Machine Learning with Applications 23 (2026) 100856

Contents lists available at ScienceDirect

Machine Learning with Applications

journal homepage: www.elsevier.com/locate/mlwa

Hybrid deep learning for anti-money laundering: Unsupervised detection of emerging schemes via feature fusion and explainable artificial intelligence

Cosmas Ochieng Kungu [a,b,*], Kennedy Senagi [a,b,c], Evans Omondi [a,d,*]

[a] Institute of Mathematical Sciences, Strathmore University, P.O. Box 59857-00200, Nairobi, Kenya
[b] @ilabAfrica, Strathmore University, P.O. Box 59857-00200, Nairobi, Kenya
[c] International Centre of Insect Physiology and Ecology, Duduville Campus, off Thika Road, Kasarani, P.O. Box 30772-00100, Nairobi, Kenya
[d] African Population and Health Research Center, P.O. Box 10787-00100, Nairobi, Kenya

ARTICLE INFO

Dataset link: https://github.com/cokungu1/TT4

Keywords:
Anomaly detection
Deep learning
Feature fusion
Financial crime
Prioritisation

ABSTRACT

Traditional rule-based anti-money laundering (AML) transaction monitoring systems suffer from high false-positive rates and rigidity in detecting complex emerging risk. This limitation has prompted changes to the Financial Action Task Force (FATF) recommendation 16, mandating the use of advanced systems for detecting money laundering schemes in cross-border payments. This study developed a hybrid framework integrating VAE-learned behavioural latent factors, GNN-captured relational network signals, and rule-based heuristics for enhanced anomaly detection. The model was evaluated on 54,258 real-world cross-border transaction records from an East African commercial bank. The One-Class SVM, optimised via a rigorous grid search proved superior compared to Isolation Forest and Local Outlier Factor benchmark, achieving a precision of 99.63% in the top 5% of prioritised alerts. Independent validation by a Kenyan financial institution confirms a batch processing speed of 1000 transactions per second on standard computer hardware (Intel Core i7, 16 GB RAM) and efficient high-priority alert triage, key requirements for deployment in financial institutions. Shapley additive explanations analysis further provided the interpretability of the feature contribution to the model performance. These results demonstrated that integration of rule-based features with deep-learning embeddings improves compliance work efficiency and proven pathway for resource-constrained financial institutions to comply with FATF regulatory demands upcoming in 2030.

1. Introduction

Money laundering, a complex financial crime process involving concealing the sources of illegal funds, poses a significant threat to the global stability of nations and sustainable development. The illegal activities, which arise from several offences, including bribery, have reached alarming levels in recent years. Data estimates suggest the illicit money flows represent between 2% and 5% of global GDP, amounting to $800 billion to $2 trillion annually (United Nations Office on Drugs and Crime, 2019). Developing countries alone lost approximately $6.6 trillion between 2003 and 2012 to financial crime, representing a significant drain on economic resources and erosion of trust in the world's financial systems (Kar & Spanjers, 2015).

Proliferation of digital payment systems such as instant payment systems, digital currencies, and virtual assets has complicated the fight against financial crime (Le & Zincir-Heywood, 2021; Wronka, 2023). While these new technologies increase financial inclusion, they also enable criminals to operate with increased anonymity across jurisdictions. To curb these threats, jurisdictions globally have implemented anti-money laundering (AML), counter-terrorism financing (CTF), and counter-proliferation financing (CPF) laws, requiring financial institutions and designated non-financial institutions to establish measures on customer due diligence and prompt reporting of suspicious transactions (de Koker, 2024). These efforts are aligned with international organisations such as the Financial Action Task Force (FATF) and the United Nations (UN), through SDG Goal 16, of creating peaceful and inclusive societies, promoting a robust and comprehensive framework and standards for combating financial crime, and enhancing the effectiveness of regulatory bodies (United Nations Office on Drugs and Crime, 2019). Despite these efforts, reports of financial crime have increased. Criminals are increasingly exploiting the vulnerabilities within financial systems to obscure the origins of funds, convert them to legitimate assets, and facilitate financing for terrorism and other extended proliferation activities (Unger, 2007).

Traditional AML systems, while currently widely used by industry players, rely heavily on rule-based approaches and are often rigid and prone to generating a high number of false-positive alerts, thus

* Corresponding authors.
E-mail addresses: cosmas.kungu@strathmore.edu (C.O. Kungu), ksenagi@strathmore.edu (K. Senagi), eomondi@strathmore.edu (E. Omondi).

https://doi.org/10.1016/j.mlwa.2026.100856
Received 21 October 2025; Received in revised form 28 January 2026; Accepted 29 January 2026
Available online 30 January 2026

increasing the cost of compliance for financial institutions (Demetis, 2010). Turksen et al. (2024) argued that rule-based tools are ineffective despite increasing the compliance transparency by providing the audit trails, as evident in cases like the recent grey-listing of Kenya by the Financial Action Task Force (FATF) due to weak financial crime compliance strategies. In addition, the rule-based tools have their inherent challenges. For example, Islam et al. (2024) developed a fraud detection system achieving an accuracy of 0.99 in precision compared to the traditional models, that is, the random forest, decision tree, naive Bayes, and k-nearest neighbour. However, it was noted that these systems lack the agility required for detecting complex money laundering schemes and are further recognised by the FATF, despite exhibiting high accuracy levels in certain contexts, creating the need to explore more agile and data-driven AML monitoring approaches. Financial Action Task Force (2025) June 2025 Plenary updated recommendation 16 on increased safety and security of cross-border payments to facilitate increased detection of financial crime, thus underscoring the need for advanced and agile monitoring systems like our hybrid system.

Recent evidence-based studies have shown that AML systems incorporating machine learning models such as random forest and naive Bayes classifiers have improved fraud detection. Random forest models, for example, as examined by Lokanan (2024) had shown promising results in fraud detection with a performance accuracy of 0.89 compared to the Matthews correlation coefficient (MCC) at 0.79, the gradient descent classifier (0.82), and the decision tree classifier (0.86) on the PaySim dataset of 1,048,575 records. However, Berkan Oztas and Deniz Cetinkaya and Festus Adedoyin and Marcin Budka and Gokhan Aksu and Huseyin Dogan (2024a) found that the effectiveness of these supervised models is hampered by the scarcity of labelled data and high class imbalance in financial crime data. Unsupervised methods, however, while effective with scarce labelled data, may overlook the complex transaction relationships (Segovia-Vargas et al., 2021).

Standalone deep learning models, such as variational autoencoders (VAEs) for latent features as explored by Fan et al. (2025), and graph neural networks (GNNs) for modelling complex relational financial networks explored by Poon et al. (2025) and Lu and Wang (2024), have offered great promise but often fail individually to capture complex money laundering schemes, especially multiple-jurisdiction transactions, and lack interpretability. For example, Graph Neural Networks (GNNs), as explored by Cardoso et al. (2022), on real datasets from anonymised bank sources with 66 features, the laundrograph model outperformed other models, such as multilayer perceptron (MLP) and light gradient boosting machine (GBM), achieving improved area under the curve (AUC) by 12 points and precision scores. Eddin et al. (2021) developed a model complementing rule-based systems in alert optimisation and predicting risks accurately on a real bank dataset of 400,000 accounts and found that the model reduced false positives by 80% and detected over 90% of true positive alerts. Further, Poon et al. (2025) proposed a line-graph-assisted multi-view graph neural networks (LineMVGNN) model which showed superior performance compared to other methods by effectively capturing money flow information in the transaction network.

Hybrid models combining rule-based systems with machine learning have demonstrated improved performance. For example, Talukder et al. (2024) study on a hybrid ensemble of dependable machine learning models, combining several models such as K-nearest neighbour (KNN), decision tree (DT), multilayer perceptron (MLP), and random forest (RF), achieved a 100% AUC and modest accuracy rate of 99.66%, 99.66%,99.73%, 98.56%, and 99.79% for the ENS, DT, RF, MLP, and KNN models. Esenogho et al. (2022), used a neural network ensemble model incorporating long short-term memory (LSTM) and the adaptive boosting model and reported impressive scores on the model performance with a sensitivity of 0.996 and a specificity of 0.998. Jensen and Iosifidis (2023) integrated a rule-based system with deep machine learning to qualify alerts and identify new alerts from the monitoring tool, found that their model reduced false positive rates by 33.3%, raised 75 new alerts on high-risk customers, prompting 26 new clients to be reported to the authorities by the Spar Nord Bank, while retaining over 98% of the false positive alerts from the dataset. However, these approaches often focus on specific crime typologies and lack comprehensive feature fusion. In addition, explainability, a critical component of AML systems for regulatory compliance, with techniques like SHAP (Shapley Additive Explanations) providing transparency in model decisions, often lacks in some of the models (Kute et al., 2021; Berkan Oztas and Deniz Cetinkaya and Festus Adedoyin and Marcin Budka and Gokhan Aksu and Huseyin Dogan, 2024a)

The FATF underscores that new technologies, including AML monitoring systems, enhance compliance with AML/CFT measures through enhanced speed, accuracy, and risk profiling capabilities. However, their report (Financial Action Task Force (FATF), 2021), identified critical gaps that our model addresses directly, such as explainability for regulatory compliance and the adaptability with evolving laundering schemes. We therefore operationalise the FATF's guidance on innovative methods for effectiveness in monitoring without compromising transparency, fusing a rule-based system with deep learning (VAE-GNN) embeddings while ensuring interpretability using SHAP.

While various financial crime studies have explored rule-based or standalone deep learning systems, few studies have integrated the three features into one framework. Our framework is the first to fuse rule-based features with VAE-learned latent transaction behaviours and GNN-captured relational networks, enabling detection of both known and emerging laundering patterns. Unlike prior studies by Cardoso et al. (2022) and Weber et al. (2019), Our hybrid model integrates a Semi-Supervised One-Class SVM (OCSVM) adaptability leverages SHAP for regulatory interpretability, enhancing transparency and contribution in the use of a hybrid AI system in resource-constrained financial settings.

The inherent challenges of the traditional AML systems have gained a global focus, prompting the amendment of the FATF recommendation 16, the travel rule, for better detection of financial crime, safety, and security of cross-border payments (Financial Action Task Force, 2025). The amended recommendation requires financial institutions to acquire and adopt advanced systems to protect against errors and fraud, standardise payment messages, and clarify roles within the payment ecosystem. This regulatory recommendation suggests a shift in reliance on traditional AML systems and creates a demand and adoption of advanced AML systems. The responsibility is now on the financial institutions to acquire and adopt these new systems to meet the new global expectations.

To contextualise our contributions, Table 1 provides a comparison of existing AML detection approaches, highlighting how our proposed hybrid framework uniquely integrates interpretability, unsupervised adaptability, and multi-view feature fusion.

This study presents a novel AML framework designed to advance the detection of complex laundering schemes and support compliance with updated FATF Recommendation 16. Our main innovation is the three-part feature fusion incorporating traditional rule-based features with the deep learning embeddings, mainly the VAE-learned latent embeddings and GNN transaction network embeddings. This hybrid approach enhances the system's ability to identify complex money laundering patterns that might otherwise go undetected. Building on this foundation, the study addresses the common challenge of limited labelled transaction data by employing a semi-supervised One-Class SVM (OCSVM) detection strategy. We operationalise the interpretation of deep learning by mapping the learned embeddings to specific AML typologies to enhance the explainability of latent features in financial compliance. Additionally, SHAP values were incorporated, highlighting the fairly distributed feature contribution to the model's decisions. This combination not only boosts detection accuracy but also increases the transparency of the AML system, making it more actionable for compliance teams. Finally, to bridge the gap between research and real-world application, the framework was validated by a Kenyan financial institution, which noted operational processing ability of more

Table 1
Summary of the AML detection approaches.

Capability	Traditional AML (Rule-based systems)	Standalone ML (GNN or VAE only)	Proposed hybrid framework
Detection Logic	Fixed thresholds	Latent or Graph patterns only	Fused Decision Boundary (Semi-Supervised Novelty Detection)
Rule Features	7 Heuristic Flags	None	Included (7 Heuristic Features)
VAE Embeddings	None	8 Latent Embeddings (Capturing non-linear behaviour)	Included (8 Behavioural Embeddings)
GNN Embeddings	None	8 Network Embeddings (Capturing structuring or cycles)	Included (8 Latent Topology Embeddings)
Interpretability	High (Direct rule matches)	Low ("Black box" representations)	High (SHAP values + Feature Mapping)
Adaptability	Low (Requires updates)	Medium (Requires distinct models)	High (Learns from non-flagged data manifold)

than 300,000 transactions under five minutes while maintaining the dynamic alert precision and prioritisation, thus addressing the industry challenges of scalability and compliance efficiency. Together, these contributions provide a scalable, interpretable, and practical solution for modern financial crime detection.

The remainder of this article is structured as follows: Section 2 provides a detailed description of the methodology for the hybrid AML system. Section 3 presents the results, including comparative evaluations. Section 4 discusses the results of the hybrid AML system. Finally, Section 5 concludes the work.

2. Methodology

This section details on the design of the hybrid AML framework for enhancing anti-money laundering (AML) detection through the fusion of rule-based heuristics and deep learning embeddings. Central to our methodology is the integration of variational autoencoders (VAEs) and graph neural networks (GNNs) to capture latent behavioural and structural anomalies, establishing a robust semi-supervised detection system. The overall research design and architectural workflow are illustrated in Fig. 1. The remainder of this section is organised into: data collection and preprocessing, feature engineering, the fusion strategy, semi-supervised anomaly detection modelling, and operational deployment.

2.1. Data collection and preprocessing

The data for this project was obtained from an anonymised bank within East Africa, a financial institution dealing with the processing of cross-border payments for its customers. The dataset comprised 54,258 SWIFT-based cross-border payment transaction metadata recorded over 9 months in 2024. The original dataset variables, such as the account numbers for the originator and beneficiary, were deducted, while the transaction unique reference numbers, the currencies, the transaction amounts, the transaction value date, the originator's (sender's) name (anonymised), the beneficiary's name (anonymised), the beneficiary country, and the bank identifier were retained per data protection protocols.

The data cleaning process involved the removal of missing critical records, such as the customer details and transaction reference numbers, to ensure the data is complete, accurate, and consistent with no missing values in the critical fields, resulting in 53,783. Categorical fields, such as the transaction flag to high-risk countries, were encoded. Further, the transaction amounts were converted to USD equivalents using fixed exchange rates for the 16 currencies and log-transformed to normalise the skewed distribution and to address heteroscedasticity and asymmetry, thereby improving the performance of VAEs (Zhang et al., 2019). The exploratory data analysis (EDA) was carried out to identify outliers and understand data patterns. According to Tukey et al. (1977), EDA helps with improving model accuracy, decision-making, and even understanding the overall data structure; thus, our EDA identified temporal features and outliers, informing further feature engineering.

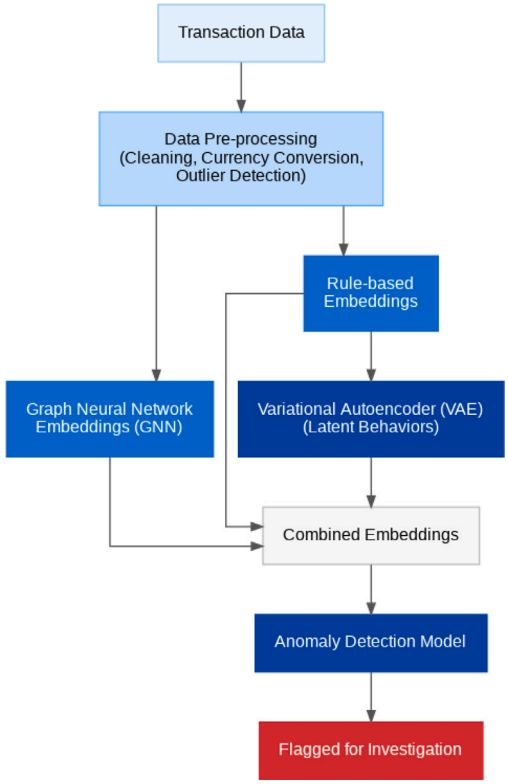

Fig. 1. Architectural workflow of the proposed Hybrid AML Framework.

2.2. Feature engineering

2.2.1. Rule-based features

Feature engineering is a critical part of developing AML systems. Raw transaction data, though abundant, is often inadequate in detecting laundering schemes. Transformation of this data is thus important for enhancing the models' ability to prevent financial crime. The seven features were engineered directly from the dataset, and they were involved in the derivation of statistical and temporal features categorised in three dimensions. The red flags were defined with a consultation with AML compliance officers, and the criteria are defined in Table 2. The red flags were used to evaluate each transaction, leading to yes

Table 2
The definition of the Temporal, Behavioural, and Risk features derived from raw SWIFT metadata.

Category	Feature variable	Scientific definition & logic
Temporal	`Days_Since_Previous`	Time delta between transaction t_i and t_{i-1} for the same ordering party: $\Delta t = (Date_i - Date_{i-1})$. *Captures rapid-fire 'smurfing' or dormancy.*
	`Days_Since_First`	Account Tenure: $(Date_{current} - Date_{min\_active})$. *Distinguishes established customers from 'burn-and-churn' mule accounts.*
Behavioural	`Log_Trans_Amount`	$\log(1 + \text{Amount}_{\text{USD}})$. *Normalises the heavy-tailed distribution of transaction values for neural network stability.*
	`Amt_Diff_From_Avg`	Absolute deviation from historical mean: $\lvert x_i - \mu_{cust} \rvert$. *Quantifies the abnormality of a transfer size relative to the customer's history.*
	`Customer_StdDev`	Standard deviation (σ) of the customer's history. *Captures volatility; high σ indicates inconsistent or high-risk pass-through behaviour.*
Risk Flags	`High_Risk_Country`	Binary flag (1) if Beneficiary Country $\in$ (OFAC List). *Deterministic regulatory flag with high predictive influence.*
	`Structuring_Flag`	Binary flag (1) if multiple transactions $< \$10k$ occur between the same Sender–Receiver pair on the same date.

or no or binary indicators (1/0) to signal the presence of suspicious transactions. Additionally, the flags provided the baseline for regulatory reporting and acted as the filtering mechanism for creating the training dataset for the default models.

- Temporal Movement: We calculated the time differences between consecutive transactions (Δt) to quantify structuring and account (in)activity.
- Behavioural Baselines: We aggregated historical customer profiles to derive mean (μ) and standard deviation (σ), which acted as baselines for flagging anomalies such as sudden account takeovers.
- Risk & Regulatory Logic: Binary flags were engineered based on the OFAC high-risk jurisdiction list and structuring typologies such as split payments below the $10,000 threshold.

It is important to note that the risk flags exert a strong influence on the classification of suspicious transactions, as they act as dominant predictors for known money laundering typologies. This aligns with findings by Lokanan (2024), who demonstrated that heuristic features such as transfer magnitude and regulatory flags consistently rank as top contributors in financial crime detection models.

2.2.2. Heuristic features for ground truth proxy

With the absence of labelled data from a verified suspicious activity report for this model, a heuristic target variable, is_rule_flag, was created. This binary feature for flagging transactions was set such that 1 is for flagging transactions meeting the predefined rule-based flags, with 0 as normal transactions, resulting in 20,823 (38.72%) transactions being flagged. While the proxy representing the known rule logic has limitations of not representing true suspicion, it served as a baseline for evaluating the model's ability to align with and potentially augment rule-based systems.

2.2.3. Variational autoencoders (VAE)

The variation autoencoders (VAEs) were developed to capture the hidden non-linear transaction behaviour representations, potentially missed by the rule-based features, addressing the scarcity of labelled transaction data for AML systems. The seven preprocessed numerical features acted as input chosen on their relevance to monitoring laundering schemes, with the final output described in Fig. 2.

The architecture followed the principles of Kingma et al. (2013) employing TensorFlow/Keras in modelling transactional behaviours. The encoder compressed the input features through dense layers of 32 units to 16 units with ReLU activations and batch normalisations to an 8-dimensional latent space via its mean and log variance. Kullback–Leibler (KL) divergence loss clipping was incorporated into the sampling layer for stability. The decoder reflected the encoder by reconstructing $\hat{x}$, through processing 7 input features from the latent samples (z) (achieving a validation loss of 2.7602) for the 30 epochs indicating that the 8-dimensional latent embeddings successfully encoded the salient variance of the input data. The VAE training involved a combination of the loss function with the reconstruction error and KL divergence followed the study by Higgins et al. (2017) for balancing the regularisation and collapse of the latent space with the loss being managed within the sampling layer.

Table 3
Correlation analysis of GNN_embeddings.

Descriptions	Average amount	Transaction count	Volume variability
GNN_Embedding_0	0.87	0.00	0.88
GNN_Embedding_1	0.03	1.00	0.13
GNN_Embedding_2	0.86	0.10	0.89
GNN_Embedding_3	−0.40	−0.83	−0.59
GNN_Embedding_4	0.77	0.50	0.78
GNN_Embedding_5	0.71	0.64	0.64
GNN_Embedding_6	0.65	−0.54	−0.05
GNN_Embedding_7	−0.54	0.21	0.18

2.2.4. Graph neural networks (GNNs)

The graph neural networks (GNNs), inspired by the convolutional neural networks, use convolutional procedures for input from the neighbours (Krizhevsky et al., 2017; Wang et al., 2020). The GNN captures the transactional network for relational anomaly detection, such as layered or circular transactions as explored by Cardoso et al. (2022), thus being important in modelling the transaction network structure between remitter and beneficiaries. The directed graph, utilising NetworkX, was used to analyse the fund flows of the dataset. The graph mapped the sender (Ordinary (sender) party name) and the beneficiary (Beneficiary (receiver) party name) to the node of a unique 31,108 transaction party. 27,123 unique directed edges were derived following aggregation of the transaction flows weighted by the USD equivalent amount and the transaction flow frequency. The initial GNN features created from the aggregated statistical properties were normalised, and the simple feed-forward neural network was utilised to form the embedding model in PyTorch. The GNN was trained for 30 epochs to minimise the Mean Squared Error (MSE) between the predicted and target node properties. The training loss decreased significantly from an initial 1.49 to 0.17, deriving a 8-dimensional GNN embeddings with ReLU activations as displayed in Table 3.

2.3. Feature fusion and selection strategy

We employed a multi-view fusion strategy in constructing a robust feature fusion space by concatenating the three distinct feature set: the 7 engineered rule flags, the 8 VAE latent embeddings, and the 8 GNN embeddings. This resulted in a unified 23 feature vector

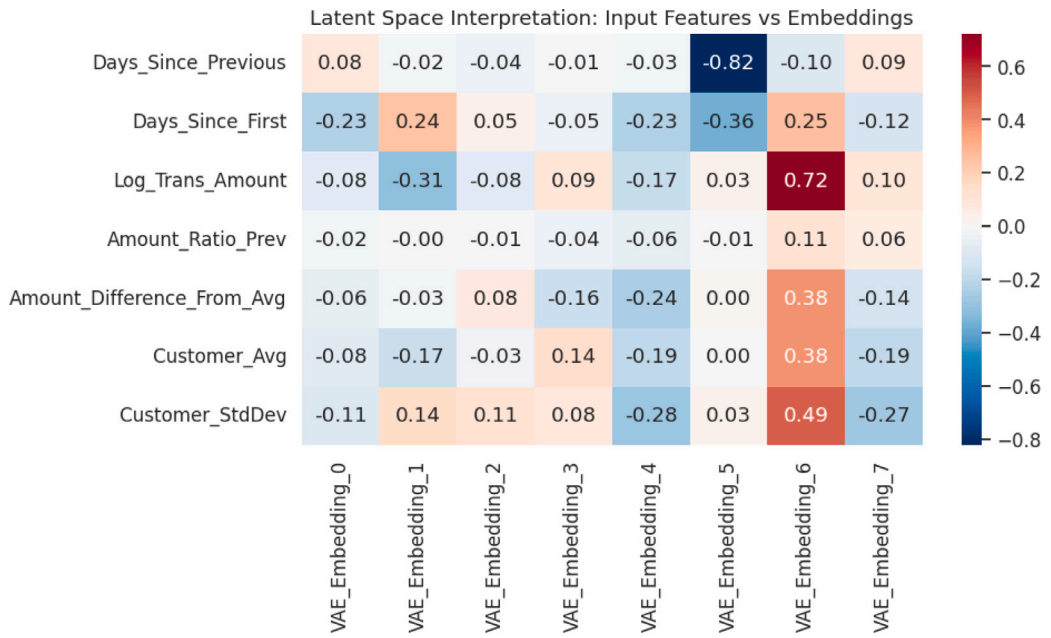

Fig. 2. Correlation heatmap between VAE latent embeddings and input features.

for each transaction, in line with the established data fusion strategies (Li et al., 2022). A Simple concatenation was preferred based on these factors. First, One-Class SVM kernel function learns the optimal non-linear relationship between features, eliminating manual weight assignment (Schölkopf et al., 2001). Standardisation of all input features prior to fusion led to elimination of scale dominance; and lastly, preserving the distinct feature columns ensure independence, critical for the SHAP explainability analysis in isolating the specific contributions of each feature (Lundberg & Lee, 2017).

However, the process of increased number of features led to a high dimensional space, thus the risk of 'curse of dimensionality' where irrelevant features may dilute the detection abilities of the model (Beyer et al., 1999; Zimek et al., 2012). We therefore, implemented a three-step filtration strategy to prevent overfitting of the models during training and ensure selection of priority risk signals, as highlighted below:

- Redundant features elimination: We employed a pairwise pearson correlation matrix and removed features with coefficients > 0.95, eliminating 5 features.
- Information gain assessment: The mutual information (MI) was calculated between each feature and the heuristic target to determine each feature's discriminative power.
- Latent signal isolation: We applied a strict filter ($0.1 \leq \text{MI} \leq 0.4$) to discard weak signals (< 0.1) and remove features that were highly proxies for the rules (> 0.4).

This design led to a reduction of the initial 23 high-dimension space to a set of 9 features (Fig. 3), comprising a balanced mix of signals for the semi-supervised model.

2.4. Anomaly detection

We employed anomaly detection models by training the Isolation Forest (IF), Local Outlier Factor (LOF), and One-Class Support Vector Machine (One-Class SVM) on the non-flagged dataset. These models were selected due to their suitability as unsupervised anomaly detection models to leverage their capabilities in the scarcity of labelled data, high dimensionality of transaction data, and reliance on different mathematical intuitions (Schölkopf et al., 2001). The training set was constructed exclusively from 26,368 'clean' (non-flagged) transactions from 43,026 training set. This allowed the models to profile the precise manifold of legitimate customer behaviour. The models were then evaluated on a stratified set of samples ($N = 19{,}776$ of 'clean' and $N = 10{,}757$ of the test set containing a mix of both normal and anomalous transactions).

2.4.1. Local outlier factor (LOF)

Local outlier factor was employed to benchmark the density-based anomaly detection. It works by calculating the local reachable density for each point based on the k nearest neighbours and assigns the LOF score, where a score greater than 1 indicates an anomaly transaction (Breunig et al., 2000). The model was chosen because it excels at detecting local anomalies and its usefulness on transaction data with varying cluster densities, such as customer segments with different transaction behaviour or sectors. However, the model is computationally expensive on a large dataset and sensitive to k (number of neighbours), presenting challenges on scalability for financial institutions with large scale dataset.

2.4.2. Isolation forest (IF)

The Isolation forest (IF) model was chosen to overcome challenges of a lack of confirmed suspicious activity labels and the challenges of the high dimensionality of the combined set of features. This model works by isolating anomalies through random feature partitioning, which is ideal for high-dimensional financial data as demonstrated in the study by Liu et al. (2012), and its computational efficiency ($\mathcal{O}(n \log n)$) enables real-time monitoring, which is critical for AML systems. However, it suffers from axis-parallel bias and is highly sensitive to high-dimensional noise (Liu et al., 2008).

2.4.3. One-class support vector machine (One-Class SVM)

One-Class SVM's ability to learn transaction boundaries and separate normal transactions from the original transaction data by using a kernel to project the data into a high-dimensional space and maximise the space around the 'normal' transactions while minimising the flagged false positives makes it an appropriate model for this study (Jurgovsky et al., 2018). This model works by mapping the input vectors into a high-dimensional feature space using a kernel function and attempts to separate the training data from the origin with a maximum margin hyperplane. Additionally, the Radial Basis Function (RBF)

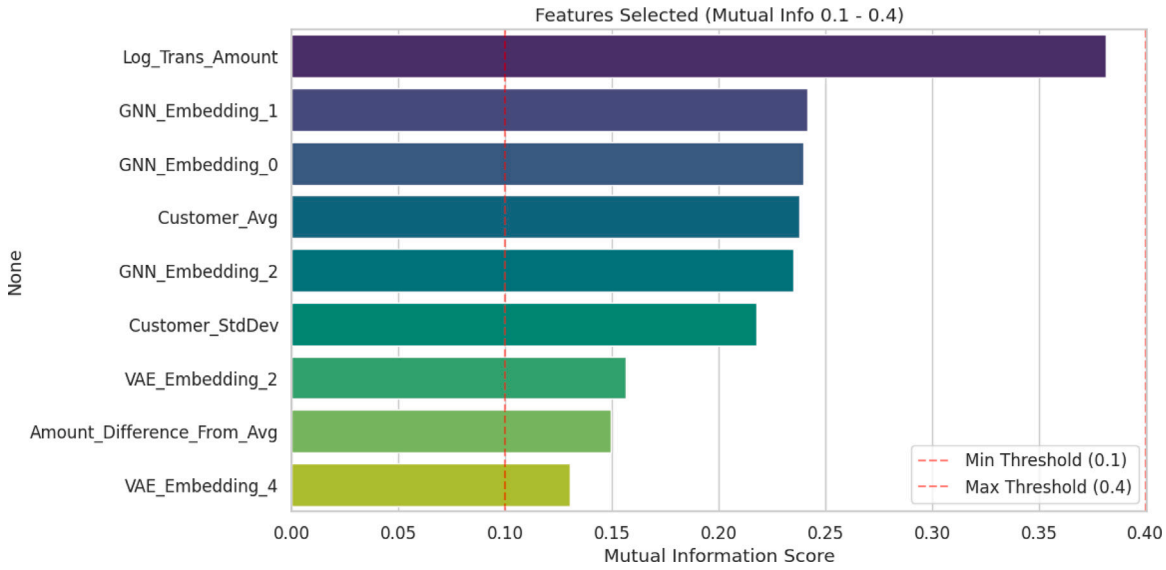

Fig. 3. Mutual information (MI) scores for the final 9 selected features.

kernel was employed to capture the non-linear interactions between the VAE and GNN embeddings. The model's sensitivity was controlled via two hyperparameters, Nu (ν) and Gamma (γ).

2.5. Hyperparameter optimisation protocol

Our best model, One-Class SVM, was hypertuned for the fused feature space by implementing a rigorous Grid Search protocol, minimising the risk of overfitting, with outcome shown in Table 9. The process involved partitioning of the non-flagged training data into an inner train set ($N = 19{,}776$) and a validation set ($N = 10{,}757$). Additionally, AP was optimised because it focuses on the quality of the top-ranked alerts, which is the operational priority. The grid search explored the following hyperparameter space:

- Nu (ν): $[0.001, 0.01, 0.05, 0.1, 0.2]$ – Controlling the upper bound of training errors.
- Gamma (γ): $[\text{scale}, \text{auto}, 0.01, 0.1]$ – Controlling the kernel influence radius.

2.6. Innovations in the AML detection model

While traditional unsupervised models like Isolation Forest are widely used, our framework introduces a significant innovation by adopting a semi-supervised novelty detection strategy utilising the One-Class SVM (OCSVM) trained on a non-flagged dataset. Furthermore, our approach is the fusion of VAE-learned latent embeddings and GNN-captured network topologies, strengthening detection of non-linear laundering schemes that evade standard rule-based features, aligning directly with updated FATF recommendation 16 (Financial Action Task Force, 2025; Financial Action Task Force and Egmont Group of Financial Intelligence Units, 2021). As advocated by Financial Action Task Force (2024), a risk-based approach, our approach provides financial institutions with an optimised and robust monitoring solution, enabling efficiency and intelligent response to evolving financial crime as well as meeting this new regulatory requirement.

2.7. Evaluation metrics

The model's performance was evaluated based on its ability to differentiate between flagged transactions and normal transactions. The following metrics were utilised while taking into account the inherent class imbalance issue and the operational need for prioritising alerts (Bolton & Hand, 2002; Davis & Goadrich, 2006; Fawcett, 2006). The metrics adopted are based on established practises in financial anomaly detection as described below:

Prioritised rank-based metrics to quantify the operational value of the system in a real-world environment:

- Precision @ Top 5% ($P@K$): This metric measures investigative efficiency by calculating the proportion of true anomalies found within the top 5% of the prioritised alert queue, aligning with industry requirements for high-priority alert triage for compliance professionals in a resource-constrained environment (Jensen & Iosifidis, 2023).
- Recall @ 95% ($R@Q$): This metric measures the proportion of the total money laundering risk captured above the defined bandwidth, reflecting the model's ability to detect anomalies within the top-ranked alerts and manageable alert review bandwidth (Yacouby & Axman, 2020).

Additionally, we employed the below metrics to help assess the model's overall discriminatory power on specific thresholds:

- Average precision (AP): This metric is important for the high class imbalance common in financial datasets. AP provides a robust assessment by summarising the precision–recall curve in addition to providing a meaningful summary in highly skewed tasks. It focuses on the model's ability to rank anomalies correctly higher than normal transactions (Saito & Rehmsmeier, 2015; Su et al., 2013).
- Area Under the ROC Curve (AUC-ROC): We employed AUC-ROC to evaluate the general separation capabilities of the model, providing a baseline comparison against standard literature (Chalapathy & Chawla, 2019).
- F1-Score: The harmonic mean of precision and recall was calculated to measure the model's balance between detection sensitivity and false alarm rates, allowing optimal decisions to justify the balance made during deployment (Cao et al., 2025; Fourure et al., 2021).

2.7.1. Model explainability and deployment

SHAP (Shapley Additive exPlanations) was employed to provide insights into the model's scoring, enhancing the interpretability and transparency of the model. SHAP integrates several model feature attributions based on Shapley value and is the most widely used because

Table 4
Provides a summary of the flagged transaction for each rule.

Rule description	Flagged transactions
Transactions above $10,000 threshold	15,328
High amount ($10,000) or significant deviation	12,491
Transaction to a high-risk country as per the FATF	6,742
Structuring pattern detected (below threshold)	1,859
Structuring pattern and high-risk country	945

it satisfies the four axioms: efficiency, symmetry, dummy player, and linearity (Li et al., 2024; Lundberg & Lee, 2017). The Kernel-explainer acted as a background dataset, sampled from the training and test datasets, to approximate the contribution of the 9 features to the anomaly scores from the hypertuned OCSVM.

The model results for compliance professionals, system administrators, regulatory authorities, and policy makers were displayed on a web application developed using the Django framework to support batch processing, alert prioritisation, and limited access to authorised compliance personnel. The model deployment in a normal computer with specifications including an Intel Core i7 processor and 16 GB of Random Access Memory (RAM), without the use of specialised GPU acceleration. The front end includes a user-friendly, simple interface. The interface features include user authorisation, allowing for authorisation, registration, and display of the top-ranked alerts generated, new alerts generated, and total alerts. The back-end interface includes the pretrained model along with feature-relevant integrations to allow for batch processing of input data, application of the necessary transformations, and displaying of the required outcome.

3. Results

The developed AML detection framework, integrating the deep learning embeddings, the VAEs, and GNNs with the rule-based features, was evaluated on a 54,258 cross-border financial transaction dataset. We followed a sequential validation protocol as follows: exploratory data analysis, interpretation of the deep learning embeddings, the validation of the feature selection strategy, model training, optimisation of the hyperparameters for the best model, and benchmarking the final model performance.

3.1. Exploratory data analysis

The initial data analysis revealed a high level of flagged suspicious activity by rules, whereby the heuristic rules identified 20,823 transactions, 39.72% of the total transactions under review, as shown in Table 4. Additionally, transaction amounts were observed to have a heavy-tailed distribution (skewness> 10), justifying the need for the use of log-transformation, VAE embeddings for compressing the high variance space into dense non-linear latent space, and the use of GNN to explicitly map these relationships.

3.2. Deep learning embeddings

The main contribution of our study is the extraction of deep learning embeddings from raw transaction data. The VAEs were successfully trained on a subset of 7 numerical features, where the training converged at 30 epochs with the final output of 8 embeddings. Additionally, the GNNs, comprising 31,108 unique parties as the nodes and 27,123 directed edges, were constructed, converging at 30 epochs to form 8 GNN embeddings. Before feature fusion, validation of learned embeddings was performed to address the challenges of the 'black box' nature of deep learning models. The process involved analysis of the correlation between the learned latent vectors and the raw input attributes of VAE and GNN embeddings as detailed in Fig. 2 and Table 5. This process demonstrates that the models successfully encoded distinct risk indicators, thus meeting the regulatory compliance requirements.

3.3. Feature importance analysis

The fusion process ensured retention of a diverse mix of features, that is, 4 Rule-based features, 2 VAE embeddings, and 3 GNN embeddings as shown in Table 5 and Fig. 3.

3.4. Evaluation of feature relevance

The efficacy of the hybrid framework is empirically demonstrated by the feature selection results shown in Fig. 3 revealing a fairly mixed mixture of rule-based heuristics and deep learning embeddings.

Notably, GNN_Embedding_1 and GNN_Embedding_0 rank immediately after the top rule-based feature, indicating that customer transaction network provides a strong risk signal. The Fig. 3 confirms that the model leverages on all the fused signal features: Rules provide the baseline magnitude, GNNs provide the network context, and VAEs provide the behavioural nuance.

3.5. Impact of feature fusion (ablation study)

The ablation study in Table 7 and Table 8 provides evidence on the contribution of each feature group in our hybrid fusion architecture.

Interestingly, Table 7 shows that the removal of deep learning embeddings resulted in a minimal performance shift in the global metrics. However, SHAP analysis reveals that specific embeddings (e.g., GNN_Embedding_1) remain locally influential for specific anomaly types related to volume variability. This suggests that deep learning embeddings provide critical context for specific AML typologies.

3.6. Performance benchmarking

Table 6 presents the comparative results showing that the One-Class SVM (OCSVM) demonstrated superior performance across the most critical operational metrics. This indicates that when the OCSVM is highly confident in anomaly detection, it is the best choice for reducing false positives in compliance workflows. However, the result also revealed that the Local Outlier Factor (LOF) achieved a slightly higher ROC-AUC (0.8419), and the OCSVM significantly outperformed both LOF and Isolation Forest in precision-based metrics. Specifically, OCSVM achieved a Precision @ 5% of 99.63%. Our framework, therefore, indicates that nearly every transaction flagged in the top 5% of the priority queue was a true anomaly, a critical requirement in reducing false positives and optimising the cost of compliance for financial organisations (Kute et al., 2021). In contrast, Isolation Forest achieved only 83.99%, suggesting that the random partitioning approach struggles to define the tight decision boundary in this semi-supervised training set-up.

3.7. Hyperparameter optimisation

The One-Class SVM was tuned via grid search with results shown in Table 9 revealing that the model performance was highly sensitive to the ν parameter. The optimal configuration was identified as $\nu = 0.01$ and $\gamma = 0.1$, achieving a validation AP of 0.8537 and validation ROC 0.8517. The diagnostic separation plot in Fig. 4 visually confirms that One-Class SVM effectively defined the optimal decision boundary around the normal training manifold, with minimal distributional overlap.

The model demonstrated stability by achieving a final precision @ 5% of 99.26%, AP of 0.8505 and ROC-AUC of 0.8458 on test result, slightly below the validation AP score of 0.8537 and ROC of 0.8517, confirming that the learned decision boundary generalises well to unseen data.

Table 5
Description of learned embeddings in relationship to the money laundering typologies for the hypertuned OCSVM.

Latent variable	Top correlation (r)	Typology detected	Contextual description
VAE Latent Space (Capturing Behavioural Anomalies)			
VAE_embedding_2	customer_StdDev (+0.11)	Customer transaction deviation history	Captures volatility.
VAE_embedding_7	customer_StdDev (−0.27)	Customer transaction deviation history	Captures volatility.
GNN Graph Space (Capturing Transaction Network Topologies)			
GNN_embedding_1	Transaction_Count (+1.00)	Abnormal activity	Identifies accounts that distribute funds to many beneficiaries or aggregate funds. And sudden activity in an account.
GNN_embedding_2	Volume_Variability (+0.89)	Layering and/or Structuring	Identifies variations from a customer's normal historical activity.
GNN_embedding_0	Volume_Variability (+0.88)	Layering and/or Structuring	Identifies accounts that distribute funds to many beneficiaries or aggregate funds. And sudden activity in an account.

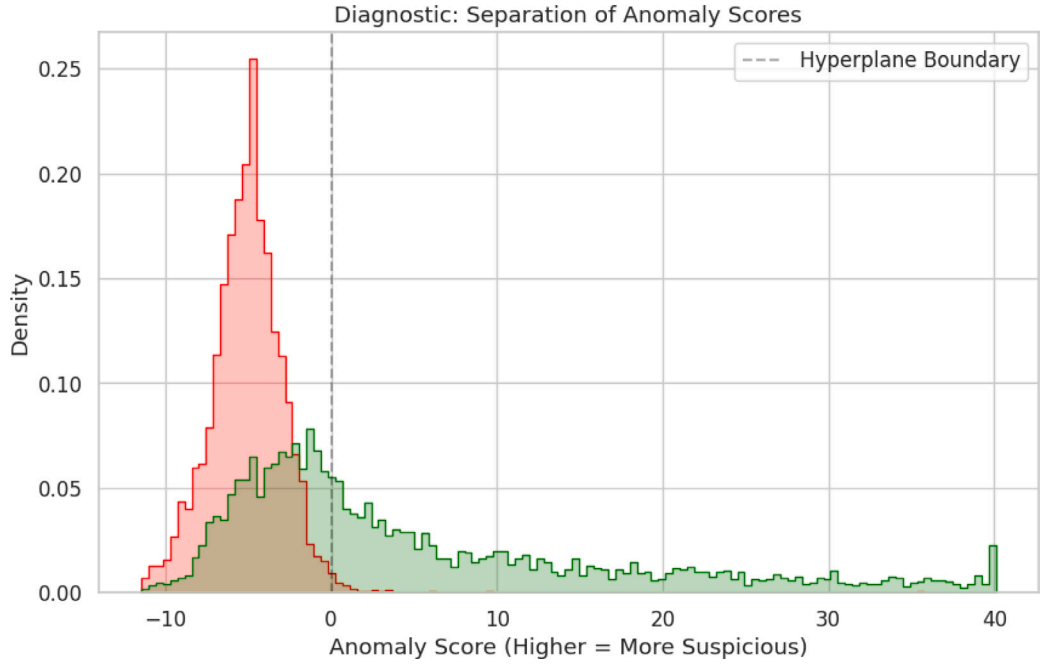

Fig. 4. Hypertuned OCSVM final test diagnostic separation plot.

Table 6
Model performance metrics.

Metrics	IsolationForest	OneClassSVM	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

Table 7
Ablation study: leave-one-group-out.

Variant	Feats	AP	ROC-AUC	F1	Impact (percentage)
Full Model	9	0.8505	0.8458	0.7614	0
GNN-Embeddings	6	0.8685	0.8622	0.7924	−2.12
VAE-Embeddings	7	0.8524	0.8496	0.7601	−0.22
Rule-Features	5	0.7188	0.7053	0.6030	−15.48

Table 8
Ablation study: 5-fold CV on optimised features.

Experiment Variant	AUC (Mean ± Std)	AP (Mean ± Std)
Full Model (Hybrid)	N/A	N/A
w/o GNN Embeddings	0.8560 ± 0.0037	0.8737 ± 0.0028
w/o VAE Embeddings	0.7973 ± 0.0071	0.8158 ± 0.0060
w/o Rule Embeddings	0.7858 ± 0.0100	0.7711 ± 0.0082

3.8. Confusion metric analysis

Fig. 5 provides the analysis showing the divergence between the hybrid OCSVM and the rule-based heuristics. The analysis shows that our model strongly aligns with the baseline rules on 2890 transactions (true-positives) and 6056 normal transactions (true negatives). However, a critical insight is that the model rejected 1275 transactions that had been flagged by the rules, thus reducing the compliance noise rather than the system detection failures. Using the dynamic filtering strategy, the OCSVM effectively addressed the issue of high false positive alerts described in Demetis (2010). Furthermore, the model flagged 536 transactions as novel anomalies that were not captured by the rules resulting in a recall of 0.69, a precision of 0.84 and an F1 score of 0.76, demonstrating that the model successfully captured the critical risk signals orthogonal to the static rules.

3.9. SHAP explanations

Fig. 6 shows the overall impact of each of the 9 features on the model's performance. The plot confirms that fused embeddings

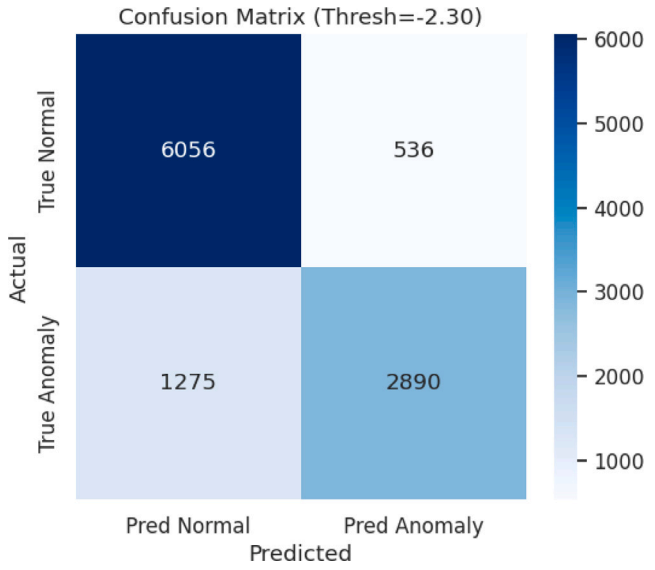

Fig. 5. Confusion Matrix with optimal threshold .2983.

Table 9
Grid search results. Lower ν values yielded higher average precision, confirming the high quality of the non-flagged training set.

Nu (ν)	Gamma (γ)	Validation AP	Val ROC-AUC
0.001	scale	0.8409	0.8286
0.001	auto	0.8280	0.8161
0.001	0.01	0.7818	0.7740
0.001	0.1	0.8268	0.8152
0.1	scale	0.8476	0.8392
0.1	auto	0.8491	0.8462
0.1	0.01	0.7039	0.7180
0.1	0.1	0.8537	0.8517
0.05	scale	0.8377	0.8407
0.05	auto	0.8411	0.8531
0.05	0.01	0.7012	0.7274
0.05	0.1	0.8410	0.8539
0.1	scale	0.8247	0.8394
0.1	auto	0.8224	0.8487
0.1	0.01	0.6932	0.7570
0.1	0.1	0.8213	0.8485
0.2	scale	0.7811	0.8246
0.2	auto	0.7541	0.8200
0.2	0.01	0.7010	0.7943
0.2	0.1	0.7503	0.8190

from VAE, GNN, and rule-based contribute significantly to the model's anomaly decisions.

3.10. Operational efficiency analysis

The framework was tested by a Kenyan financial institution to assess scalability. On the specified Intel Core i7/16 GB environment, the system processed a batch of 300,000 transactions in under 5 min, equating to a throughput of 1000 transactions per second (TPS). The deployed AML system consists of the following components, as shown in Fig. 7: the front page with details on the login page, the welcoming page with brief information on the AML, top alerts, high-priority alerts, new alerts, and all alerts. The input section involves uploading a batch transaction CSV file, and the transactions are modelled on a trained model. The model comprises 9 features from the GNN, VAE, and rule-based systems, where the file is run together with the trained model, thus allowing for uniform processing of alerts in Python and Django frameworks. After processing, the transactions are forwarded, ranked, and saved. These results highlight the dual strengths of the solution in technical robustness and compliance efficiency.

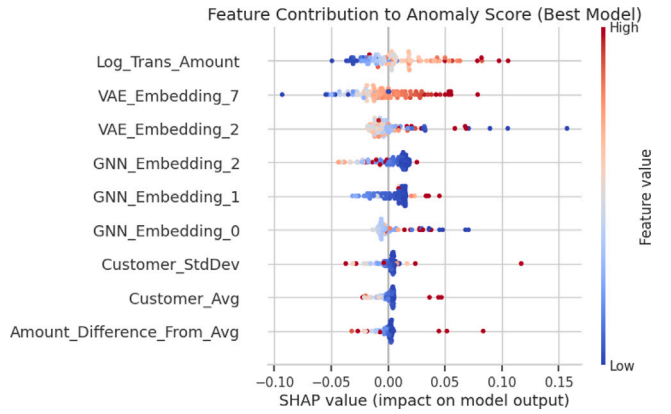

Fig. 6. Shap feature contribution analysis illustrating the most contributing features to the AML model.

4. Discussion

4.1. Addressing the limitation of ground truth validation

A critical limitation in AML research is the unavailability of confirmed Suspicious Activity Reports (SARs) due to confidentiality, necessitating the use of proxy labels during the testing phase, which introduces circular evaluation (Weber et al., 2019). Our model addresses this limitation through the adoption of the semi-supervised novelty detection strategy, where the training of the One-Class SVM involved only the non-flagged transactions to learn the manifold of legitimate behaviour rather than mimicking the heuristic rules (Erfani et al., 2016; Pimentel et al., 2014). The model's ability to transcend circularity is evidence in the 3.8: where the model identified 536 novel anomalies missed by the rules thus no 100% alignment with the rules, and rejected 1275 false-positive alerts. Additionally, the SHAP explainability analysis confirms that deep learning embeddings contribute significant decision weight orthogonal to rule-based features. This finding aligns with similar validations strategies on proxy-based evaluation by Cardoso et al. (2022), Weber et al. (2019), demonstrating that the fused deep learning embeddings contributes to the independent anomaly logic, overcoming the challenges of circularity and providing a decision boundary for evaluation rather than the binary proxy labels.

4.2. Data understanding and the need for deep learning embeddings

The EDA reveals that cross-border transfers are characterised by a highly tailed distribution (skewness> 10) and a sparse network. Similarly, the heuristic rules flagged 39.72% of the transactions as anomalies. These findings are consistent with studies from Rodin and Jelinska (2025) on the limitations of rule-based systems. The limitations of the rule-based AML system include the generation of a high number of false positive alerts, the inability to detect complex money laundering schemes, and rigidity in the face of new threats or money laundering schemes (Oztas et al., 2024b). These observations align with a study by Weber et al. (2019) on the need for deep learning approaches that are robust to reduce false positives while improving compliance and detection rates. Our findings show that the VAE and GNN embeddings adequately addressed by compressing the high-variance transaction amounts into a dense gaussian latent space, and aggregating neighbour statistics, effectively creating a dense topology embedding. These results align with the study findings of Chalapathy and Chawla (2019) that deep learning embeddings are a prerequisite for detecting anomalies that escape heuristic flags.

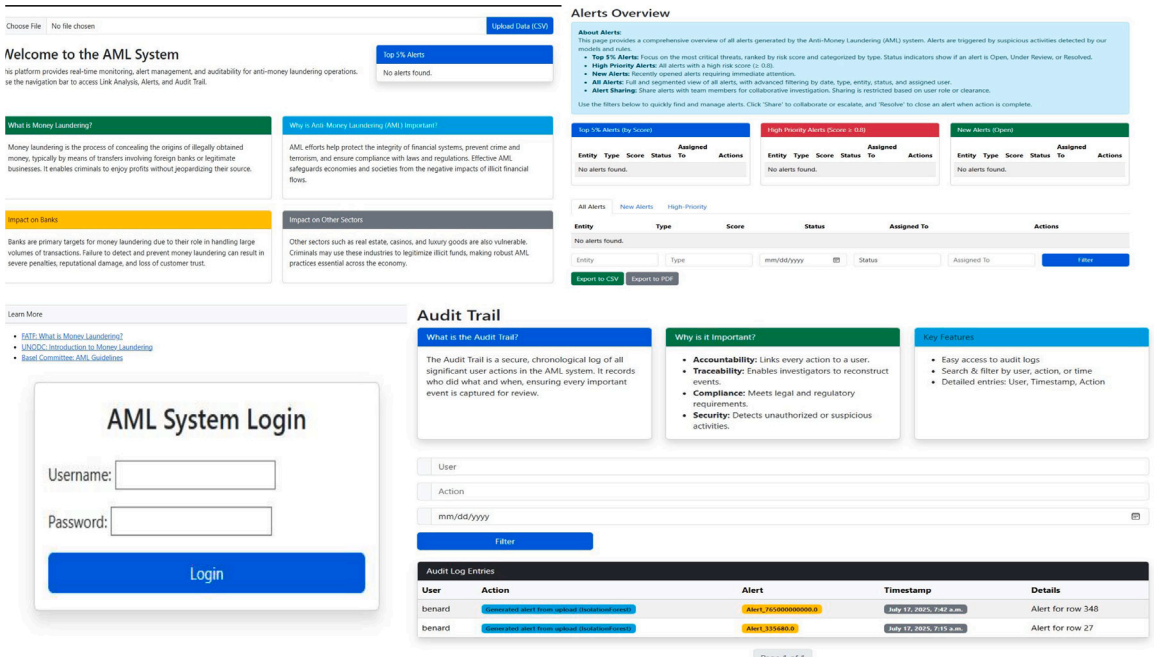

Fig. 7. Deployed AML web application interface showing details on system login, the welcoming page, audit trails, and the alerts overview interface.

4.3. The feature fusion strategy

The successful implementation of the concatenated fusion strategy detailed in Section 2.3 can be attributed to the empirical success of the model. While deep learning models require massive data for optimal outcomes, the anomaly detection models, especially the kernel-based methods, are affected by the noise in the feature space. By employing the mutual information filtering method, we ensured the final input features contained only the risk signals that provide foundation compliance through rule-based features, temporal movements from the VAE Embeddings, and structuring signals from the GNN signals.

This feature reduction prevented the curse of dimensionality from diluting the Euclidean distance metrics used by the RBF kernel, allowing the model to converge on a tighter decision boundary (Toennies, 2024). As evidenced by the SHAP analysis (Fig. 6), the model dynamically shifts reliance between these 9 views corroborating the findings by Li et al. (2024)'s argument that multi-view learning reduces uncertainty in anomaly detection. Further, the ablation study in 3.5 empirically confirms that our hybrid architecture fusion strategy achieved superior results to its individual components. This is attributable to individual features, where rule-based features provided the baseline risk signal, forming the foundation of regulatory compliance. The findings follow a study by Weber et al. (2019), who demonstrated that VAEs can significantly reduce false positives for AML systems. Additionally, Pareja et al. (2020) demonstrated that dynamic GNNs adapt well to the evolving laundering schemes, thus improving efficiency. Integration of the features from the rule-based system ensures that the known anomalies are not missed, with GNN embeddings for modelling complex transaction networks and VAE embeddings for modelling hidden customer behaviours, forming the anomaly pipeline. These fused features increased the data richness, thus improving the model performance in financial crime anomaly detection. The model's capabilities and the findings address the current FATF call for adaptive systems that are characterised by high speed, accuracy and explainability (Financial Action Task Force, 2025).

The SHAP summary plot in Fig. 6 confirms that the hybrid nature of the decision logic conf. The most influential features pushing transactions toward the anomaly threshold include:

- Log_Trans_Amount (Rule): High transaction values are the strongest global predictor of risk.
- Amount_Difference_From_Avg (Rule): Large deviations from customer history serve as a critical behavioural trigger.
- VAE_Embedding_7 (Deep Learning): Consistently ranked among the top features, this embedding captures non-linear variances in customer spending patterns that rigid rules miss. Its high SHAP value confirms that the model relies on latent deep learning signals to refine the crude thresholds provided by the rules.

Therefore a simple concatenation process adopted led to the preservation of the orthogonal risk signals from the deep learning embeddings.

4.4. The superiority of boundary learning (OCSVM)

The main contribution of this study is the superior precision @5% performance of One-Class SVM by 99.63% over Isolation Forest (83.99%) and Local Outlier Factor(89.76%) on the optimised feature space. Our framework, therefore, builds upon the work by Cardoso et al. (2022) by incorporating deep learning embeddings into the rule-based features, thus improving operational efficiency. Training the model exclusively on non-flagged data and framing the task as finding the boundary of normality. The OCSVM boundary-based approach utilising the radial basis function (RBF) kernel (Schölkopf et al., 2001), successfully constructs a non-linear hypersphere around the legitimate transactions. This confirms that for financial data where the normal manifold is dense and clustered, boundary-based methods trained on non-flagged data outperform partition-based ensembles.

In contrast, the Isolation Forest, which relies on random partitioning, struggled to define this boundary as tightly. This suggests that the latent behavioural manifold of legitimate customers is dense and clustered, favouring boundary-based methods (SVM) over partition-based methods (Trees) when non-flagged training data is available. The lower performance of the model is also attributable to the model's struggle to isolate anomalies defined by the deep learning embeddings (Liu et al., 2012).

The Local Outlier Factor (LOF) achieved the overall AUC-ROC of 0.8459. This performance can be attributed to the model density-based mechanism, allowing to identify outliers through comparison of

the density of transaction embeddings with those of k-nearest neighbours (Breunig et al., 2000). Given the mixed feature fusion leading to a high-dimensional space, a certain laundering scheme may emerge, enabling LOF to rank these deviations higher, thus the superior LOF ROC-AUC performance ranking.

4.5. Operational efficiency

The confusion matrix 5 validates the transition from rule-based compliance to risk-based compliance as mandated by the Financial Action Task Force (2025). This is because the model rejected 1275 rule-flagged alerts, a reduction in regulatory noise. The model recognised that these transactions, despite meeting the rule-based threshold, were consistent with the customer's latent behavioural history. Jensen and Iosifidis (2023) identifies this reduction in false positives as the primary driver for operational efficiency in AML. Our model also identified 536 new alerts representing emerging risks. Since the model was not trained to mimic the rules, these anomalies signify novel typologies — such as complex structuring below reporting limits — that violate the learned network topology. These findings has a direct implications following the June 2025 FATF update on recommendation 16 (Financial Action Task Force, 2025). The FATF's goal is to ensure standardisation on payment messaging for ease of detection of financial crime. These results suggest that the VAE and GNN features enriched the model, proving that this hybrid methodology is important for effective compliance with new global standard.

Beyond accuracy, the system demonstrated high efficiency. Validation logs confirmed the framework processed batches of 300,000 transactions in under 5 min, equating to a throughput of ≈ 1000 transactions per second. This aligns with FATF guidelines (Financial Action Task Force (FATF), 2021) requiring 'rapid and effective' monitoring capabilities, proving that the inclusion of complex GNN embeddings does not create an operational bottleneck.

While FATF update on Recommendation 16 aims to ensure that AML defences adapt with the changing payment landscape and supports the G20 cross-border road-map on transparent and inclusive payments (Financial Action Task Force, 2025). Our approach addresses the customer transaction behaviours by formalising the latent-space taxonomy of laundering schemes, moving our work beyond simple rule-based flags to allow detection of complex schemes, thus providing comfort to the cross-border payment system that the FATF aims to build.

Furthermore, the results show that 99.2551% of the top-priority transactions flagged by the model aligned with heuristic flags. The findings suggest that the OCSVM anomaly score can effectively flag suspicious transactions. Presenting compliance officers with prioritised alerts improves compliance efficiency and reduces costs to the financial institutions.

While the model shows superior performance, we acknowledge that the lack of confirmed Suspicious Activity Reports (SARs) necessitated the use of heuristic flags as a testing proxy. However, by adopting a semi-supervised training strategy — learning only from non-flagged data — we mitigated the risk of circular logic, ensuring the model learned to recognise normality rather than merely predicting the rules.

5. Conclusion

This study presented a hybrid AML detection framework, combining rule-based features with deep learning embeddings, specifically the VAE for latent relationships on numerical features and the GNNs for the transactional relationship network to address the recently updated FATF recommendation 16. By employing a Semi-Supervised Novelty Detection approach, we demonstrated that training on non-flagged data allows for the precise profiling of legitimate financial behaviour. We acknowledge that the confidentiality of confirmed SARs led to the use of heuristic rule-flags as a ground truth proxy for evaluation. However, by training the models on non-flagged data, we mitigated the circularity bias inherent in supervised approaches, ensuring the model learned to recognise the manifold of legitimate behaviour rather than merely predicting regulatory flags. Validated by a Kenyan financial institution, the system offers a scalable, transparent, and highly precise solution for meeting the evolving challenges of financial crime compliance.

CRediT authorship contribution statement

Cosmas Ochieng Kungu: Conceptualisation, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualisation, Writing – original draft, Writing – review & editing. **Kennedy Senagi:** Formal analysis, Investigation, Methodology, Supervision, Validation, Visualisation, Writing – original draft, Writing – review & editing. **Evans Omondi:** Formal analysis, Investigation, Methodology, Supervision, Validation, Visualisation, Writing – original draft, Writing – review & editing.

Declaration of generative AI and AI-assisted technologies in scientific writing

During the preparation of this work, the authors used [ChatGPT and Google Ai Studio] to [check the grammar and linguistic faults]. After using this tool, the authors reviewed and edited the content as needed. Therefore, they take full responsibility for the publication content.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors acknowledge the support from Strathmore University.

Data availability

The dataset and the code for the study can be accessed here: https://github.com/cokungu1/TT4.

References

Beyer, K., Goldstein, J., Ramakrishnan, R., & Shaft, U. (1999). When is "nearest neighbor" meaningful? In *International conference on database theory* (pp. 217–235). Springer.

Bolton, R. J., & Hand, D. J. (2002). Statistical fraud detection: A review. *Statistical Science*, *17*(3), 235–255.

Breunig, M. M., Kriegel, H. P., Ng, R. T., & Sander, J. (2000). LOF: identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD international conference on management of data* (pp. 93–104).

Cao, J., Zheng, W., Ge, Y., & Wang, J. (2025). DriftShield: Autonomous fraud detection via actor-critic reinforcement learning with dynamic feature reweighting. *IEEE Open Journal of the Computer Society*.

Cardoso, M., Saleiro, P., & Bizarro, P. (2022). Laundrograph: Self-supervised graph representation learning for anti-money laundering. In *Proceedings of the third ACM international conference on AI in finance* (pp. 130–138).

Chalapathy, R., & Chawla, S. (2019). Deep learning for anomaly detection: A survey. arXiv preprint arXiv:1901.03407.

Davis, J., & Goadrich, M. (2006). The relationship between Precision-Recall and ROC curves. In *Proceedings of the 23rd international conference on machine learning* (pp. 233–240).

de Koker, L. (2024). The FATF's combating of financing of proliferation standards: private sector implementation challenges. In *Financial crime and the law: identifying and mitigating risks* (pp. 123–166). Springer.

Demetis, D. S. (2010). *Technology and anti-money laundering: A systems theory and risk-based approach*. Edward Elgar Publishing.

Eddin, A. N., Bono, J., Aparício, D., Polido, D., Ascensão, J. T., Bizarro, P., & Ribeiro, P. (2021). Anti-money laundering alert optimization using machine learning with graphs. arXiv preprint arXiv:2112.07508.

Erfani, S. M., Rajasegarar, S., Karunasekera, S., & Leckie, C. (2016). High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning. *Pattern Recognition*, *58*, 121–134.

Esenogho, E., Mienye, I. D., Swart, T. G., Aruleba, K., & Obaido, G. (2022). A neural network ensemble with feature engineering for improved credit card fraud detection. *IEEE Access*, *10*, 16400–16407.

Fan, J., Shar, L. K., Zhang, R., Liu, Z., Yang, W., Niyato, D., Mao, B., & Lam, K. Y. (2025). Deep learning approaches for anti-money laundering on mobile transactions: Review, framework, and directions. arXiv preprint arXiv:2503.10058.

Fawcett, T. (2006). ROC graphs with instance-varying costs. *Pattern Recognition Letters*, *27*(8), 882–891.

Financial Action Task Force (2024). *International Standards on Combating Money Laundering and the Financing of Terrorism and Proliferation: The FATF Recommendations*. Paris, France: FATF, URL: https://www.fatf-gafi.org/en/publications/fatfrecommendations/fatf-recommendations.html. Updated February 2024. Originally published 2012.

Financial Action Task Force (2025). Update to recommendation 16: Payment transparency. URL: https://www.fatf-gafi.org/en/publications/Fatfrecommendations/update-Recommendation-16-payment-transparency-june-2025.html. (Accessed 6 August 2025).

Financial Action Task Force and Egmont Group of Financial Intelligence Units (2021). Digital transformation of AML/CFT for operational agencies. URL: https://www.fatf-gafi.org/en/publications/Fatfgeneral/Digital-transformation-aml-cft.html.

Financial Action Task Force (FATF) (2021). Opportunities and challenges of new technologies for AML/CFT. URL: https://www.fatf-gafi.org/en/publications/Digitaltransformation/Opportunities-challenges-new-technologies-for-aml-cft.html. (Accessed 15 July 2024).

Fourure, D., Javaid, M. U., Posocco, N., & Tihon, S. (2021). Anomaly detection: How to artificially increase your f1-score with a biased evaluation protocol. In *Joint European conference on machine learning and knowledge discovery in databases* (pp. 3–18). Springer.

Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M., Mohamed, S., & Lerchner, A. (2017). Beta-vae: Learning basic visual concepts with a constrained variational framework. In *International conference on learning representations*.

Islam, S., Haque, M. M., & Karim, A. N. M. R. (2024). A rule-based machine learning model for financial fraud detection. *International Journal of Electrical & Computer Engineering (2088-8708)*, *14*(1).

Jensen, R. I. T., & Iosifidis, A. (2023). Qualifying and raising anti-money laundering alarms with deep learning. *Expert Systems with Applications*, *214*, Article 119037.

Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P. E., He-Guelton, L., & Caelen, O. (2018). Sequence classification for credit-card fraud detection. *Expert Systems with Applications*, *100*, 234–245.

Kar, D., & Spanjers, J. (2015). Illicit financial flows from developing countries: 2004–2013. *Global Financial Integrity*, 1–10.

Kingma, D. P., Welling, M., et al. (2013). Auto-encoding variational bayes.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90.

Kute, D. V., Pradhan, B., Shukla, N., & Alamri, A. (2021). Deep learning and explainable artificial intelligence techniques applied for detecting money laundering–a critical review. *IEEE Access*, *9*, 82300–82317.

Le, D. C., & Zincir-Heywood, N. (2021). Exploring anomalous behaviour detection and classification for insider threat identification. *International Journal of Network Management*, *31*(4), Article e2109.

Li, M., Sun, H., Huang, Y., & Chen, H. (2024). Shapley value: from cooperative game to explainable artificial intelligence. *Autonomous Intelligent Systems*, *4*(1), 2.

Li, Q., Yan, G., & Yu, C. (2022). A novel multi-factor three-step feature selection and deep learning framework for regional GDP prediction: evidence from China. *Sustainability*, *14*(8), 4408.

Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008). Isolation forest. In *2008 eighth ieee international conference on data mining* (pp. 413–422). IEEE.

Liu, F. T., Ting, K. M., & Zhou, Z. H. (2012). Isolation-based anomaly detection. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *6*(1), 1–39.

Lokanan, M. E. (2024). Predicting money laundering using machine learning and artificial neural networks algorithms in banks. *Journal of Applied Security Research*, *19*(1), 20–44.

Lu, H., & Wang, H. (2024). Graph contrastive pre-training for anti-money laundering. *International Journal of Computational Intelligence Systems*, *17*(1), 307.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*.

Oztas, B., Cetinkaya, D., Adedoyin, F., Budka, M., Aksu, G., & Dogan, H. (2024a). Transaction monitoring in anti-money laundering: A qualitative analysis and points of view from industry. *Future Generation Computer Systems*, *159*, 161–171. http://dx.doi.org/10.1016/j.future.2024.05.027, URL: https://www.sciencedirect.com/science/article/pii/S0167739X24002607.

Oztas, B., Cetinkaya, D., Adedoyin, F., Budka, M., Aksu, G., & Dogan, H. (2024b). Transaction monitoring in anti-money laundering: A qualitative analysis and points of view from industry. *Future Generation Computer Systems*, *159*, 161–171.

Pareja, A., Domeniconi, G., Chen, J., Ma, T., Suzumura, T., Kanezashi, H., Kaler, T., Schardl, T., & Leiserson, C. (2020). Evolvegcn: Evolving graph convolutional networks for dynamic graphs. In *Proceedings of the AAAI conference on artificial intelligence*: *Vol. 34*, (04), (pp. 5363–5370).

Pimentel, M. A., Clifton, D. A., Clifton, L., & Tarassenko, L. (2014). A review of novelty detection. *Signal Processing*, *99*, 215–249.

Poon, C. H., Kwok, J., Chow, C., & Choi, J. H. (2025). LineMVGNN: Anti-money laundering with line-graph-assisted multi-view graph neural networks. *AI*, *6*(4), 69.

Rodin, I., & Jelinska, J. (2025). Enhanced anti-money laundering transaction monitoring via fuzzy equivalence in rule-based systems. In *Conference of the European society for fuzzy logic and technology* (pp. 263–274). Springer.

Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PloS One*, *10*(3), Article e0118432.

Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., & Williamson, R. C. (2001). Estimating the support of a high-dimensional distribution. *Neural Computation*, *13*(7), 1443–1471.

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

Su, W., Yuan, Y., & Zhu, M. (2013). Threshold-free evaluation of medical tests for classification and prediction: Average precision versus area under the roc curve. arXiv preprint arXiv:1310.5103.

Talukder, M. A., Hossen, R., Uddin, M. A., Uddin, M. N., & Acharjee, U. K. (2024). Securing transactions: A hybrid dependable ensemble machine learning model using iht-lr and grid search. *Cybersecurity*, *7*(1), 32.

Toennies, K. D. (2024). Decision boundaries in feature space. In *An introduction to image classification: from designed models to end-to-end learning* (pp. 109–137). Springer.

Tukey, J. W., et al. (1977). *Exploratory data analysis*: *Vol. 2*, Springer.

Turksen, U., Benson, V., & Adamyk, B. (2024). Legal implications of automated suspicious transaction monitoring: enhancing integrity of AI. *Journal of Banking Regulation*, 1–19.

Unger, B. (2007). The scale and impacts of money laundering. In *The scale and impacts of money laundering*. Edward Elgar Publishing.

United Nations Office on Drugs and Crime (2019). Crime prevention, criminal justice, the rule of law and the sustainable development goals. https://www.unodc.org/e4j/en/mun/crime-prevention-and-sdgs.html. (Accessed 26 April 2019).

Wang, Y., Lv, K., Huang, R., Song, S., Yang, L., & Huang, G. (2020). Glance and focus: a dynamic approach to reducing spatial redundancy in image classification. *33*, (pp. 2432–2444). https://Www.Fatf-Gafi.Org/En/Publications/Digitaltransformation/Digital-Transformation-Aml-Cft.Html.

Weber, M., Domeniconi, G., Chen, J., Weidele, D. K. I., Bellei, C., Robinson, T., & Leiserson, C. E. (2019). Anti-money laundering in bitcoin: Experimenting with graph convolutional networks for financial forensics. arXiv preprint arXiv:1908.02591.

Wronka, C. (2023). Financial crime in the decentralized finance ecosystem: new challenges for compliance. *Journal of Financial Crime*, *30*(1), 97–113.

Yacouby, R., & Axman, D. (2020). Probabilistic extension of precision, recall, and f1 score for more thorough evaluation of classification models. In *Proceedings of the first workshop on evaluation and comparison of NLP systems* (pp. 79–91).

Zhang, Z., Jiang, T., Zhan, C., & Yang, Y. (2019). Gaussian feature learning based on variational autoencoder for improving nonlinear process monitoring. *Journal of Process Control*, *75*, 136–155.

Zimek, A., Schubert, E., & Kriegel, H. P. (2012). A survey on unsupervised outlier detection in high-dimensional numerical data. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, *5*(5), 363–387.