Handling Imbalanced Financial Data Using Cost-Sensitive Deep Learning for AML

Amgoth Naveen
PG Scholar, Department of Computer Science and Engineering
Hyderabad, Telangana, India - 506381
amgothnaveennaik@gmail.com

Muntha Raju
Professor and Head, Department of CSE
Nalla Malla Reddy Engineering College
raju.cse@nmrec.edu.in

Sirisha Veluri
Guide, Department of CSE
Nalla Malla Reddy Engineering College
sirisha.cse@nmrec.edu.in

Mamatha Samson
Coordinator, Department of CSE
Nalla Malla Reddy Engineering College, Hyderabad
mamata2001@gmail.com

***Abstract*— Traditional rule-based anti-money laundering (AML) monitoring systems generate large numbers of false-positive alerts and are rigid when new laundering strategies emerge. This paper presents a hybrid AML framework that combines seven engineered rule-based risk signals with variational-autoencoder (VAE) behavioural embeddings and graph neural network (GNN) transaction-network embeddings. The framework was evaluated on 54,258 real-world SWIFT-based cross-border payment records from an East African commercial bank. After cleaning, currency conversion, log transformation, deep representation learning, feature fusion, correlation filtering, and mutual-information selection, the fused feature space was reduced from 23 to 9 informative attributes. Isolation Forest, Local Outlier Factor, and One-Class Support Vector Machine (OCSVM) were compared under a semi-supervised novelty-detection protocol. OCSVM achieved 99.63% precision within the top 5% of ranked alerts, while LOF obtained the highest ROC-AUC of 0.8459. The framework identified 536 novel anomalies not captured by rule-based heuristics and rejected 1,275 rule-generated alerts. SHAP explanations expose the influence of rule, VAE, and GNN features, while institutional validation showed approximately 1,000 transactions per second on standard Intel Core i7/16 GB hardware.**

Keywords*— *Anti-money laundering, anomaly detection, variational autoencoder, graph neural network, One-Class SVM, feature fusion, explainable artificial intelligence.

I. INTRODUCTION

Money laundering hides the origin, ownership, or movement of illegally obtained funds and remains a major threat to financial stability. The base study cites estimates of illicit financial flows equal to 2–5% of global GDP and explains that expanding digital payments, cross-border transfer channels, virtual assets, and instant-payment systems have enlarged the operational surface available to financial criminals. Financial institutions must therefore detect suspicious activity while satisfying customer-due-diligence, AML/CFT, and transaction-reporting obligations [5], [11]–[14].

Conventional AML monitoring still depends heavily on deterministic rules and fixed thresholds. Such systems are transparent and auditable, but they are costly because one threshold can raise large volumes of alerts requiring manual investigation. They also adapt slowly when criminals alter transaction timing, counterparties, routing, jurisdiction, or transfer size. The source paper identifies false-positive burden, limited adaptability, and difficulty detecting complex multi-jurisdiction schemes as central weaknesses of traditional monitoring [6], [17], [25].

Machine-learning methods offer a complementary path. Unsupervised anomaly detection can learn regular transaction behaviour without requiring a large verified set of suspicious labels, while deep representation learning can encode behavioural and relational structure that is difficult to capture through rules alone. Prior work has used Isolation Forest, LOF, One-Class SVM, autoencoders, GNNs, and hybrid ensembles for fraud or AML monitoring [1]–[4], [7]–[10], [23], [24]. Operational AML, however, requires more than a high global score: models must prioritise the small number of alerts investigators can realistically review, expose reasons behind those alerts, and execute at financial-institution scale.

The proposed framework addresses these requirements by fusing three complementary views of the same transaction: explicit rule-based indicators representing known compliance typologies, VAE embeddings capturing nonlinear behavioural variation, and GNN embeddings representing sender–beneficiary network topology. The combined representation is filtered to retain informative attributes and supplied to semi-supervised anomaly detectors trained on non-flagged behaviour. SHAP explains anomaly scores, and the trained workflow is deployed through a Django application for batch processing and ranked alert review.

The main contribution is therefore an end-to-end monitoring strategy that links domain rules, latent behaviour, network structure, boundary learning, explainability, and operational deployment. The architecture was validated on real cross-border transaction metadata and compared using alert-prioritisation and threshold-independent metrics. Particular emphasis is placed on Precision@5%, because a high-quality top alert queue directly reduces investigation effort in a resource-constrained compliance environment [18].

II. RELATED WORK

Statistical and machine-learning fraud detection has a long history, with prior studies showing that highly imbalanced financial data require evaluation measures that reflect the rarity of anomalous events [1]. Density-based LOF identifies points whose local density is lower than that of neighbouring samples [2], whereas Isolation Forest isolates anomalies through random partitioning [23], [24]. One-Class SVM estimates the support of normal behaviour and can construct a nonlinear boundary around legitimate observations, a useful property when suspicious labels are scarce [8].

Graph-based AML research is increasingly important because laundering often involves relationships rather than isolated transactions. Cardoso et al. used self-supervised graph representation learning for AML [3], while Eddin et al. explored graph-based alert optimisation [7]. More recent work investigates graph contrastive pre-training and multi-view GNNs for capturing money-flow relationships [26], [28]. These studies show that network representations can expose layering, structuring, and unusual counterparty activity that may be invisible in transaction-level threshold rules.

Deep behavioural models offer a second complementary view. Variational autoencoders, founded on latent probabilistic encoding [20], compress high-variance inputs into lower-dimensional representations that can describe nonlinear transaction behaviour. Reviews of deep learning for AML emphasise latent embeddings when verified labels are limited [10], [21]. Related fraud studies have also demonstrated feature-engineered neural ensembles and sequence modelling [9], [19], but the operational challenge remains the combination of detection quality with explainability and manageable alert volume.

Interpretability is especially important in regulated monitoring. SHAP assigns feature contributions using Shapley-value principles and has become a widely used method for explaining complex model outputs [22], [27]. The FATF has similarly stressed that adoption of new AML technologies should improve effectiveness without compromising transparency, accountability, and risk-based decision making [11]–[14]. These requirements motivate hybrid systems in which explicit compliance rules remain visible while learned representations provide additional evidence for emerging behaviour.

The base study positions its framework between purely rule-driven and purely learned systems. It preserves known indicators such as high-risk-country and structuring flags, adds VAE and GNN signals, selects a balanced subset of fused features, and then models the boundary of normality using OCSVM. The design is intended to reduce rule noise while still detecting anomalies orthogonal to existing rules. This combination addresses the source paper's stated gaps: scarcity of confirmed SAR labels, lack of comprehensive feature fusion, poor interpretability of standalone deep learning, and the need for efficient deployment.

III. MATERIALS AND METHODS

The proposed system follows the workflow shown in Fig. 1. Cross-border transaction metadata are cleaned and normalised, transformed into rule-based, behavioural, and relational representations, fused and filtered, then supplied to anomaly detectors. The strongest boundary model is explained with SHAP and integrated into a web interface supporting batch upload, alert scoring, prioritisation, and investigation.

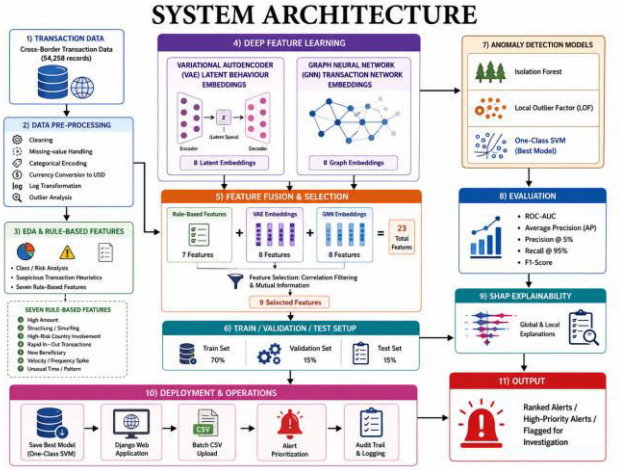

Fig. 1. Architectural Workflow of the Proposed Hybrid AML Framework

A) Dataset Collection and Pre-processing

The dataset contains 54,258 anonymised SWIFT-based cross-border payment transactions recorded over nine months in 2024 by an East African commercial bank. Direct identifiers such as account numbers were removed, while transaction reference, currency, amount, value date, anonymised originator and beneficiary names, beneficiary country, and bank identifier information were retained according to the source study's data-protection protocol. Records missing critical customer or transaction-reference fields were removed, leaving 53,783 usable observations.

Sixteen currencies were converted to USD equivalents using fixed exchange rates. Because transaction amounts were highly skewed, the amounts were log-transformed before neural representation learning. Exploratory analysis identified temporal patterns, outliers, and network characteristics that informed feature engineering. Categorical risk indicators were encoded into machine-readable form.

Table.1 Rule-Based Flag Statistics Reported in the Dataset

Rule Description	Flagged Transactions
Transactions above $10,000 threshold	15,328
High amount or significant deviation	12,491
Transaction to a high-risk country	6,742
Structuring pattern below threshold	1,859
Structuring + high-risk country	945

The individual rule counts in Table 1 overlap because a transaction may satisfy more than one criterion. The rules therefore serve as domain signals and as a proxy for known suspicious patterns rather than as mutually exclusive ground-truth classes. This distinction is important for the later semi-supervised design, which deliberately avoids training the OCSVM to reproduce the rule labels directly.

B) Rule-Based and Deep Feature Engineering

Seven engineered rule-oriented features represent temporal movement, behavioural deviation, and regulatory risk. The source defines indicators including days since the previous transaction, account tenure, log transaction amount, deviation from a customer's historical average, customer standard deviation, high-risk-country status, and a structuring flag for repeated transfers below a reporting threshold. These variables retain familiar compliance logic while creating numerical inputs for subsequent modelling.

A VAE was trained on seven numerical inputs to learn nonlinear behavioural factors. The encoder used dense layers of 32 and 16 units before producing an eight-dimensional latent representation. The decoder reconstructed the original feature space, and the model trained for 30 epochs with a reported validation loss of 2.7602. The VAE embeddings capture behavioural patterns that may not be directly represented by the handcrafted thresholds.

In parallel, a transaction graph mapped anonymised senders and beneficiaries to 31,108 nodes and 27,123 directed edges. Aggregated transaction flow statistics were used as graph inputs. The GNN embedding model generated eight network features; during 30 training epochs its loss decreased from 1.49 to 0.17. The learned components were subsequently related to transaction count and volume variability, providing a link between latent graph dimensions and AML typologies.

C) Feature Fusion and Selection

The initial fused representation concatenated 7 rule-based features, 8 VAE latent embeddings, and 8 GNN embeddings, producing 23 candidate attributes per transaction. The study then applied a three-stage filter: pairwise Pearson correlation removed features above 0.95, mutual information measured discriminative relevance, and only signals in the selected information range were retained. This reduced the 23-dimensional representation to 9 features containing a mixture of rule, VAE, and GNN evidence.

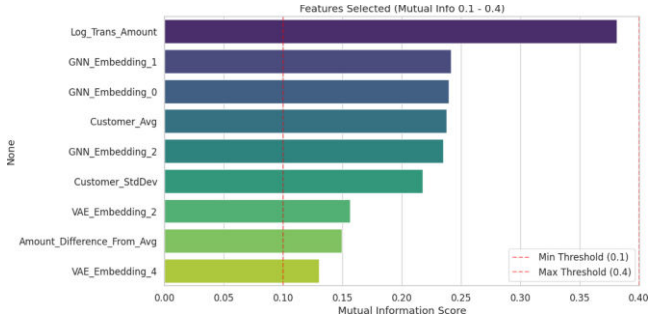

Fig. 2. Mutual Information Scores of the Final Nine Selected Features

Fig. 2 shows that Log_Trans_Amount is the strongest selected feature, followed by multiple GNN components and customer-history variables. The retained feature list includes GNN_Embedding_1, GNN_Embedding_0, Customer_Avg, GNN_Embedding_2, Customer_StdDev, VAE_Embedding_2, Amount_Difference_From_Avg, and VAE_Embedding_4. The composition supports the intended multi-view design because no single representation family occupies the entire selected space.

D) Semi-Supervised Anomaly Detection and Optimisation

Isolation Forest, Local Outlier Factor, and One-Class SVM were evaluated as anomaly detectors. The training strategy focused on non-flagged transactions so that each model learned the structure of legitimate behaviour instead of directly copying heuristic rules. This addresses the lack of verified Suspicious Activity Report labels highlighted in the source study and changes the learning objective from rule classification to novelty detection.

The OCSVM used an RBF kernel to form a nonlinear boundary around the normal transaction manifold. Grid search varied Nu over 0.001, 0.01, 0.05, 0.1, and 0.2 and Gamma over scale, auto, 0.01, and 0.1. The best validation result reported in the source used Nu = 0.01 and Gamma = 0.1, with validation AP = 0.8537 and validation ROC-AUC = 0.8517. The final test reported AP = 0.8505 and ROC-AUC = 0.8458, indicating only a small change from validation.

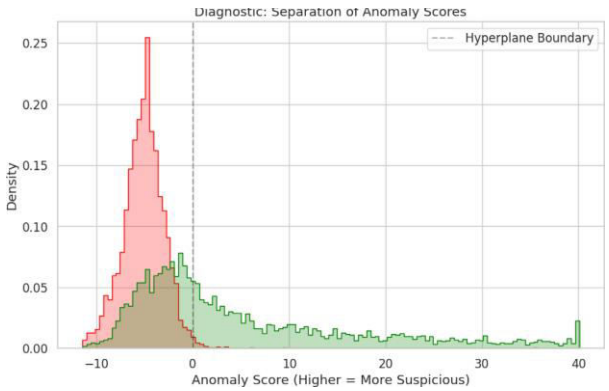

Fig. 3. Diagnostic Separation of OCSVM Anomaly Scores

The diagnostic distribution in Fig. 3 visualises the separation between the dense normal score region and higher anomaly scores. The learned hyperplane boundary is positioned near the overlap region, reflecting the OCSVM objective of enclosing the dominant normal manifold while allowing a limited fraction of observations to fall outside the boundary.

E) Evaluation, Explainability and Deployment

The evaluation emphasised ranking quality and discrimination under class imbalance. Precision@5% measures the proportion of true anomalies in the top 5% of prioritised alerts; Recall@95% measures the proportion of total risk captured above the defined review bandwidth. Average Precision summarises the precision–recall curve, ROC-AUC measures threshold-independent separation, and F1-Score balances precision and recall [4], [15].

SHAP KernelExplainer was applied to the nine-feature OCSVM input space to estimate how individual attributes influence anomaly scores. For operational use, the model was deployed through a Django web application supporting user authentication, CSV batch upload, feature transformation, ranked alerts, audit trails, and restricted compliance-user access. Testing used standard Intel Core i7 hardware with 16 GB RAM and no specialised GPU acceleration.

IV. EXPERIMENTAL RESULTS

A) Feature and Embedding Analysis

The exploratory analysis showed a heavy-tailed transaction-amount distribution and substantial rule-driven alert volume. The source reports 20,823 transactions flagged by the heuristic proxy. VAE and GNN training converged in 30 epochs and produced eight behavioural and eight relational embeddings, respectively. After fusion and selection, the nine retained features contained four rule-based features, two VAE embeddings, and three GNN embeddings, demonstrating that the selected anomaly space was not dominated by only one representation family.

The GNN variables were strongly associated with transaction-network properties. GNN_Embedding_1 correlated with transaction count, while GNN_Embedding_0 and GNN_Embedding_2 were associated with volume variability. VAE embeddings were related to customer transaction volatility. These relationships provide an interpretable mapping from latent dimensions to AML typologies and support the use of fused representation learning.

B) Performance Comparison

Table 2 reproduces the benchmark values reported for Isolation Forest, OCSVM, and LOF. OCSVM dominates the operational precision measure with Precision@5% = 0.9963, meaning that almost all transactions in the highest-priority 5% of the alert queue align with the anomaly proxy. LOF provides the highest ROC-AUC (0.8459) and the highest benchmark F1 score (0.7409), while OCSVM leads Average Precision (0.8478).

Table.2 Performance Evaluation of AML Anomaly Detection Models

Metric	Isolation Forest	One-Class SVM	LOF
ROC-AUC	0.7826	0.8388	0.8459
Avg Precision (AP)	0.6742	0.8478	0.8019
Precision @ 5%	0.8492	0.9963	0.8976
Recall @ 95%	0.1095	0.1285	0.1157
F1 Score	0.6496	0.7361	0.7409

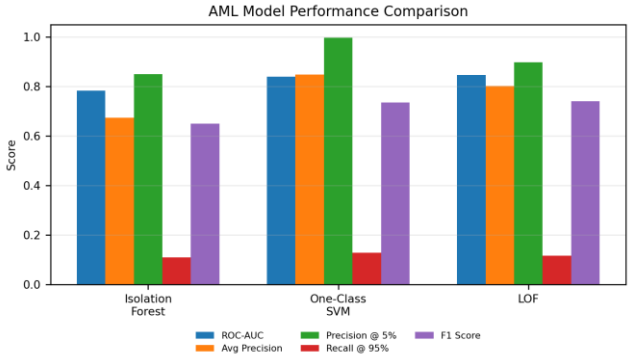

Fig. 4. Performance Comparison of AML Anomaly Detection Models

The grouped chart in Fig. 4 shows the different strengths of the three detectors. Isolation Forest trails the other methods on ranking precision and AP. LOF separates classes well on ROC-AUC but does not match OCSVM on the critical top-alert precision measure. Because financial investigators generally review only a constrained number of alerts, the source paper prioritises OCSVM for the deployed workflow even though LOF is marginally higher on ROC-AUC and F1.

C) Confusion Matrix Analysis

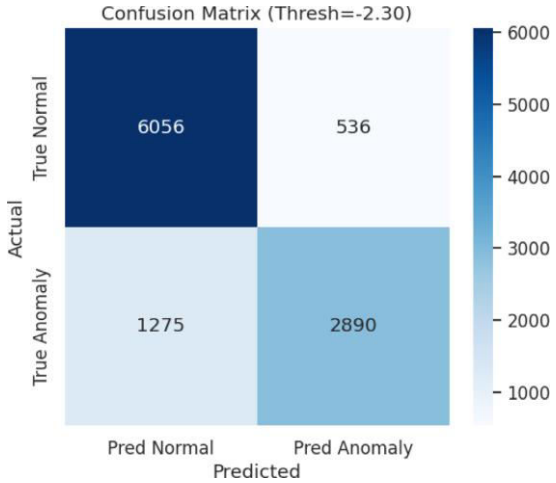

Fig. 5. Confusion Matrix of the Optimised Hybrid OCSVM

The confusion matrix in Fig. 5 compares model decisions with the heuristic proxy at the selected threshold. The model agrees with the rules on 6,056 normal transactions and 2,890 anomalous transactions. It rejects 1,275 rule-flagged transactions, interpreted by the source as reduced compliance noise, and identifies 536 additional anomalies not captured by fixed rules. At this threshold the paper reports precision of approximately 0.84, recall of 0.69, and F1 of 0.76. The result is important because the model does not simply reproduce the rule logic; it can both suppress rule alerts and surface new patterns.

D) Ablation and Robustness Analysis

The ablation study evaluates whether performance depends only on one feature family. Table 3 shows the reported leave-one-group-out results. Removing rule features causes the largest decline, reducing AP from 0.8505 to 0.7188 and ROC-AUC from 0.8458 to 0.7053. Removing the VAE embeddings has a small global effect. Removing GNN embeddings increases the listed global scores in this particular split, but the source study notes that GNN components remain locally influential for specific network-driven anomaly types.

Table.3 Leave-One-Group-Out Ablation Study

Variant	Features	AP	ROC-AUC	F1
Full Model	9	0.8505	0.8458	0.7614
w/o GNN Embeddings	6	0.8685	0.8622	0.7924
w/o VAE Embeddings	7	0.8524	0.8496	0.7601
w/o Rule Features	5	0.7188	0.7053	0.6030

Five-fold cross-validation in the base paper provides additional context: the variant without GNN embeddings reports AUC 0.8560 ± 0.0037 and AP 0.8737 ± 0.0028; without VAE embeddings, AUC 0.7973 ± 0.0071 and AP 0.8158 ± 0.0060; and without rule embeddings, AUC 0.7858 ± 0.0100 and AP 0.7711 ± 0.0082. These values indicate that handcrafted rule information is the most critical global component, while learned embeddings supply complementary behavioural and relational context rather than uniformly increasing every aggregate metric.

E) Explainability and Feature Contribution

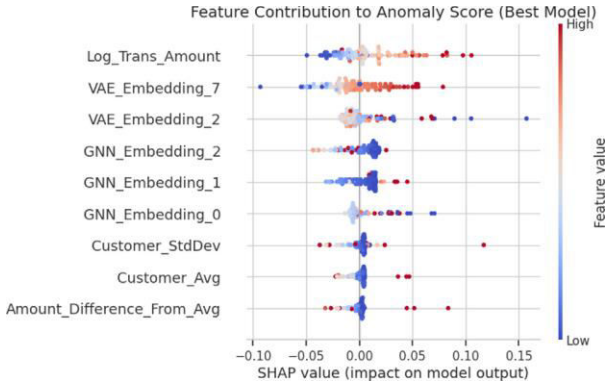

Fig. 6. SHAP Feature Contribution to the OCSVM Anomaly Score

The SHAP analysis confirms that decisions use both explicit and learned signals. Log_Trans_Amount is the strongest global contributor, while Amount_Difference_From_Avg is an important behavioural rule feature. VAE_Embedding_7 also receives substantial contribution, showing that nonlinear behaviour captured by the VAE influences anomaly ranking beyond rule thresholds. GNN embeddings add network context to individual decisions. This feature-level evidence supports the source paper's claim that simple concatenation preserves orthogonal risk signals from the three representation families.

F) Boundary Learning and Alert Prioritisation

The strongest operational finding is the OCSVM Precision@5% of 99.63%, compared with 83.99% for Isolation Forest and 89.76% for LOF. The base paper attributes this advantage to boundary learning: when legitimate transactions form a dense manifold, an RBF OCSVM can construct a tight nonlinear boundary around normality. Random partitioning in Isolation Forest is less suited to this structure, while LOF is effective at local density deviations but produces lower top-alert precision.

The study also reports that 99.2551% of top-priority transactions flagged by the model aligned with heuristic flags in a later operational analysis. This high overlap gives investigators confidence that the priority queue preserves known high-risk patterns, while the 536 newly identified anomalies demonstrate that the model is not limited to the static rule set. The simultaneous rejection of 1,275 rule alerts and discovery of additional anomalies is the core operational argument for transitioning from purely rule-based to risk-based monitoring.

G) Operational Efficiency and Deployment

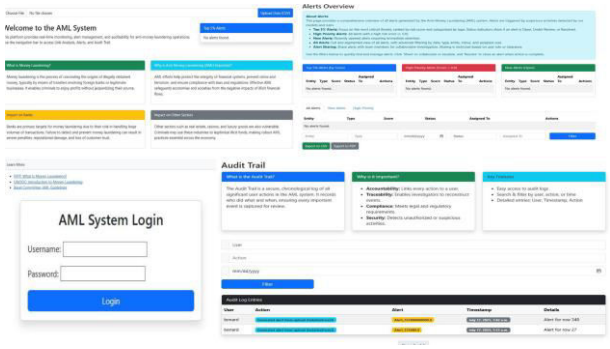

Fig. 7. Deployed AML Web Application Interface

Institutional deployment testing demonstrated processing of more than 300,000 transactions in under five minutes, corresponding to roughly 1,000 transactions per second on standard Intel Core i7/16 GB hardware. The application exposes login, a welcome page, alert overview, high-priority and new alerts, audit trails, and batch CSV upload. The back end applies the trained nine-feature model, ranks transactions, and stores the resulting alert state for authorised compliance users.

This throughput is important because it shows that the added VAE and GNN representations do not create an operational bottleneck under batch processing. The source paper positions the result as especially relevant for resource-constrained institutions that need advanced monitoring without specialised GPU infrastructure. Together with SHAP explanations, the deployment demonstrates that the framework addresses the practical dimensions of speed, alert triage, access control, and interpretability in addition to model discrimination.

H) Practical Implications, Limitations and Regulatory Relevance

The results have direct implications for alert-review design. Traditional rules remain valuable because they encode known typologies and regulatory thresholds, but the source paper demonstrates that a fixed rule set can create substantial investigation noise. The hybrid model treats those rules as one evidence stream rather than the final decision mechanism. By ranking transactions according to an anomaly score, compliance teams can focus review effort on the most suspicious cases while retaining traceable rule indicators for audit and investigation.

The confusion analysis illustrates this operational shift. The 1,275 rule-flagged transactions that the OCSVM classifies as normal are not described by the source as model failures; instead, they represent cases in which the transaction may cross a static threshold but remain consistent with the learned behavioural manifold. Conversely, the 536 transactions newly marked as anomalous represent activity that did not activate the known rules. This two-way correction is the practical value of the hybrid design: it can suppress noisy threshold alerts and surface emerging deviations at the same time.

The feature-fusion strategy is also important for model governance. The nine final inputs preserve explicit attributes such as transaction magnitude and customer-history deviation while adding latent VAE and GNN variables. SHAP then provides a common explanation space in which investigators can see whether a decision is primarily driven by an understandable regulatory signal, a behavioural embedding, a network embedding, or a combination. This structure reduces the interpretability gap associated with standalone deep models and supports the source paper's objective of transparent AI-assisted compliance.

The study connects these technical findings to the revised FATF Recommendation 16 on payment transparency. The updated direction emphasises more standardised and secure cross-border payment information and encourages financial institutions to adopt monitoring technologies capable of improving the detection of fraud and financial crime. The proposed framework aligns with that direction by combining faster batch analysis, risk-based prioritisation, and explainable decision support without discarding the auditability of established rule logic [11]–[14].

The source study nevertheless recognises an important validation limitation. Confirmed Suspicious Activity Reports are confidential and were not available as a verified ground truth, so heuristic rule flags were used as a testing proxy. This creates the possibility of circular evaluation if a model is trained directly to reproduce those labels. The semi-supervised approach mitigates that concern by fitting OCSVM on non-flagged transactions and learning the boundary of normality instead of a binary rule classifier. The discovery of 536 anomalies outside the rule set further indicates that the learned boundary is not identical to the heuristic target.

For deployment beyond the evaluated institution, the framework would benefit from validation on additional banks, currencies, customer populations, and confirmed investigation outcomes. Such validation is especially important because transaction behaviour and graph topology can shift across jurisdictions and product types. Even with this limitation, the reported throughput, feature-level explanations, and high top-alert precision demonstrate a practical pathway for institutions seeking to supplement existing AML rules with adaptive, resource-aware anomaly detection.

I) Validation Stability and Operational Thresholding

The source paper separates model selection from the final operational test so that alert-ranking quality is not judged on the same data used for tuning. For OCSVM, non-flagged observations were divided into an inner training set of 19,776 transactions and a validation set of 10,757 transactions. Grid search varied the upper-bound parameter Nu and the RBF-kernel influence parameter Gamma. Average Precision was treated as the principal optimisation criterion because the compliance objective is not merely to classify every transaction correctly, but to rank the most suspicious cases near the top of a constrained review queue. The selected setting reported validation AP of 0.8537 and validation ROC-AUC of 0.8517. On the held-out test evaluation, the source reports AP of 0.8505 and ROC-AUC of 0.8458, showing only a small reduction from validation and supporting the stability of the learned normality boundary.

This distinction between global discrimination and prioritised precision is important in AML operations. LOF records the highest benchmark ROC-AUC, which indicates useful overall separation across thresholds, whereas OCSVM provides the strongest Precision@5%. A compliance team that can investigate only a small percentage of transactions gains greater operational benefit from a model that concentrates genuine anomalies at the top of the queue. The benchmark therefore should not be interpreted as a single winner on every statistic; rather, the source selects OCSVM because its boundary-learning behaviour best matches the intended high-priority triage workflow.

The evaluation metrics can be written compactly for the ranked setting. If K denotes the reviewed top portion of the queue, precision at K is the ratio of true anomalies inside that reviewed set to K. Recall is the fraction of all proxy-positive anomalies recovered inside the permitted review bandwidth, while the F1 score is the harmonic mean of precision and recall. Average Precision summarises the precision-recall curve and is particularly informative for the imbalanced AML setting. ROC-AUC provides a threshold-independent comparison of the ordering between normal and anomalous samples. Considering these measures together prevents the high top-alert precision from being reported without the broader context of recall and global discrimination.

J) Error Interpretation and Compliance Workflow

The confusion matrix should also be interpreted as an operational comparison against the heuristic proxy rather than as a conventional fully supervised ground-truth matrix. The 6,056 normal agreements and 2,890 anomalous agreements show substantial overlap between the learned boundary and the existing rules. The 1,275 cases predicted normal despite rule flags represent the model's potential to reduce alert noise, while the 536 cases predicted anomalous despite no rule flag represent novel candidates for investigation. Because confirmed Suspicious Activity Reports were unavailable, these disagreement groups cannot automatically be labelled true system errors; they instead identify the transactions where the hybrid model and the fixed-rule baseline provide different evidence.

This disagreement is central to the proposed workflow. Rule logic supplies auditable indicators of known typologies, VAE features describe nonlinear customer behaviour, and GNN features add relational context from sender-beneficiary networks. After mutual-information selection, OCSVM maps those signals into an anomaly score and produces a ranked review list. SHAP then provides a feature-level explanation for why a transaction obtained a particular score. The deployed interface exposes the ranked alerts together with audit information, enabling a compliance analyst to move from detection to review without losing the traceability of the original rule signals.

The resulting architecture is therefore intended to augment rather than discard established AML controls. Known thresholds remain visible for regulatory audit, while the learned representations expand the evidence available for prioritisation. The source paper's institutional validation indicates that this richer decision process remains computationally practical on standard hardware. The main remaining requirement for future validation is access to independently confirmed investigation outcomes across multiple institutions, which would allow the disagreement cases to be evaluated against stronger ground truth and would provide a clearer estimate of generalisation across banking environments.

V. CONCLUSION

This paper presents a conference-format summary of the hybrid AML framework evaluated in the base study. The method combines interpretable rule features with VAE behavioural embeddings and GNN relational embeddings, then applies feature selection and semi-supervised novelty detection. Across the benchmark models, One-Class SVM provides the strongest alert-prioritisation precision, reaching 99.63% Precision@5%, while LOF attains the highest ROC-AUC of 0.8459. The confusion analysis further shows that the model can reduce rule-generated noise by rejecting 1,275 rule alerts while detecting 536 additional anomalies not captured by the heuristic baseline.

The principal advantage of the framework is the integration of known compliance logic with adaptive learned representations. SHAP explanations show that explicit transaction variables and deep embeddings both contribute to anomaly scores, which improves transparency for compliance teams. Operational validation indicates a throughput of approximately 1,000 transactions per second on standard hardware, supporting deployment in institutions without specialised accelerators.

The source study also identifies a key limitation: confirmed Suspicious Activity Reports were unavailable, so heuristic rule flags were used as an evaluation proxy. Training only on non-flagged data reduces circularity because the OCSVM learns normal behaviour instead of directly predicting the rule labels, but independent validation against confirmed suspicious cases remains important. Broader multi-institution testing can therefore strengthen evidence of generalisation while preserving the proposed risk-based and explainable monitoring design.

From a system perspective, the work demonstrates that feature fusion can be implemented without sacrificing deployment practicality. The model retains a small nine-feature inference space after selection, exposes its decisions with SHAP, and operates through a web-based batch-review workflow. This combination is useful for compliance teams because it connects model output directly to a prioritised queue instead of treating anomaly detection as an isolated offline experiment.

Overall, the findings support a transition from rigid threshold monitoring toward explainable risk-based AML surveillance. The rule layer preserves known compliance obligations, the VAE captures nonlinear customer behaviour, the GNN captures transactional relationships, and OCSVM provides a flexible normality boundary. The resulting architecture is therefore most valuable as an augmentation to existing AML controls, with future evaluation against independently confirmed cases providing the next step toward broader institutional adoption.

REFERENCES

[1] R. J. Bolton and D. J. Hand, "Statistical fraud detection: A review," Statistical Science, vol. 17, no. 3, pp. 235–255, 2002.

[2] M. M. Breunig, H. P. Kriegel, R. T. Ng, and J. Sander, "LOF: Identifying density-based local outliers," in Proc. ACM SIGMOD Int. Conf. Management of Data, 2000, pp. 93–104.

[3] M. Cardoso, P. Saleiro, and P. Bizarro, "Laundrograph: Self-supervised graph representation learning for anti-money laundering," in Proc. 3rd ACM Int. Conf. AI in Finance, 2022, pp. 130–138.

[4] R. Chalapathy and S. Chawla, "Deep learning for anomaly detection: A survey," arXiv:1901.03407, 2019.

[5] L. de Koker, "The FATF's combating of financing of proliferation standards: Private sector implementation challenges," in Financial Crime and the Law: Identifying and Mitigating Risks, Springer, 2024, pp. 123–166.

[6] D. S. Demetis, Technology and Anti-Money Laundering: A Systems Theory and Risk-Based Approach. Edward Elgar Publishing, 2010.

[7] A. N. Eddin et al., "Anti-money laundering alert optimization using machine learning with graphs," arXiv:2112.07508, 2021.

[8] S. M. Erfani, S. Rajasegarar, S. Karunasekera, and C. Leckie, "High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning," Pattern Recognition, vol. 58, pp. 121–134, 2016.

[9] E. Esenogho, I. D. Mienye, T. G. Swart, K. Aruleba, and G. Obaido, "A neural network ensemble with feature engineering for improved credit card fraud detection," IEEE Access, vol. 10, pp. 16400–16407, 2022.

[10] J. Fan et al., "Deep learning approaches for anti-money laundering on mobile transactions: Review, framework, and directions," arXiv:2503.10058, 2025.

[11] Financial Action Task Force, International Standards on Combating Money Laundering and the Financing of Terrorism and Proliferation: The FATF Recommendations. Paris, France: FATF, updated Feb. 2024.

[12] Financial Action Task Force, "Update to Recommendation 16: Payment transparency," Jun. 2025.

[13] Financial Action Task Force and Egmont Group of Financial Intelligence Units, "Digital transformation of AML/CFT for operational agencies," 2021.

[14] Financial Action Task Force, "Opportunities and challenges of new technologies for AML/CFT," 2021.

[15] D. Fourure, M. U. Javaid, N. Posocco, and S. Tihon, "Anomaly detection: How to artificially increase your F1-score with a biased evaluation protocol," in Joint European Conf. Machine Learning and Knowledge Discovery in Databases, 2021, pp. 3–18.

[16] I. Higgins et al., "Beta-VAE: Learning basic visual concepts with a constrained variational framework," in Int. Conf. Learning Representations, 2017.

[17] S. Islam, M. M. Haque, and A. N. M. R. Karim, "A rule-based machine learning model for financial fraud detection," Int. J. Electrical & Computer Engineering, vol. 14, no. 1, 2024.

[18] R. I. T. Jensen and A. Iosifidis, "Qualifying and raising anti-money laundering alarms with deep learning," Expert Systems with Applications, vol. 214, Art. no. 119037, 2023.

[19] J. Jurgovsky et al., "Sequence classification for credit-card fraud detection," Expert Systems with Applications, vol. 100, pp. 234–245, 2018.

[20] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," arXiv:1312.6114, 2013.

[21] D. V. Kute, B. Pradhan, N. Shukla, and A. Alamri, "Deep learning and explainable artificial intelligence techniques applied for detecting money laundering—A critical review," IEEE Access, vol. 9, pp. 82300–82317, 2021.

[22] M. Li, H. Sun, Y. Huang, and H. Chen, "Shapley value: From cooperative game to explainable artificial intelligence," Autonomous Intelligent Systems, vol. 4, no. 1, p. 2, 2024.

[23] F. T. Liu, K. M. Ting, and Z. H. Zhou, "Isolation forest," in Proc. 8th IEEE Int. Conf. Data Mining, 2008, pp. 413–422.

[24] F. T. Liu, K. M. Ting, and Z. H. Zhou, "Isolation-based anomaly detection," ACM Trans. Knowledge Discovery from Data, vol. 6, no. 1, pp. 1–39, 2012.

[25] M. E. Lokanan, "Predicting money laundering using machine learning and artificial neural networks algorithms in banks," Journal of Applied Security Research, vol. 19, no. 1, pp. 20–44, 2024.

[26] H. Lu and H. Wang, "Graph contrastive pre-training for anti-money laundering," Int. J. Computational Intelligence Systems, vol. 17, no. 1, p. 307, 2024.

[27] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," Advances in Neural Information Processing Systems, vol. 30, 2017.

[28] C. H. Poon, J. Kwok, C. Chow, and J. H. Choi, "LineMVGNN: Anti-money laundering with line-graph-assisted multi-view graph neural networks," AI, vol. 6, no. 4, p. 69, 2025.

[29] M. Weber et al., "Anti-money laundering in bitcoin: Experimenting with graph convolutional networks for financial forensics," arXiv:1908.02591, 2019.

[30] C. Wronka, "Financial crime in the decentralized finance ecosystem: New challenges for compliance," Journal of Financial Crime, vol. 30, no. 1, pp. 97–113, 2023.