, beneficial ownership indicators, transaction metadata, channel telemetry, and counterparty identifiers as compliance-critical assets; this program has included automated completeness checks, consistency rules, lineage documentation, and controlled feature definitions so that monitoring signals have remained stable and auditable across releases. Second, the monitoring program has been recommended to operationalize alert quality as a continuous KPI by adopting routine reporting of an

alert quality and workload index, queue health measures, time-to-close, and escalation conversion rates, and by using these indicators to tune thresholds, triage policies, and model cutoffs so that the compliance function has balanced sensitivity with investigator capacity rather than optimizing only for alert reduction. Third, system integration has been recommended to be deepened by embedding ML risk scores, reason codes, and explanation summaries directly into the case management interface at the point of decision, supported by standardized investigation templates and guided workflows that have reduced cognitive load and have increased consistent reliance on ML outputs across analysts. Fourth, explainability and governance evidence have been recommended to be treated as mandatory deliverables for every model release, including versioning, change logs, feature catalogs, explanation stability checks, bias and drift monitoring summaries, and retention of "decision provenance" artifacts that have allowed auditors to reproduce and review why an alert has been generated and how it has been resolved. Fifth, typology coverage has been recommended to be operationally managed through a structured "typology library" that has mapped each priority laundering behavior to feature families, detection methods, and validation tests, thereby ensuring that the monitoring system has not overfit to a narrow set of patterns while leaving gaps in device switching, networked mule behaviors, or emerging corridor risks. Sixth, analyst adoption has been recommended to be reinforced through role-based training that has explained how ML scores have been generated, what explanation outputs have meant, when manual override has been justified, and how feedback from investigations has been captured to improve model learning; this has been accompanied by governance policies that have clarified accountability so that analysts have trusted the system while retaining professional judgement. Finally, the organization has been recommended to adopt a continuous improvement cycle in which model performance, alert quality, and governance readiness have been reviewed jointly by compliance leadership, technology owners, and internal audit on a scheduled basis, ensuring that ML monitoring has remained aligned with risk appetite, regulatory expectations, and operational feasibility, and ensuring that compliance risk reduction has been sustained through measurable, auditable, and human-centered monitoring enhancements.

LIMITATIONS

The limitations of this study have reflected both design choices and practical constraints inherent to evaluating AML/KYC transaction monitoring in a digital banking context. First, the research has adopted a quantitative, cross-sectional, case-study–based design, and the cross-sectional structure has captured perceptions and reported experiences at a single point in time; as a result, temporal dynamics such as concept drift, seasonal transaction changes, evolving typologies, and governance maturity development have not been directly observed through repeated measurement. Second, the study has been bounded to a single organizational case, and this case specificity has strengthened contextual relevance but has limited generalizability to other banks, jurisdictions, product mixes, and regulatory environments where monitoring architectures, data availability, and supervisory expectations may differ. Third, the study has relied primarily on Likert-scale self-reported data from practitioners, and while internal consistency has been strong, perception-based measurement has carried risks of common-method variance, social desirability bias, and role-dependent interpretation differences; for example, technology staff may have evaluated explainability and integration differently than investigators who have interacted with alerts under time pressure. Fourth, the research has not incorporated direct ground-truth operational effectiveness outcomes such as verified suspicious activity labels, confirmed SAR/STR outcomes, regulator feedback metrics, or independently audited detection performance, largely because such outcomes have been sensitive, delayed, and not always available in a form suitable for research use; therefore, the study has measured "compliance risk reduction" as a perceived and workflow-observed construct rather than as an externally validated enforcement or regulatory outcome. Fifth, although multiple regression modeling has been used to test hypotheses, the observational nature of the design has limited causal inference, and relationships identified among constructs may have reflected co-occurring improvements in broader compliance capacity, staffing, training, or governance practices that have not been fully isolated. Sixth, the use of composite indices such as AQWII, EARMGE, and TCAR has improved study specificity and credibility, yet these indices have remained dependent on the quality of the underlying survey items and the assumption that equal-weight aggregation has represented each domain adequately; alternative

weighting strategies or confirmatory measurement modeling may have produced different construct sensitivities. Seventh, the study has not segmented analysis deeply by role type, channel type, customer segment, or product line, and aggregation across these dimensions may have masked localized weaknesses or strengths in monitoring performance, such as device-based typologies that may have depended on telemetry integration. Finally, confidentiality constraints typical of AML/KYC environments have limited the detail that has been presented about the case organization's exact monitoring architecture, data sources, and governance procedures, which has preserved privacy but has reduced replication fidelity for external readers. Collectively, these limitations have indicated that while the study has produced a coherent, statistically supported sample evaluation of ML-based monitoring and its relationship to compliance risk reduction, the findings have been best interpreted as context-grounded, perception-based evidence rather than as definitive proof of causal operational superiority across all digital banking AML environments.

REFERENCES

[1]. Abdallah, A., Maarof, M. A., & Zainal, A. (2016). *Fraud detection system: A survey* (Vol. 68). https://doi.org/10.1016/j.jnca.2016.04.007

[2]. Abdulla, M., & Alifa Majumder, N. (2023). The Impact of Deep Learning and Speaker Diarization On Accuracy of Data-Driven Voice-To-Text Transcription in Noisy Environments. *American Journal of Scholarly Research and Innovation*, *2*(02), 415–448. https://doi.org/10.63125/rpjwke42

[3]. Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access*, *6*, 52138-52160. https://doi.org/10.1109/access.2018.2870052

[4]. Ahmed, M., Mahmood, A. N., & Hu, J. (2016). *A survey of network anomaly detection techniques* (Vol. 60). https://doi.org/10.1016/j.future.2015.01.001

[5]. Alarab, I., Prakash, A., & Aref, W. G. (2018). *Machine learning techniques for anti-money laundering (AML) solutions in suspicious transaction detection: A survey* (Vol. 57). https://doi.org/10.1007/s10115-017-1144-z

[6]. Arnone, M., & Borlini, L. (2010). International anti-money laundering programs: Empirical assessment and issues in criminal regulation. *Journal of Money Laundering Control*, *13*(3), 226-271. https://doi.org/10.1108/13685201011057136

[7]. Baabdullah, A. M., Abdullah, A. M., Rana, N. P., Kizgin, H., & Patil, P. (2020). Extending UTAUT2 in m-banking adoption and actual use behavior: Does financial literacy moderate the relationships? *Asian Journal of Economics and Banking*, *5*(2), 136-157. https://doi.org/10.1108/ajeb-10-2020-0085

[8]. Bahnsen, A. C., Aouada, D., Stojanovic, A., & Ottersten, B. (2016). *Feature engineering strategies for credit card fraud detection* (Vol. 51). https://doi.org/10.1016/j.eswa.2015.12.030

[9]. Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). *Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI* (Vol. 58). https://doi.org/10.1016/j.inffus.2019.12.012

[10]. Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). Methodologies for data quality assessment and improvement. *ACM Computing Surveys*, *41*(3), Article 16. https://doi.org/10.1145/1541880.1541883

[11]. Benbasat, I., & Wang, W. (2005). Trust in and adoption of online recommendation agents. *Journal of the Association for Information Systems*, *6*(3). https://doi.org/10.17705/1jais.00065

[12]. Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. (2011). *Data mining for credit card fraud: A comparative study* (Vol. 50). https://doi.org/10.1016/j.dss.2010.08.008

[13]. Bussmann, N., Giudici, P., Marinelli, D., & Papenbrock, J. (2020). *Explainable AI in fintech risk management* (Vol. 3). https://doi.org/10.3389/frai.2020.00026

[14]. Carcillo, F., Dal Pozzolo, A., Le Borgne, Y.-A., Caelen, O., Mazzer, Y., & Bontempi, G. (2018). *Streaming active learning strategies for real-life credit card fraud detection*. https://doi.org/10.1007/s41060-018-0116-z

[15]. Carcillo, F., Le Borgne, Y.-A., Caelen, O., Kessaci, Y., Oblé, F., & Bontempi, G. (2021). *Combining unsupervised and supervised learning in credit card fraud detection* (Vol. 557). https://doi.org/10.1016/j.ins.2019.05.042

[16]. Chang, C.-C., & Lin, C.-J. (2011). *LIBSVM: A library for support vector machines* (Vol. 2). https://doi.org/10.1145/1961189.1961199

[17]. Dal Pozzolo, A., Boracchi, G., Caelen, O., Alippi, C., & Bontempi, G. (2018). *Credit card fraud detection: A realistic modeling and a novel learning strategy* (Vol. 29). https://doi.org/10.1109/tnnls.2017.2736643

[18]. Dalla Pellegrina, L., & Masciandaro, D. (2009). The risk-based approach in the new European anti-money laundering legislation: A law and economics view. *Review of Law & Economics*, *5*(2), 931-952. https://doi.org/10.2202/1555-5879.1422

[19]. Fahimul, H. (2022). Corpus-Based Evaluation Models for Quality Assurance Of AI-Generated ESL Learning Materials. *Review of Applied Science and Technology*, *1*(04), 183–215. https://doi.org/10.63125/m33q0j38

[20]. Fahimul, H. (2023). Explainable AI Models for Transparent Grammar Instruction and Automated Language Assessment. *American Journal of Interdisciplinary Studies*, *4*(01), 27-54. https://doi.org/10.63125/wttvnz54

[21]. Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861-874. https://doi.org/10.1016/j.patrec.2005.10.010

[22]. Faysal, K., & Tahmina Akter Bhuya, M. (2023). Cybersecure Documentation and Record-Keeping Protocols For Safeguarding Sensitive Financial Information Across Business Operations. *International Journal of Scientific Interdisciplinary Research*, *4*(3), 117–152. https://doi.org/10.63125/cz2gwm06

[23]. Fiore, U., De Santis, A., Perla, F., Zanetti, P., & Palmieri, F. (2019). *Using generative adversarial networks for improving classification effectiveness in credit card fraud detection* (Vol. 479). https://doi.org/10.1016/j.ins.2017.12.030

[24]. Fritz-Morgenthal, S., & Tenney, E. (2022). *Trustworthy AI: A perspective from the insurance industry* (Vol. 5). https://doi.org/10.3389/frai.2022.779799

[25]. Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Pedreschi, D., & Giannotti, F. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys*, *51*(5), Article 93. https://doi.org/10.1145/3236009

[26]. Habibullah, S. M., & Aditya, D. (2023). Blockchain-Orchestrated Cyber-Physical Supply Chain Networks with Byzantine Fault Tolerance For Manufacturing Robustness. *Journal of Sustainable Development and Policy*, *2*(03), 34-72. https://doi.org/10.63125/057vwc78

[27]. Hadji Misheva, B., & Papenbrock, J. (2022). Editorial: Explainable, trustworthy, and responsible AI for the financial service industry. *Frontiers in Artificial Intelligence*, *5*, 902519. https://doi.org/10.3389/frai.2022.902519

[28]. Hammad, S. (2022). Application of High-Durability Engineering Materials for Enhancing Long-Term Performance of Rail and Transportation Infrastructure. *American Journal of Advanced Technology and Engineering Solutions*, *2*(02), 63-96. https://doi.org/10.63125/4k492a62

[29]. Hammad, S., & Muhammad Mohiul, I. (2023). Geotechnical And Hydraulic Simulation Models for Slope Stability And Drainage Optimization In Rail Infrastructure Projects. *Review of Applied Science and Technology*, *2*(02), 01–37. https://doi.org/10.63125/jmx3p851

[30]. Haque, B. M. T., & Md. Arifur, R. (2020). Quantitative Benchmarking of ERP Analytics Architectures: Evaluating Cloud vs On-Premises ERP Using Cost-Performance Metrics. *American Journal of Interdisciplinary Studies*, *1*(04), 55-90. https://doi.org/10.63125/y05j6m03

[31]. Haque, B. M. T., & Md. Arifur, R. (2021). ERP Modernization Outcomes in Cloud Migration: A Meta-Analysis of Performance and Total Cost of Ownership (TCO) Across Enterprise Implementations. *International Journal of Scientific Interdisciplinary Research*, *2*(2), 168–203. https://doi.org/10.63125/vrz8hw42

[32]. Haque, B. M. T., & Md. Arifur, R. (2023). A Quantitative Data-Driven Evaluation of Cost Efficiency in Cloud and Distributed Computing for Machine Learning Pipelines. *American Journal of Scholarly Research and Innovation*, *2*(02), 449–484. https://doi.org/10.63125/7tkcs525

[33]. He, P. (2005). The suspicious transactions reporting system. *Journal of Money Laundering Control*, *8*(3), 252-259. https://doi.org/10.1108/13685200510620948

[34]. Jabed Hasan, T., & Waladur, R. (2022). Advanced Cybersecurity Architectures for Resilience in U.S. Critical Infrastructure Control Networks. *Review of Applied Science and Technology*, *1*(04), 146–182. https://doi.org/10.63125/5rvjav10

[35]. Jahangir, S., & Hammad, S. (2024). A Meta-Analysis of OSHA Safety Training Programs and their Impact on Injury Reduction and Safety Compliance in U.S. Workplaces. *International Journal of Scientific Interdisciplinary Research*, *5*(2), 559–592. https://doi.org/10.63125/8zxw0h59

[36]. Jahangir, S., & Muhammad Mohiul, I. (2023). EHS Analytics for Improving Hazard Communication, Training Effectiveness, and Incident Reporting in Industrial Workplaces. *American Journal of Interdisciplinary Studies*, *4*(02), 126-160. https://doi.org/10.63125/ccy4x761

[37]. Jha, S., Guillen, M., & Westland, J. C. (2012). Employing transaction aggregation strategy to detect credit card fraud. *Expert Systems with Applications*, *39*(16), 12650-12657. https://doi.org/10.1016/j.eswa.2012.05.018

[38]. Kaur, H., Nori, H., Jenkins, S., Caruana, R., Wallach, H., & Wortman Vaughan, J. (2020). Interpreting interpretability: Understanding data scientists' use of interpretability tools for machine learning. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems,

[39]. Khajeh-Hosseini, A., Greenwood, D., & Sommerville, I. (2016). An exploration of the determinants for decision to migrate existing resources to cloud computing. *Journal of Cloud Computing*, *5*, 23. https://doi.org/10.1186/s13677-016-0072-x

[40]. Krivko, M. (2010). A hybrid model for plastic card fraud detection systems. *Expert Systems with Applications*, *37*(8), 6070-6076. https://doi.org/10.1016/j.eswa.2010.02.119

[41]. Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation forest. 2008 Eighth IEEE International Conference on Data Mining,

[42]. Masud, R., & Hammad, S. (2024). Computational Modeling and Simulation Techniques For Managing Rail–Urban Interface Constraints In Metropolitan Transportation Systems. *American Journal of Scholarly Research and Innovation*, *3*(02), 141–178. https://doi.org/10.63125/pxet1d94

[43]. Md Ashraful, A., Md Fokhrul, A., & Md Fardaus, A. (2020). Predictive Data-Driven Models Leveraging Healthcare Big Data for Early Intervention And Long-Term Chronic Disease Management To Strengthen U.S. National Health Infrastructure. *American Journal of Interdisciplinary Studies*, *1*(04), 26-54. https://doi.org/10.63125/1z7b5v06

[44]. Md Fokhrul, A., Md Ashraful, A., & Md Fardaus, A. (2021). Privacy-Preserving Security Model for Early Cancer Diagnosis, Population-Level Epidemiology, And Secure Integration into U.S. Healthcare Systems. *American Journal of Scholarly Research and Innovation*, *1*(02), 01–27. https://doi.org/10.63125/q8wjee18

[45]. Md Harun-Or-Rashid, M., Mst. Shahrin, S., & Sai Praveen, K. (2023). Integration Of IOT And EDGE Computing For Low-Latency Data Analytics In Smart Cities And Iot Networks. *Journal of Sustainable Development and Policy*, *2*(03), 01-33. https://doi.org/10.63125/004h7m29

[46]. Md Harun-Or-Rashid, M., & Sai Praveen, K. (2022). Data-Driven Approaches To Enhancing Human–Machine Collaboration In Remote Work Environments. *International Journal of Business and Economics Insights*, *2*(3), 47-83. https://doi.org/10.63125/wt9t6w68

[47]. Md, K., & Sai Praveen, K. (2024). Hybrid Discrete-Event And Agent-Based Simulation Framework (H-DEABSF) For Dynamic Process Control In Smart Factories. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *4*(1), 72–96. https://doi.org/10.63125/wcqq7x08

[48]. Md. Akbar, H., & Farzana, A. (2023). Predicting Suicide Risk Through Machine Learning–Based Analysis of Patient Narratives and Digital Behavioral Markers in Clinical Psychology Settings. *Review of Applied Science and Technology*, *2*(04), 158–193. https://doi.org/10.63125/mqty9n77

[49]. Md. Arifur, R., & Haque, B. M. T. (2022). Quantitative Benchmarking of Machine Learning Models for Risk Prediction: A Comparative Study Using AUC/F1 Metrics and Robustness Testing. *Review of Applied Science and Technology*, *1*(03), 32–60. https://doi.org/10.63125/9hd4e011

[50]. Md. Towhidul, I., Alifa Majumder, N., & Mst. Shahrin, S. (2022). Predictive Analytics as A Strategic Tool For Financial Forecasting and Risk Governance In U.S. Capital Markets. *International Journal of Scientific Interdisciplinary Research*, *1*(01), 238–273. https://doi.org/10.63125/2rpyze69

[51]. Mostafa, K. (2023). An Empirical Evaluation of Machine Learning Techniques for Financial Fraud Detection in Transaction-Level Data. *American Journal of Interdisciplinary Studies*, *4*(04), 210-249. https://doi.org/10.63125/60amyk26

[52]. 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* (Vol. 50). https://doi.org/10.1016/j.dss.2010.08.006

[53]. Omar, N., & Johari, Z. A. (2015). An international analysis of FATF recommendations and compliance by DNFBPs. *Procedia Economics and Finance*, *28*, 14-23. https://doi.org/10.1016/s2212-5671(15)01076-x

[54]. Petter, S., & McLean, E. R. (2009). A meta-analytic assessment of the DeLone and McLean IS success model: An examination of IS success at the individual level. *Information & Management*, *46*(3), 159-166. https://doi.org/10.1016/j.im.2008.12.006

[55]. Ramdani, B., & Kawalek, P. (2007). SME adoption of enterprise systems in the Northwest of England: An environmental, technological and organizational perspective. In T. McMaster, D. Wastell, E. Ferneley, & J. I. DeGross (Eds.), *Organizational Dynamics of Technology-Based Innovation: Diversifying the Research Agenda* (pp. 409-429). Springer. https://doi.org/10.1007/978-0-387-72804-9_27

[56]. Ratul, D., & Subrato, S. (2022). Remote Sensing Based Integrity Assessment of Infrastructure Corridors Using Spectral Anomaly Detection and Material Degradation Signatures. *American Journal of Interdisciplinary Studies*, *3*(04), 332-364. https://doi.org/10.63125/1sdhwn89

[57]. Rauf, M. A. (2018). A needs assessment approach to english for specific purposes (ESP) based syllabus design in Bangladesh vocational and technical education (BVTE). *International Journal of Educational Best Practices*, *2*(2), 18-25.

[58]. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,

[59]. Rifat, C., & Jinnat, A. (2022). Optimization Algorithms for Enhancing High Dimensional Biomedical Data Processing Efficiency. *Review of Applied Science and Technology*, *1*(04), 98–145. https://doi.org/10.63125/2zg6x055

[60]. Rifat, C., & Khairul Alam, T. (2022). Assessing The Role of Statistical Modeling Techniques in Fraud Detection Across Procurement And International Trade Systems. *American Journal of Interdisciplinary Studies*, *3*(02), 91-125. https://doi.org/10.63125/gbdq4z84

[61]. Rifat, C., & Rebeka, S. (2023). The Role of ERP-Integrated Decision Support Systems in Enhancing Efficiency and Coordination In Healthcare Logistics: A Quantitative Study. *International Journal of Scientific Interdisciplinary Research*, *4*(4), 265–285. https://doi.org/10.63125/c7srk144

[62]. Rifat, C., & Rebeka, S. (2024). Integrating Artificial Intelligence and Advanced Computing Models to Reduce Logistics Delays in Pharmaceutical Distribution. *American Journal of Health and Medical Sciences*, *5*(03), 01–35. https://doi.org/10.63125/t1kx4448

[63]. Rocha-Salazar, J.-J., Segovia-Vargas, M.-J., & Camacho-Miñano, M.-M. (2021). *Money laundering and terrorism financing detection using neural networks and an abnormality indicator* (Vol. 169). https://doi.org/10.1016/j.eswa.2020.114470

[64]. Sai Praveen, K. (2024). AI-Enhanced Data Science Approaches For Optimizing User Engagement In U.S. Digital Marketing Campaigns. *Journal of Sustainable Development and Policy*, *3*(03), 01-43. https://doi.org/10.63125/65ebsn47

[65]. Sánchez, D., Vila, M. A., Cerda, L., & Serrano, J. M. (2009). Association rules applied to credit card fraud detection. *Expert Systems with Applications*, *36*(2), 3630-3640. https://doi.org/10.1016/j.eswa.2008.02.001

[66]. Shehwar, D., & Nizamani, S. A. (2024). Power Dynamics in Indian Ocean: US Indo-Pacific Strategic Report and Prospects for Pakistan's National Security. *Government: Research Journal of Political Science*, *13*.

[67]. Shoflul Azam, T., & Md. Al Amin, K. (2024). Quantitative Study on Machine Learning-Based Industrial Engineering Approaches For Reducing System Downtime In U.S. Manufacturing Plants. *International Journal of Scientific Interdisciplinary Research*, *5*(2), 526–558. https://doi.org/10.63125/kr9r1r90

[68]. Simonova, A. (2011). The risk-based approach to anti-money laundering: Problems and solutions. *Journal of Money Laundering Control*, *14*(4), 346-358. https://doi.org/10.1108/13685201111173820

[69]. Tripp, J. F., Liu, Z., & Tan, Y. (2015). Technology, humanness, and trust: Rethinking trust in technology. *Journal of the Association for Information Systems*, *16*(10), 880-918. https://doi.org/10.17705/1jais.00411

[70]. Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly*, *36*(1), 157-178. https://doi.org/10.2307/41410412

[71]. Vorobyev, I., & Krivitskaya, A. (2022). Reducing false positives in bank anti-fraud systems based on rule induction in distributed tree-based models. *Computers & Security*, *120*, 102786. https://doi.org/10.1016/j.cose.2022.102786

[72]. Whitrow, C., Hand, D. J., Juszczak, P., Weston, D., & Adams, N. M. (2009). Transaction aggregation as a strategy for credit card fraud detection. *Data Mining and Knowledge Discovery*, *18*, 30-55. https://doi.org/10.1007/s10618-008-0116-z

[73]. Zaman, M. A. U., Sultana, S., Raju, V., & Rauf, M. A. (2021). Factors İmpacting the Uptake of İnnovative Open and Distance Learning (ODL) Programmes in Teacher Education. *Turkish Online Journal of Qualitative Inquiry*, *12*(6).

[74]. Zhang, Y., & Trubey, P. (2019). *Machine learning and sampling scheme: An empirical study of money laundering detection* (Vol. 54). https://doi.org/10.1007/s10614-018-9864-z

[75]. Zhang, Z., Zhou, X., Zhang, X., Wang, L., & Wang, P. (2018). A model based on convolutional neural network for online transaction fraud detection. *Security and Communication Networks*, *2018*, 5680264. https://doi.org/10.1155/2018/5680264

[76]. Zhu, K., Kraemer, K. L., & Xu, S. (2006). The process of innovation assimilation by firms in different countries: A technology diffusion perspective on e-business. *Management Science*, *52*(10), 1557-1576. https://doi.org/10.1287/mnsc.1050.0487