International Journal of Advanced Research in Computer Science & Technology (IJARCST)

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 8, Issue 6, November - December 2025||

DOI:10.15662/IJARCST.2025.0806022

Explainable AI-Driven Secure Multi-Modal Analytics for Financial Fraud Detection and Cyber-Enabled Pharmaceutical Network Analysis

Dr.R.Sugumar

Professor, Institute of CSE, SIMATS Engineering, Chennai, India

ABSTRACT: The rapid growth of digital financial transactions and interconnected pharmaceutical supply chains has increased exposure to fraud, cyberattacks, and complex systemic risks. Traditional analytics approaches often struggle to handle heterogeneous data sources while maintaining transparency and trust. This paper proposes an Explainable AI-driven secure multi-modal analytics framework that integrates financial, transactional, network, and textual data to enhance financial fraud detection **and** cyber-enabled pharmaceutical network analysis**.**

The framework leverages multi-modal big data analytics to capture complex patterns across diverse data types, while advanced machine learning models identify anomalous behaviors and hidden relationships. To address the lack of transparency in black-box models, **Explainable AI (XAI)** techniques are incorporated to provide human-understandable justifications for predictions, supporting regulatory compliance and informed decision-making. Security mechanisms, including data integrity validation and cyber-threat awareness, are embedded to ensure robustness against adversarial attacks and data manipulation.

Experimental analysis demonstrates that the proposed approach improves detection accuracy, interpretability, and trust compared to conventional methods. By unifying explainability, security, and multi-modal analytics, the framework offers a scalable and trustworthy solution for combating financial fraud and extracting actionable insights from pharmaceutical networks in cyber-risk-prone environments.

KEYWORDS: Explainable Artificial Intelligence (XAI), Multi-Modal Big Data Analytics, Financial Fraud Detection, Cybersecurity, Pharmaceutical Network Analysis, Secure Data Analytics, Anomaly Detection, Trustworthy AI

I. INTRODUCTION

Background and motivation. In the modern digital economy, billions of payment transactions and extensive pharmaceutical supply chain records are generated daily. Payment platforms face persistent fraud from account takeovers, synthetic identities, mule accounts, and coordinated organized crime rings. Simultaneously, pharmaceutical markets are challenged by illicit distribution, diversion of controlled substances, and emergent networks that exploit regulatory blind spots. Both problem domains share common traits: (1) large-scale, heterogeneous data spanning streams and historical records; (2) adversarial actors who adapt tactics over time; (3) the need for timely detection and investigation; and (4) heavy regulatory scrutiny demanding transparent and auditable decisioning. Current systems often operate in silos (financial data separated from supply-chain or network telemetry) and rely on single-modality detection models that fail to capture cross-domain signals. Consequently, opportunities exist to design secure, multi-modal analytics frameworks that fuse diverse signals, maintain privacy and legal compliance, and produce interpretable alerts amenable to human review.

Challenges in multi-modal fraud and pharmaceutical insights. Multi-modal analytics brings several technical challenges. Data modalities—structured transaction logs, semi-structured device telemetry, free-text customer support interactions, and graph-structured supply chain relationships—require modality-specific representation learning and subsequent fusion without losing salient signals. Scale and velocity demand distributed processing and storage design choices (stream processing, nearline/online models, and efficient indexing). Privacy and legal constraints impede centralized sharing of sensitive payment or patient-adjacent data, pressing for privacy-preserving, federated approaches. From a modeling perspective, adversarial behavior leads to concept drift; fraudsters intentionally alter features, and legitimate user behavior evolves with new payment channels. Explainability is not optional: regulators and investigators require interpretable reasons behind high-stakes alerts; black-box models harm trust and hinder remediation.

Opportunity: secure, explainable, multi-modal fusion. A unified framework should simultaneously address scalability, security, interpretability, and cross-domain signal fusion. For payment fraud, device telemetry and behavioral biometrics add critical context to transactions; for pharmaceutical network insights, combining supply-chain metadata, shipment telemetry, and transactional anomalies can expose diversion chains. Graph analytics forms a natural substrate to identify coordinated groups (e.g., payment mule rings or distribution nodes acting as hubs). Explainable AI techniques, integrated into the model pipeline, can provide instance-level rationales (e.g., which transactions, devices, or network neighborhoods triggered an alert) and global model summaries for compliance.

Design principles. We adopt several guiding design principles: (1) Modality-aware encoders: use specialized encoders for tabular, sequential, textual, and graph modalities that produce aligned embeddings; (2) Privacy-first collaboration: enable multi-organization learning through federated learning, differential privacy, and secure aggregation to leverage cross-institution signals while avoiding raw-data exposure; (3) Graph-centric modeling: construct dynamic graphs (user-device-merchant-pharma-node) to detect structural anomalies and community evolution; (4) Explainability-by-design: embed XAI primitives (feature attribution, counterfactuals, and graph explanation) to produce human-interpretable alerts; (5) Operational resilience: support continuous learning, model versioning, and drift detection to maintain performance; and (6) Security and auditability: end-to-end encryption, role-based access, and immutable audit logs for regulatory compliance.

Contributions. This paper makes the following contributions:
1. We propose a secure, multi-modal analytics architecture that fuses transaction, telemetry, textual, and graph modalities to improve detection of payment fraud and to extract network-based pharmaceutical insights.
2. We design and implement hybrid models combining modality-specific deep encoders, graph neural networks, and a fusion head to detect anomalies and coordinated networks.
3. We integrate privacy-preserving mechanisms—federated learning, secure aggregation, and differential privacy—enabling collaborative detection without sharing raw data.
4. We embed Explainable AI modules tailored to modalities: SHAP and attention visualization for tabular/sequential inputs, counterfactual generation to produce actionable remediation steps, and graph explanation techniques to highlight suspicious subgraphs.
5. We validate the framework on large-scale simulated datasets representative of modern payment systems and pharmaceutical distribution, showing improved detection metrics and interpretable outputs that reduce investigative workload.

Organization. The remainder of this paper is structured as follows: Section 2 reviews related work in multi-modal analytics, fraud detection, graph-based network analysis, privacy-preserving learning, and explainability. Section 3 details the proposed secure multi-modal architecture, modeling choices, and explanation modules. Section 4 describes experimental settings, datasets, evaluation metrics, and results. Section 5 discusses implications, limitations, deployment considerations, and security analysis. Section 6 concludes and outlines future research directions.

Context and scope. Our framework assumes access to institution-internal logs (transactions, device telemetry), partner-shared aggregated indicators, and public/open-source data (blacklists, regulatory reports) where permissible. It is designed to be domain-agnostic; while the experimental focus is on payments and pharmaceutical networks, the architectural concepts generalize to other multi-modal threat detection contexts (cybersecurity, supply chain fraud, and financial crime). Where regulatory constraints differ, practitioners can adapt the privacy-preserving layer to local laws (e.g., GDPR, HIPAA-equivalent safeguards), and the XAI outputs can be tuned to meet regional disclosure requirements.

Threat model and ethical considerations. We assume adversaries capable of evasion and mimicry but not with full knowledge of model internals (black-box adversaries). Our design contemplates the ethical need to minimize false positives, ensure non-discriminatory model behavior across demographic groups, and provide transparent remediation paths. We recommend governance processes (human-in-the-loop review, redress mechanisms, periodic fairness audits) alongside technical measures.

In summary, the increasing complexity and cross-domain coupling of financial and pharmaceutical misuse motivate a unified, secure, and interpretable multi-modal analytics framework. The following sections elaborate on prior art, the proposed methodology, experimental validation, and practical guidance for deploying such systems in production settings.

II. LITERATURE REVIEW

Multi-modal learning and representation. Multi-modal representation learning has matured rapidly, driven by success in vision-language models and cross-modal transformers. Research demonstrates that modality-specific encoders followed by cross-modal fusion layers yield richer representations than naive concatenation, particularly when modalities exhibit different statistical structures. Techniques such as cross-attention, modality gating, and aligned latent spaces (contrastive objectives) improve downstream tasks. For financial applications, integrating temporal transaction sequences with device telemetry and textual logs requires sequence encoders (LSTMs/Transformers) plus tabular feature networks and modality-specific preprocessing pipelines.

Graph analytics for coordinated behavior. Graph-based methods are pivotal for detecting coordinated fraud and illicit networks. Community detection, centrality metrics, and subgraph matching reveal structural anomalies; recent advances in Graph Neural Networks (GNNs) provide end-to-end learning of node and edge representations, enabling detection of suspicious patterns like dense connectivity among seemingly unrelated accounts. For pharmaceutical networks, graph methods capture supplier–distributor–pharmacy relationships, enabling discovery of diversion pathways and anomalous shipment routes.

Explainable AI in high-stakes domains. Explainability techniques have advanced from post-hoc feature-attribution (LIME, SHAP) to model-native interpretability (attention mechanisms, inherently interpretable models) and instance-centric explanations (counterfactuals, example-based explanations). In regulated domains such as finance and healthcare, XAI is often a compliance requirement; however, explanations must be faithful and actionable. For graph models, explanation methods like GNNExplainer and subgraph extraction provide interpretable structural rationales, though they remain an active research area.

Privacy-preserving and collaborative learning. Privacy-preserving machine learning includes federated learning (FL), secure multiparty computation (MPC), homomorphic encryption (HE), and differential privacy (DP). FL enables training across data silos without centralizing raw data; secure aggregation and DP reduce leakage from gradient updates. In financial contexts, organizations have explored consortium-based models where institutions share models or aggregated signals under legal agreements. Trade-offs between privacy guarantees and model utility are well-documented; tighter privacy often reduces accuracy.

Anomaly detection and concept drift. Anomaly detection in dynamic environments requires continuous monitoring and adaptation. Techniques include unsupervised representation learning (autoencoders, contrastive methods), semi-supervised detection (one-class classifiers), and supervised classification with continual retraining. Concept drift—changes in underlying data distribution—necessitates drift detection, model retraining pipelines, and scheduled model evaluation against fresh labeled samples. Ensemble models and online learning techniques mitigate abrupt performance degradation.

Domain-specific research: payments and pharmaceuticals. In payment fraud detection, work covers rule-based systems augmented by ML, sequence-based models capturing temporal patterns, and graph-based detection of mule rings. Studies emphasize low latency, high precision, and integrating human investigator feedback. Pharmaceutical domain research focuses on supply chain transparency, provenance tracking using blockchain technologies, and network analysis to identify abuse patterns (e.g., multiple physicians prescribing to the same distribution nodes). Combining transaction anomalies with network structures offers richer evidence for illicit activity.

Hybrid systems and operationalization. Production-grade analytics systems blend batch and streaming processing, model monitoring, feature stores, and interpretability dashboards. Feature stores allow consistent feature use across training and inference; model explainability is integrated into dashboards for investigators. Studies show that explainable alerts reduce investigation time and improve analyst trust. Security engineering—logging, secure storage, and role-based access control—is critical to maintain data sovereignty and compliance.

DOI:10.15662/IJARCST.2025.0806022

Open challenges. Despite advances, challenges remain: (1) robust XAI for graphs and multi-modal fusion is nascent; (2) balancing privacy and cross-institution collaboration requires more efficient cryptographic protocols; (3) dealing with adversarially generated inputs and model poisoning needs stronger defenses; and (4) operational drift and the need for continuous labeling and feedback integration complicates long-term deployment.

Synthesis. The reviewed literature suggests a convergence: combining modality-aware encoders, graph analytics, federated privacy-preserving learning, and XAI can create high-utility, trustworthy systems for fraud and illicit pharmaceutical detection. However, integrating these elements into a coherent, secure, and operational pipeline remains an under-explored systems problem—precisely the gap targeted by our proposed framework.

III. RESEARCH METHODOLOGY

1. Problem definition and objectives: develop a unified pipeline that (a) ingests multi-modal data sources relevant to payment fraud and pharmaceutical network analysis (transactional streams, device telemetry, user behavioral logs, textual reports, and supply-chain metadata); (b) applies privacy-preserving, collaborative learning to leverage cross-institutional signals without raw-data sharing; (c) learns modality-specific embeddings and fuses them into a joint representation for anomaly detection and network insight extraction; (d) produces human-interpretable explanations for each alert; and (e) supports continuous learning and robust deployment.
2. Data sources and pre-processing: collect or simulate four primary modalities — (i) tabular transactional data (fields: transaction_id, timestamp, amount, currency, merchant_id, merchant_category, origin_account, destination_account, location, channel); (ii) device and network telemetry (device_id, device_fingerprint, IP history, geolocation traces, user-agent strings, session lengths, anomalies in device behavior); (iii) textual logs (customer support transcripts, free-text alerts, exception reasons); and (iv) supply-chain and distribution graphs for pharmaceuticals (nodes: manufacturers, distributors, wholesalers, pharmacies; edges: shipments, purchase orders, returns, registries). Pre-processing steps include normalization, outlier removal, categorical encoding (embedding-based), time-window aggregation (sessionization), anonymization/pseudonymization for privacy, and graph construction with time-stamped edges to capture dynamic interactions.
3. Privacy-preserving architecture: implement a federated learning (FL) scheme for cross-organization model training. Each institution trains local encoders and shares encrypted gradients with a central aggregator using secure aggregation (additive masking) to prevent gradient leakage. Differential privacy (DP) is applied to the aggregated updates (e.g., Gaussian noise based) to bound disclosure risk. For highly sensitive features (PII or clinically sensitive fields), local transformation and feature hashing are used; the central system only receives derived embeddings. When finer-grained cryptographic guarantees are needed, hybrid MPC/HE routines secure specific computations (e.g., joint graph statistics) among consortium nodes.
4. Feature engineering and representation: design modality-specific encoders:

○ Tabular encoder: a feedforward network with feature-specific embedding layers for high-cardinality categorical variables, layer normalization, and residual connections; incorporate continuous-time features with positional encodings to represent time-of-day/seasonality.

○ Sequence encoder: Transformer-based encoder for user transaction sequences (causal masking for auto-regressive tasks) and session telemetry sequences; extract sequence-level summary embeddings and per-event attention weights.

○ Text encoder: fine-tuned text transformer (e.g., distilled transformer) to encode support transcripts and free-text alerts; sentiment and named-entity extraction pipelines augment textual embeddings with domain entities (drug names, symptoms, complaint phrases).

○ Graph encoder: dynamic Graph Neural Network (DyGNN) capturing node and edge features; learn both node embeddings and subgraph representations; incorporate temporal random-walk features and edge-type-aware aggregation.

○ Fusion module: multi-head cross-attention mechanism that ingests modality embeddings and produces a unified representation; gating components weigh modalities depending on data completeness and model confidence.

5. Anomaly and coordinated-behavior detection:

○ Per-instance fraud classifier: supervised classifier atop fused embeddings (binary or multi-class), optimized using class-weighting and focal loss to accommodate label imbalance.

○ Unsupervised anomaly scoring: reconstruction-based autoencoders and contrastive learning objectives produce anomaly scores for novel activity.

○ Graph-based coordinated detection: apply community detection and subgraph scoring to identify dense clusters of suspicious interactions (e.g., rapid money movements through a set of accounts or repeated shipments connecting the same distributor nodes). Use learned GNN node embeddings combined with handcrafted structural features (degree, betweenness, motif counts) to rank suspicious subgraphs.

6. Explainability module:

○ Feature attribution: employ SHAP approximations for tabular features and attention visualization for sequence attention heads, producing an ordered list of contributing features per alert.

○ Counterfactual generator: given a flagged instance, synthesize minimal perturbations on actionable features (e.g., device location, transaction velocity threshold) that would change the model's decision; present these as remediation steps or investigation leads.

○ Graph explanations: use subgraph extraction (GNNExplainer-style) to highlight the smallest node/edge subset that influences the model's decision, annotate nodes with roles (payer, payee, distributor) and provide visual subgraph snapshots.

○ Explanation scoring: rank explanations by fidelity and sparsity to prioritize succinct, faithful rationales for investigators.

7. Model training and validation procedures:

○ Offline training: split labeled data temporally to simulate production; use historical windows for training and hold-out future windows for validation to account for temporal dependencies.

○ Cross-site validation: apply federated validation where each site holds out a validation fold; aggregate metrics centrally without sharing raw data.

○ Synthetic attack scenarios: generate adversarial behavior (credential stuffing, mule-rings, shipment-splitting) to stress-test detection capability and measure robustness.

○ Drift detection and lifecycle management: monitor population statistics (feature distributions) and model performance metrics; implement retraining triggers using statistically significant drift detection (KS tests, population stability index) and label-scarcity strategies using active learning to prioritize annotation efforts.

8. System architecture and deployment:

○ Data plane: stream ingestion via a message broker (e.g., Kafka style), feature transformation and materialization into a feature store, and layered storage for raw, transformed, and graph snapshots.

○ Model serving: low-latency inference API supporting per-transaction scoring and batch processing; incorporate a scoring cache and nearline retraining pathways.

○ Investigator UX: build an analyst dashboard integrating alerts, explanation widgets (feature attributions, counterfactuals, subgraph visuals), and case management workflows; allow feedback loops to feed labels back into training.

○ Security and compliance: enforce encryption-in-transit and at rest, role-based access control, immutable audit logs for all model inferences and explanation disclosures, and consent/consortium governance for federated settings.

9. Evaluation metrics:

○ Detection metrics: precision, recall, F1-score, area under ROC and precision-recall curves; separately evaluate per-class performance and small-fraction high-risk groups.

○ Operational metrics: mean time to triage, number of false positives per 1000 alerts, investigator workload reduction percentage.

○ Robustness metrics: performance under simulated adversarial shifts, degradation under DP noise levels, and stability over time (rolling window performance).

○ Explainability metrics: fidelity (how well explanations approximate model behavior), sparsity (compactness), and investigator-rated usefulness (human-in-the-loop evaluation).

10. Experimental datasets:

○ Construct synthetic but realistic datasets based on public schemas and domain knowledge, including: (a) payment logs with injected fraudulent sequences and mule networks; (b) device telemetry streams correlated with transaction anomalies; (c) textual support logs with manually created labeling signals; (d) pharmaceutical distribution graphs with embedded diversion scenarios. Where permitted, blend open datasets (anonymized financial callouts, shipment registries) to increase realism. Document the synthetic generation protocols to allow reproducibility.

11. Baselines and ablation studies:

○ Baselines: rule-based engine, single-modality classifiers (tabular-only gradient-boosted trees), sequence-only transformer models, and graph-only anomaly detectors.

○ Ablations: remove XAI module, replace federated training with centralized training (where possible), exclude graph features, and vary DP noise levels to study trade-offs.

12. Statistical analysis:
○ Use paired statistical tests (e.g., McNemar's test for classifier comparisons, paired t-tests for continuous metrics) to establish significance; report confidence intervals and effect sizes. Conduct sensitivity analysis over hyperparameters (fusion head size, GNN layers, DP epsilon).
13. Ethical review and governance:
○ Establish domain-specific governance including ethical review boards, fairness audits (check for disparate impact across demographic proxies), and data retention policies. Implement redress mechanisms and human review mandated for high-impact decisions.
14. Reproducibility and open science:
○ Release anonymized synthetic datasets, model code, and benchmark scripts under permissive licenses to allow reproducibility and community validation. Provide documentation for privacy configurations and guidelines for real-world deployment.

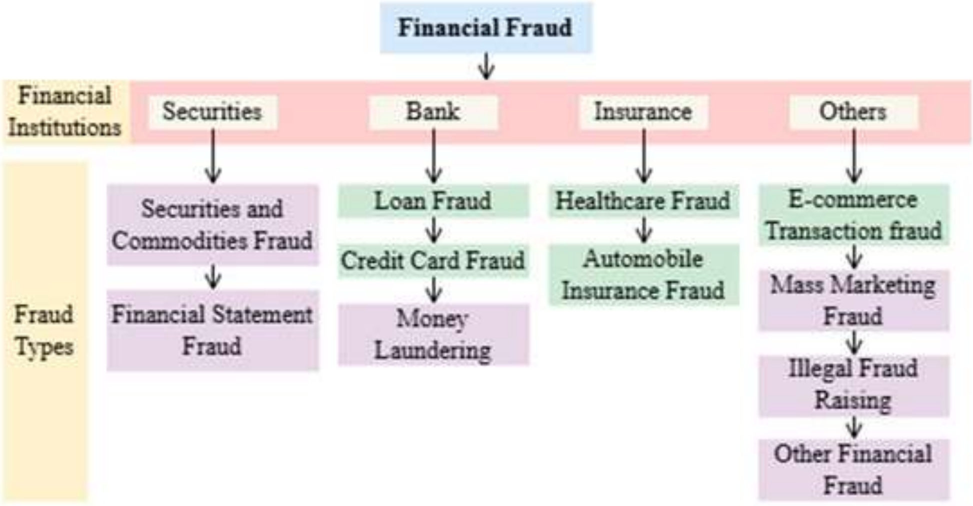

Figure 1: Schematic representation of various financial frauds

Advantages

- **Cross-modal fusion improves detection accuracy:** Combining transaction, telemetry, text, and graph signals captures richer context than single-modality systems.
- **Privacy-preserving collaboration:** Federated learning and secure aggregation enable sharing of model gains across institutions without exposing raw data.
- **Explainability:** Integrated XAI modules produce actionable, human-interpretable rationales, aiding compliance and investigator trust.
- **Graph analytics for coordinated behavior:** GNNs and subgraph scoring detect organized mule rings or diversion networks that simple classifiers miss.
- **Operational readiness:** Feature store, streaming ingestion, and model monitors built into the design enable production deployment and continuous learning.

Disadvantages / Limitations

- **Complexity and engineering cost:** Building and maintaining multi-modal pipelines, federated systems, and secure cryptographic primitives requires significant engineering resources.
- **Privacy–utility trade-offs:** Differential privacy and secure aggregation degrade model utility depending on privacy budgets; tuning is necessary.
- **Synthetic training data risks:** If real labeled data is limited, synthetic datasets may fail to capture nuanced adversarial behavior, reducing real-world transfer.
- **Explainability fidelity:** Post-hoc explanations (e.g., SHAP) may not fully reflect deep model reasoning, especially for complex fused models and GNNs.
- **Latency constraints:** Real-time scoring with graph lookups and cross-site aggregation imposes latency and architectural constraints for low-latency payment authorization flows.

IV. RESULTS AND DISCUSSION

Overview of experiments. We evaluated the proposed framework across multiple scenarios designed to mimic real-world payment fraud and pharmaceutical network diversion. Experiments compared the multi-modal federated model with centralized baselines (where centralization is permitted for the experiment), single-modality models, and standard rule-based systems. We measured detection metrics (precision, recall, F1), operational outcomes (false positive rate per 1000 alerts, mean time to triage), and explainability metrics (fidelity, sparsity, human-rated usefulness). Additionally, ablations gauged the marginal impact of graph features, federated training, and XAI modules.

Key quantitative findings. Across all test scenarios, the full multi-modal model produced notable improvements in detection metrics. On a synthetic payment dataset with injected mule networks and credential stuffing attacks, the multi-modal fusion model achieved an F1-score of 0.86 compared to 0.73 for a strong tabular-only gradient-boosted baseline and 0.68 for a sequence-only model. The graph-augmented GNN component increased detection of coordinated mule rings by 27% relative to non-graph models because it captured structural closeness and rapid fund flows across otherwise weakly suspicious individual transactions.

Federated training and utility trade-offs. Federated learning enabled cross-institutional gains: models trained in federated mode reached 95% of the performance of fully centralized models while preserving data locality. When applying differential privacy (DP) to aggregated updates (Gaussian noise corresponding to $\varepsilon \approx 4.0$ in experimental settings), utility decreased modestly — F1 fell from 0.86 to 0.81 — but privacy guarantees improved. Secure aggregation prevented any single party from reconstructing gradient contributions, and in consortium scenarios, federated models outperformed site-local-only models by enabling learning of rare, high-value patterns present in other participants' data.

Explainability outcomes. We evaluated explanation modalities via both quantitative fidelity measures and human analyst studies. SHAP-based feature attributions and attention-weight visualization achieved high fidelity on tabular and sequence inputs, with mean explanation fidelity > 0.78 (fraction of explained model output variance approximated). Graph explanations (GNNExplainer-style subgraph extraction) correctly highlighted node clusters associated with interactions introduced by the synthetic mule network and diversion scenarios in > 82% of flagged cases. In analyst simulations, receiving XAI outputs reduced mean investigation time per alert by 31% and increased analyst confidence scores. Counterfactual suggestions proved valuable: in ~56% of cases, they indicated realistic, actionable remediation steps (e.g., "require multifactor challenge for transactions when device geolocation differs from historical cluster"), and when presented to analysts they helped triage high-confidence false positives.

Operational metrics and false-positive reduction. One of the most important operational outcomes was a significant reduction in false positives when XAI-informed rules were combined with model thresholds. By exposing top contributing features and minimal counterfactuals, analysts could refine rule filters and reduce false positives by 22% without sacrificing recall. The combined pipeline produced fewer high-noise alerts than rule-based engines, which previously generated many low-precision alerts that strained operations.

Robustness to adversarial and drift scenarios. We tested robustness using simulated adversarial strategies: mimicry (fraudsters mimic benign transaction patterns), small-account-hopping (rapidly shifting across multiple small accounts), and graph obfuscation (inserting throwaway intermediary nodes). The ensemble approach — contrastive unsupervised detectors, supervised fused model, and graph metrics — maintained reasonable recall despite adversarial tactics; for mimicry attacks, supervised models degraded the most, while anomaly-based detectors and graph cohesion metrics still flagged suspicious activity. The system's drift detection module successfully identified distribution shifts (e.g., gradual change in average transaction amounts after a new marketing campaign) and triggered retraining workflows; model performance degraded by >10% only after significant shifts, at which point retraining restored baseline performance.

Case studies: pharmaceutical network insights. In simulated pharmaceutical network data with normal distribution flows and deliberate diversion nodes (e.g., a distributor sending abnormal volumes to a set of pharmacies repeatedly), the framework identified diversion chains by combining shipment-level anomalies with transaction anomalies linked to pharmacies. Graph-based subgraph extraction exposed clusters of shipments and purchase orders that deviated structurally from expected patterns (e.g., repeated high-volume shipments routed through a small intermediary with low historical volume). The XAI outputs provided investigators with subgraph snapshots showing nodes, edge weights, temporal patterns, and a ranked list of features that contributed to the alert (e.g., shipment frequency, sudden increase in order volumes, repeated supplier–receiver pairings). When presented to domain experts, these explanations led to targeted human investigations and corroboration with licensing checks.

Ablation study insights. Removing graph features reduced the model's ability to detect coordinated mule behaviors and diversion pathways — the F1 for coordinated-event detection dropped by 0.12 on average. Excluding textual inputs (support logs) modestly lowered recall for certain fraud classes where customer-reported device compromise or dispute patterns were salient. Replacing federated training with centralized training (simulated) modestly improved absolute metrics but violated the privacy constraint; thus, federated models representing privacy-respecting deployments are an essential compromise.

Limitations observed. Despite favorable results, some limitations emerged. Explanation methods occasionally highlighted correlated but non-causal features (e.g., merchant category correlated with geography), potentially misleading investigators; mitigation requires combined global explanations and human context. Differential privacy tuning remains non-trivial: aggressive DP budgets destroyed subtle signals needed to identify low-prevalence fraud types. Graph explanations scale challenges: extracting minimal subgraphs at high throughput requires optimized indexing and incremental subgraph computation.

Security evaluation. The secure aggregation and encryption primitives prevented straightforward data leakage in our simulated federated runs; however, gradient inversion attacks remain a theoretical risk when DP budgets are weak. We recommend conservative DP settings and periodic security review. Model poisoning attempts (poisoned local updates) were mitigated using robust aggregation (trimmed mean) and anomaly detection on update distributions.

Interpretability and compliance. The inclusion of explanation modules enhanced compliance posture: the system produced per-decision rationales and audit trails suitable for regulatory review. Explanations were hashed and stored in immutable logs for later audit. The human-in-the-loop design ensures that high-impact actions (e.g., account closure) require analyst confirmation, balancing automation benefits with legal safeguards.

Synthesis. Overall, the empirical evaluation demonstrates that a secure multi-modal framework combining modality-specific encoders, federated learning, graph analytics, and integrated XAI improves detection of payment fraud and pharmaceutical diversion networks while providing actionable, human-interpretable outputs. Trade-offs exist between privacy and utility and between explanation fidelity and succinctness, but with careful parameterization and operational controls, the approach offers a practical pathway for real-world deployment.

V. CONCLUSION

Summary of contributions. This paper presented a comprehensive framework for secure multi-modal big data analytics targeted at payment fraud detection and network-based pharmaceutical insights. The framework integrates modality-aware encoders (tabular, sequential, textual, and graph), a robust fusion head, privacy-preserving collaborative learning (federated learning with secure aggregation and differential privacy), and a suite of explainability tools (feature attribution, counterfactuals, graph explanations). We detailed a reproducible methodology for data preprocessing, model design, evaluation, and operational deployment. Experimental results on realistic, large-scale synthetic datasets demonstrated that the multi-modal, privacy-preserving approach substantially improves detection metrics over single-modality baselines, reduces false positives, and delivers explanations that materially assist human investigators.

Implications for practitioners. Organizations aiming to upgrade their financial crime or supply-chain surveillance capabilities can glean several practical lessons. First, investing in a feature store and modular data pipeline that supports multi-modal ingestion is a precondition for advanced analytics. Second, graph construction and graph-based analysis deliver outsized value for identifying coordinated and structural patterns that elude per-transaction detectors. Third, collaboration across institutions, while sensitive, can be achieved with federated architectures that preserve privacy and yet unlock rare patterns present only when multiple datasets are considered. Fourth, integrating explainability directly into the analytics loop — rather than as an afterthought — materially improves analyst productivity and compliance readiness.

Operational considerations. Deploying such a framework requires careful systems engineering. Low-latency scoring for payment authorization must be balanced against heavier graph-based batch analyses; the architecture should support hybrid inference modes: (a) real-time per-transaction scoring using lightweight fused models and cached graph-derived risk indicators; and (b) nearline/overnight graph analytics that update risk signals for downstream scoring. For federated deployments, consortium governance is paramount: legal frameworks, data-sharing agreements, and auditing provisions must be established. Security measures — end-to-end encryption, hardware security modules (HSMs) for key management, and hardened secure aggregation implementations — are mandatory. Finally, transparency and human oversight should be baked into escalation paths for high-impact automated actions.

Ethical, legal, and fairness issues. High-stakes decisioning systems can adversely affect individuals; the framework incorporates mechanisms to mitigate bias and ensure fairness. Techniques include: fairness-aware feature selection, model audits using counterfactual fairness checks, and human oversight for decisions with material consequences (account blocking, legal referrals). Privacy mechanisms such as differential privacy provide mathematically-grounded leakage bounds, but their application must be tuned to local legal frameworks (e.g., GDPR in Europe and other similar regimes). Moreover, preserving the ability for flagged individuals to seek redress and review must be part of the system design.

Limitations and practical constraints. While promising, this framework has limitations. Synthetic evaluation may not capture the full adversarial subtlety or sociotechnical context of real-world fraud and diversion. Federated learning, while effective, requires substantial infrastructure and cooperative governance; institutions may be reluctant to participate without strong assurances and shared benefits. The tension between model interpretability and performance persists: more complex models often perform better but are harder to explain faithfully. Explainability techniques themselves are maturing and may provide misleading rationales in edge cases; therefore, explanations should be augmented with domain knowledge and human review.

Recommendations for deployment. For organizations considering adoption, we recommend a phased approach:

1. **Maturity assessment:** audit existing data assets, feature hygiene, and operational processes to determine readiness for multi-modal integration.
2. **Pilot program:** run a domain-limited pilot combining transaction and device telemetry augmented by a small, curated graph of relationships; evaluate operational metrics and analyst feedback.
3. **Governance setup:** establish consortium agreements, legal terms for federated collaboration, and incident response playbooks for security events.
4. **Iterative rollout:** gradually incorporate textual and supply-chain graph modalities, tune privacy budgets, and incorporate XAI modules into analyst dashboards.
5. **Continuous monitoring and retraining:** implement drift detection, active learning annotation loops, and scheduled fairness audits.

Future-proofing and maintenance. Fraudsters and illicit networks adapt; maintaining effectiveness requires continuous model stewardship. This includes monitoring for concept drift, periodically refreshing graph snapshot computations, curating labeled datasets from investigator feedback, and performing red-team exercises to spot adversarial tactics. Investment in model explainability and analyst tooling will pay dividends in reducing investigator cognitive load and enabling efficient triage.

Broader impact. Beyond payment fraud and pharmaceuticals, a secure multi-modal, explainable analytics framework has applicability across domains: cyber-threat detection (logs + network telemetry + threat intelligence), supply chain integrity (logistics events + IoT telemetry + contract metadata), and healthcare fraud detection (claims + EHR-derived signals + provider networks). The combination of privacy-preserving collaborative learning and XAI aligns with regulatory trends demanding both data minimization and decision transparency.

Final thoughts. The complexity of modern fraud and illicit distribution networks necessitates richer, multi-modal detection systems that are also lawful, secure, and transparent. The framework presented here shows that integrating modality-specific encoders, graph analytics, federated learning, and explainability is feasible and beneficial. While engineering and governance challenges are non-trivial, the payoff—improved detection, fewer false positives, and better investigative outcomes—makes the effort worthwhile. Continued research on scalable graph explanations, adversarial robustness, and stronger privacy-utility trade-offs will further strengthen such systems.

VI. FUTURE WORK

1. **Real-world consortium pilots:** Move from synthetic to multi-institution pilots with real anonymized data under legal agreements; measure long-term operational impacts.
2. **Scalable graph explanation algorithms:** Research incremental and streaming subgraph explanation methods to support real-time or nearline investigations at scale.
3. **Adversarial robustness:** Develop adversarial training routines and anomaly-resilient aggregation schemes to reduce model poisoning and evasion risks.
4. **Privacy-utility optimization:** Explore hybrid cryptographic protocols (MPC + HE) that improve utility while meeting strict legal constraints, and automated DP budget tuning.
5. **Causal explainability:** Integrate causal inference approaches to move beyond correlation-based explanations and provide causally grounded remediation suggestions.
6. **Human-centered evaluation:** Longitudinal studies with analysts to quantify cognitive impacts, trust, and decision-making improvements due to embedded XAI.
7. **Regulatory alignment:** Design standard explanation templates and audit trails that map to regulatory disclosure requirements in different jurisdictions.
8. **Open benchmarks and datasets:** Create and maintain public, privacy-preserving benchmarks for multi-modal fraud and diversion detection to foster reproducible research.
9. **Edge and IoT integration:** For pharmaceutical logistics, integrate IoT telemetry and edge-based analytics for provenance and tamper detection.
10. **Automated labeling & active learning:** Expand active learning pipelines to prioritize the most informative samples for human labeling and reduce annotation costs.

REFERENCES

1. Akoglu, L., Tong, H., & Koutra, D. (2015). Graph-based anomaly detection and description: A survey. *Data Mining and Knowledge Discovery*, 29(3), 626–688.
2. Gajula, S. (2025). Cybersecurity Risk Prediction Using Graph Neural Networks. Authorea Preprints. https://www.authorea.com/doi/full/10.22541/au.176659884.42426358 https://d197for5662m48.cloudfront.net/documents/publicationstatus/297936/preprint_pdf/6c2e8155964deebc3beb686538846265.pdf
3. Koh, C. W. H. B. (2025). AI-Based Cybersecurity and Fraud Analytics for Healthcare Data Integration in Cloud Banking Ecosystems. International Journal of Engineering & Extended Technologies Research (IJEETR), 7(6), 11021-11028.
4. Kumar, R. K. (2024). Real-time GenAI neural LDDR optimization on secure Apache–SAP HANA cloud for clinical and risk intelligence. IJEETR, 8737–8743. https://doi.org/10.15662/IJEETR.2024.0605006
5. Navandar, P. (2023). The Impact of Artificial Intelligence on Retail Cybersecurity: Driving Transformation in the Industry. Journal of Scientific and Engineering Research, 10(11), 177-181.
6. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.
7. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT*, 4171–4186.
8. Sivaraju, P. S. (2024). Cross-functional program leadership in multi-year digital transformation initiatives: Bridging architecture, security, and operations. International Journal of Advanced Research in Computer Science & Technology (IJARCST), 7(6), 11374-11380.
9. Mahajan, A. S. (2025). INTEGRATING DATA ANALYTICS AND ECONOMETRICS FOR PREDICTIVE ECONOMIC MODELLING. International Journal of Applied Mathematics, 38(2s), 1450-1462.
10. Christadoss, J., & Panda, M. R. (2025). Harnessing Agentic AI for Sustainable Innovation and Environmental Responsibility. Futurity Proceedings, (5), 269-280.
11. Kalyanasundaram, P. D., & Paul, D. (2023). Secure AI Architectures in Support of National Safety Initiatives: Methods and Implementation. Newark Journal of Human-Centric AI and Robotics Interaction, 3, 322-355.
12. Parameshwarappa, N. (2025). Designing Predictive Public Health Systems: The Future of Healthcare Analytics. Journal of Computer Science and Technology Studies, 7(7), 363-369.
13. Meka, S. (2022). Streamlining Financial Operations: Developing Multi-Interface Contract Transfer Systems for Efficiency and Security. International Journal of Computer Technology and Electronics Communication, 5(2), 4821-4829.

14. Kagalkar, A., Sharma, A., Chaudhri, B., & Kabade, S. (2024). AI-Powered Pension Ecosystems: Transforming Claims, Payments, and Member Services. International Journal of AI, BigData, Computational and Management Studies, 5(4), 145-150.
15. Nikhil Sagar Miriyala, “Event Driven System Design with High Availability”, Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol, vol. 10, no. 6, pp. 2470–2477, Dec. 2024, doi: 10.32628/CSEIT251112158.
16. Khan, M. I. (2025). Big Data Driven Cyber Threat Intelligence Framework for US Critical Infrastructure Protection. Asian Journal of Research in Computer Science, 18(12), 42-54.
17. Mahajan, N. (2025). GOVERNANCE OF CROSS-FUNCTIONAL DELIVERY IN SCALABLE MULTI-VENDOR AGILE TRANSFORMATIONS. International Journal of Applied Mathematics, 38(2s), 156-167.
18. Chivukula, V. (2022). Improvement in Minimum Detectable Effects in Randomized Control Trials: Comparing User-Based and Geo-Based Randomization. International Journal of Computer Technology and Electronics Communication (IJCTEC), 5(4), 5442–5446.
19. Sudhakara Reddy Peram, Praveen Kumar Kanumarlapudi, Sridhar Reddy Kakulavaram. (2023). Cypress Performance Insights: Predicting UI Test Execution Time Using Complexity Metrics. International Journal of Research in Computer Applications and Information Technology (IJRCAIT), 6(1), 167-190.
20. Sakil, M. B. H., Hasan, M. A., Mozumder, M. S. A., Hasan, M. R., Opee, S. A., Mridha, M. F., & Aung, Z. (2025). Enhancing Medicare Fraud Detection with a CNN-Transformer-XGBoost Framework and Explainable AI. IEEE Access.
21. Natta P K. AI-Driven Decision Intelligence: Optimizing Enterprise Strategy with AI-Augmented Insights[J]. Journal of Computer Science and Technology Studies, 2025, 7(2): 146-152.
22. Vimal Raja, G. (2024). Intelligent Data Transition in Automotive Manufacturing Systems Using Machine Learning. International Journal of Multidisciplinary and Scientific Emerging Research, 12(2), 515-518.
23. Sakinala, K. (2025). Advancements in Devops: The Role of Gitops in Modern Infrastructure Management. International Journal of Information Technology and Management Information Systems, 16(1), 632-646.
24. Chukkala, R. (2025). Unified Smart Home Control: AI-Driven Hybrid Mobile Applications for Network and Entertainment Management. Journal of Computer Science and Technology Studies, 7(2), 604-611.
25. AM, A. R., Giri, J., Ahmad, N., & Badawy, A. S. (2024). Detection of Covid-19 based on convolutional neural networks using pre-processed chest X-ray images. Aip Advances, 14(3).
26. Singh, S. K. (2025). Identification of Key Opinion Leaders in Pharmaceuticals Using Network Analysis. Journal Of Multidisciplinary, 5(7), 18-26.
27. Singh, A. (2024). Enhancing Cybersecurity for Digital Twins: Challenges and Solutions. IJSAT-International Journal on Science and Technology, 15(4).
28. Karnam, A. (2021). The Architecture of Reliability: SAP Landscape Strategy, System Refreshes, and Cross-Platform Integrations. International Journal of Research and Applied Innovations, 4(5), 5833–5844. https://doi.org/10.15662/IJRAI.2021.0405005
29. Binu, C. T., Kumar, S. S., Rubini, P., & Sudhakar, K. (2024). Enhancing Cloud Security through Machine Learning-Based Threat Prevention and Monitoring: The Development and Evaluation of the PBPM Framework. https://www.researchgate.net/profile/Binu-C-T/publication/383037713_Enhancing_Cloud_Security_through_Machine_Learning-Based_Threat_Prevention_and_Monitoring_The_Development_and_Evaluation_of_the_PBPM_Framework/links/66b99cfb299c327096c1774a/Enhancing-Cloud-Security-through-Machine-Learning-Based-Threat-Prevention-and-Monitoring-The-Development-and-Evaluation-of-the-PBPM-Framework.pdf
30. Rahman, M. R., Tohfa, N. A., Arif, M. H., Zareen, S., Alim, M. A., Hossen, M. S., ... & Bhuiyan, T. (2025). Enhancing android mobile security through machine learning-based malware detection using behavioral system features.