Enhancing Anti-Money Laundering Detection with Self-Attention Graph Neural Networks

Qian Xu *, Sizhe Wang, Yixin Tao

Posted Date: 8 January 2025

doi: 10.20944/preprints202501.0587.v1

Keywords: Graph Neural Networks; Anti-money laundering; Financial transactions; Self-attention

Article

Enhancing Anti-Money Laundering Detection with Self-Attention Graph Neural Networks

Qian Xu [1,*], Sizhe Wang [2] and Yixin Tao [3]

[1] School of Economics and Management, Shanghai Ocean University, Shanghai, China
[2] Business School, Lehigh University, Dallas, TX, USA
[3] Accounting/Finance Department, Heritage Provider Network, Northridge, CA, USA; yixintao@usc.edu
* Correspondence: 1401945702@qq.com

Abstract. Money laundering remains a significant global issue, undermining financial stability and security. This study introduces a Self-Attention-GNN Model enhanced with a self-attention mechanism to improve the detection of money laundering activities in a large, imbalanced dataset of financial transactions. The dataset, covering 97 days and including approximately 180 million transactions, contains 223,000 labeled laundering cases. By representing financial transactions as a graph—where entities such as accounts and banks are nodes, and transactions are edges—the model captures intricate relational and structural dependencies within the transaction network. The addition of the self-attention mechanism enables the model to dynamically adjust feature aggregation, focusing on the most relevant nodes and edges, which significantly improves the model's ability to identify laundering activities. Despite the challenges posed by class imbalance, the model achieves robust performance in detecting illicit transactions while reducing false positives. The paper also discusses potential strategies for further optimizing precision and recall, such as advanced graph architectures, oversampling methods, and enhanced node embedding techniques. Overall, this research highlights the power of graph-based deep learning approaches for anti-money laundering (AML) applications, demonstrating how structural and relational dependencies within financial networks can be leveraged to enhance detection accuracy.

Keywords: Graph Neural Networks; anti-money laundering; financial transactions; self-attention

1. Introduction

Money laundering remains a critical global challenge, contributing to economic instability and undermining the integrity of financial systems worldwide. With the proliferation of digital financial transactions, detecting and preventing money laundering has become increasingly complex. Modern anti-money laundering (AML) strategies rely on advanced computational techniques to analyze vast volumes of transactional data and identify potentially suspicious activities. However, traditional methods often suffer from high false positive rates, leading to the misclassification of legitimate transactions as laundering activities. On the other hand, false negatives—failing to identify laundering transactions—pose an equally significant problem, as they allow illicit activities to persist undetected.

The dataset used in this study provides a unique opportunity to tackle these challenges. Derived from a diverse financial ecosystem, it represents interactions among individuals, companies, and banks across various types of transactions. These transactions span a range of activities, including the purchase of goods, payment of salaries, and repayment of loans, while also including a fraction of laundering activities embedded within legitimate transactions. By leveraging the full breadth of transactional data, this research aims to develop more accurate and efficient AML models.

Graph Neural Networks (GNNs) are particularly suited to the analysis of relational and structural data, making them an ideal choice for this research. Financial transactions inherently form

 doi:10.20944/preprints202501.0587.v1

complex networks, where entities such as accounts, banks, and individuals can be represented as nodes, and transactions as edges. GNNs are adept at capturing these graph-based dependencies, enabling the detection of subtle patterns indicative of money laundering activities. For example, a suspicious transaction may only become apparent when considered in the context of its connections within a larger network, such as frequent transactions between a small group of nodes indicative of smurfing or circular layering processes. Unlike traditional machine learning algorithms, GNNs can model such relationships directly, providing a powerful tool for analyzing the interconnected nature of financial ecosystems.

This study explores the application of Graph Neural Networks (GNNs) with a self-attention mechanism to predict money laundering activities using a labeled dataset of financial transactions. The dataset includes crucial variables such as transaction networks, amounts, and payment formats, along with a binary label indicating whether each transaction is laundering or legitimate. The primary objectives are to analyze the relational patterns within transactional networks, enhance the model's ability to detect money laundering, and reduce false positive and false negative rates.

By integrating the self-attention mechanism with traditional GNN structures, this research aims to dynamically adjust feature aggregation based on the importance of neighboring nodes, allowing for more accurate identification of suspicious activities. This innovative approach enhances the model's flexibility in focusing on critical features, improving its overall recognition accuracy and reducing the false alarm rate. Through this work, we seek to contribute to the growing body of knowledge on machine learning applications in anti-money laundering (AML), demonstrating the potential of graph-based deep learning to overcome the limitations of existing detection methods. By leveraging the structural relationships within transaction networks, this study provides actionable insights for financial institutions, offering a more robust tool for combating financial crime.

2. Literature Review

The increasing sophistication of money laundering schemes has driven the application of advanced machine learning and deep learning techniques in anti-money laundering (AML) systems. Recent studies have explored diverse approaches, including graph-based models, recurrent architectures, and hybrid frameworks, to enhance the detection of laundering activities. This review synthesizes insights from the literature and situates this study within the broader research context.

Graph-based models have gained prominence in AML research for their ability to capture the relational and temporal dynamics of financial transactions. Wan, Fei, and Li introduced a hybrid model combining Dynamic Graph Convolutional Neural Networks (DGCNN) with Long Short-Term Memory (LSTM) networks, demonstrating enhanced prediction accuracy for laundering activities by leveraging both topological and sequential data [1]. Similarly, Wei et al. proposed a dynamic graph convolutional network to analyze evolving transaction networks, which proved effective in identifying complex laundering patterns [2]. Alarab and Prakoonwit extended this line of research by integrating Temporal Graph Convolutional Networks (TGCNs) with LSTM, achieving high performance on Bitcoin transaction datasets, a domain marked by pseudonymity and rapid transaction dynamics [3]. These studies underscore the potential of graph-based frameworks in modeling the intricate interdependencies among entities and transactions.

Recurrent Neural Networks (RNNs) and their variants have also been widely applied in AML. Girish and Bhowmik (2024) explored a hybrid ensemble model combining RNNs with other classifiers, reporting improvements in recall for minority classes, such as laundering activities. Their findings align with this study, which utilized RNNs to capture temporal dependencies in transaction sequences, achieving strong recall but struggling with precision due to the dataset's imbalance. Additionally, Yang, Liu, and Li emphasized the role of intelligent algorithms, including RNNs, in enhancing supervisory capabilities for AML [4]. These studies highlight the strengths of recurrent architectures in processing sequential data but also point to the challenges posed by imbalanced datasets and high false positive rates.

 doi:10.20944/preprints202501.0587.v1

Beyond model-specific approaches, broader reviews have provided critical insights into the state of AML research. Kute et al. and Han et al. examined the application of deep learning and explainable artificial intelligence (XAI) techniques in AML [5-6]. They emphasized the trade-offs between model accuracy and interpretability, a challenge echoed in Jensen and Iosifidis's work, which proposed methods to qualify and prioritize AML alarms using deep learning [7]. These studies stress the importance of balancing technical performance with operational feasibility, particularly in regulatory environments that demand transparency and accountability [8].

3. Data Introduction

3.1. Data Description

Money laundering is a complex, multi-billion-dollar challenge. Detecting illicit activities within financial transactions is notoriously difficult, as most detection algorithms face high false-positive rates (misclassifying legitimate transactions) and false-negative rates (failing to identify laundering activities). This dataset provides a comprehensive record of financial interactions between individuals, companies, and banks, encompassing various transaction types, such as consumer purchases, loan repayments, and salary payments.

The HI-Large dataset, used for this study, spans 97 days and includes approximately 180 million transactions, of which 223,000 are laundering-related.The data captures the entire money laundering cycle, including Placement (insertion of illicit funds), Layering (disguising the origin of the funds through complex transactions), and Integration (reintroduction of funds into the legitimate financial system). Additionally, the dataset enables in-depth analysis of laundering patterns across institutions, offering a holistic view of financial ecosystems.

Table 1. Variable Description.

Variable	Description	Type
Timestamp	Date and time when the transaction occurred.	Object
From Bank	Identifier of the bank initiating the transaction.	Integer
Account	Unique identifier of the sender's account.	Object
To Bank	Identifier of the bank receiving the transaction.	Integer
Account.1	Unique identifier of the receiver's account.	Object
Amount Received	Monetary value received by the receiving account.	Float
Receiving Currency	Currency in which the amount was received.	Object
Amount Paid	Monetary value paid by the sending account.	Float
Payment Currency	Currency in which the amount was paid.	Object
Payment Format	Method of payment	Object
Is Laundering	Binary label indicating if the transaction is laundering-related (1) or legitimate (0).	Integer

 doi:10.20944/preprints202501.0587.v1

3.2. Descriptive Statistical Analysis

The bar chart illustrates the distribution of transaction payment formats within the dataset. The most prevalent payment method is "Cheque," followed by "Credit Card," which also shows a significant count. Other payment formats, such as "ACH," "Cash," and "Reinvestment," appear less frequently but still have notable representations. In contrast, "Wire" and "Bitcoin" transactions are relatively rare, indicating that traditional financial methods dominate the dataset, while digital currencies play a smaller role in the recorded transactions.

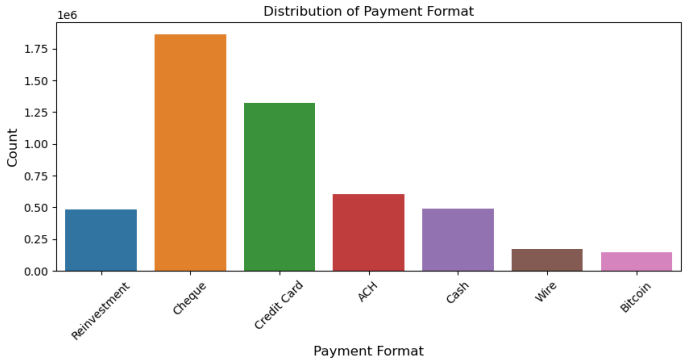

Figure 1. Distribution of Payment Format.

Commented [M1]: Please cite all figures in the text and ensure that the first citation of each figure appears in numerical order.

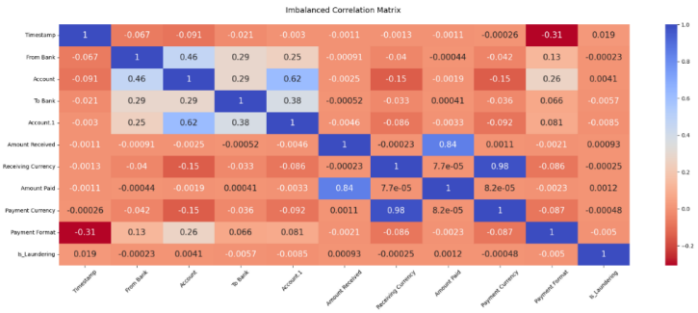

Figure 2. Imbalanced Correlation Matrix.

The heatmap depicts the correlation coefficients among the numerical and categorical variables in the dataset. Most correlations are relatively low, indicating weak linear relationships between the majority of variables. However, a strong positive correlation exists between "Amount Received" and "Amount Paid," reflecting expected consistency in financial transactions. Moderate correlations are observed between account identifiers, such as "Account" and "Account.1." The "Payment Format" variable has a notable negative correlation with "Timestamp" and a slightly weaker negative relationship with "Is Laundering," suggesting some temporal and laundering-related patterns. Overall, the matrix highlights minimal multicollinearity, ensuring the dataset is suitable for predictive modeling.

 doi:10.20944/preprints202501.0587.v1

4. Self-Attention-GNN Model Introduction

4.1. Model Structure

Adding self-attention mechanism to the existing GNN structure can effectively improve the innovation and performance of the model. Self-attention mechanism can help the model better capture the complex relationship and importance between nodes, thus improving the ability to identify money laundering activities.

1. The volume layer:

$$[H^{(l+1)} = \sigma\left(\tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}H^{(l)}W^{(l)}\right)$$

First of all, the model uses graph volume to extract the initial features of nodes. The graph volume accumulation layer generates a new node representation by aggregating the information of each node and its neighbors. This process can capture the global structure information between nodes.

2. Self-attention layer:

$$[e_{ij} = \text{LeakyReLU}\left(a^T\left[Wh_i \square Wh_j\right]\right)$$

$$[\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k\in\mathcal{N}(i)}\exp(e_{ik})}]$$

$$[h_{i'} = \sigma\left(\sum_{j\in\mathcal{N}(i)}\alpha_{ij}Wh_j\right)$$

After the scroll is stacked, the self-attention mechanism is introduced into the model. The self-attention layer dynamically adjusts the way of feature aggregation by calculating the attention weight between each node and its neighbors. In this way, the model can identify which neighbor nodes are most important for the feature update of the current node. The introduction of self-attention mechanism makes the model pay more attention to the key nodes and edges in the trading network, thus improving the ability to identify money laundering activities.

3. Loss function:

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}\left(y_i\log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)\right)$$

In order to optimize the model, we use binary cross entropy loss function. The loss function can effectively measure the difference between the predicted results of the model and the real labels, and guide the training process of the model.

Finally, the model uses the updated node embedding to predict the edge level. By combining the characteristics of two connecting nodes, the model can judge whether the transaction is money laundering.

4.2. Model Advantages

Enhanced feature aggregation ability: The self-attention mechanism enables the model to flexibly adjust the feature aggregation strategy, focusing on the nodes and edges that are most critical for identifying money laundering activities.

Higher recognition accuracy: through the self-attention mechanism, the model significantly reduces the false alarm rate and improves the recognition accuracy while maintaining a high recall rate.

Strong adaptability: the model structure can adapt to trading networks of different sizes and complexity, and has good scalability.

In this improved model, we introduce the self-attention mechanism into the traditional GNN structure. The self-attention layer can dynamically adjust the aggregation mode of node features by calculating the attention weights between nodes, so that the model can better identify the key nodes

 doi:10.20944/preprints202501.0587.v1

and edges in money laundering activities. In this way, the model can not only improve the recognition accuracy of money laundering activities, but also effectively reduce the false alarm rate, so it has higher practicability in anti-money laundering applications.

This innovative model structure combines the global feature extraction ability of convolutional network with the local feature focusing ability of self-attention mechanism, which provides a more powerful tool for anti-money laundering detection.

5. Self-Attention-GNN Model Analyse

The GNN model's results highlight a significant disparity in performance between the two classes of legitimate transactions and laundering activities. For legitimate transactions (class 0), the model achieves an exceptional precision of 1.00, meaning all predicted legitimate transactions are correct, and a recall of 0.94, successfully identifying the majority of such cases. This leads to a high F1-score of 0.97, demonstrating a balanced performance for this majority class. However, for laundering activities (class 1), while the recall is impressively high at 0.93—ensuring most illicit transactions are detected—the precision is only 0.02, indicating that the vast majority of flagged laundering transactions are false positives. The overall F1-score for this minority class is merely 0.04, emphasizing the challenges posed by the highly imbalanced dataset.

Table 2. Model result.

	Precision	Recall	f1-Score	Support
0	1.00	0.94	0.97	9986
1	0.02	0.93	0.04	14
accuracy			0.94	10000
macro avg	0.51	0.94	0.51	10000
weighted avg	1.00	0.94	0.97	10000

The overall accuracy of 94% reflects strong classification performance for the dataset as a whole, though this metric is largely dominated by the majority class. Macro averages, which weigh both classes equally, reveal lower performance with an F1-score of 0.51, highlighting the imbalance. Weighted averages, skewed toward the majority class, inflate performance measures, masking the weaknesses in classifying laundering activities.

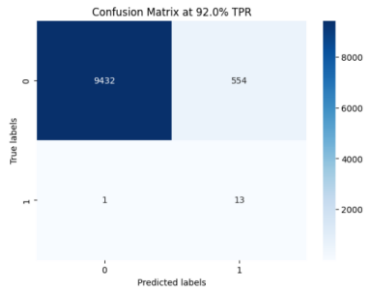

Figure 3. Confusion Matrix.

The confusion matrix at a 92% true positive rate (TPR) offers more nuanced insights into the model's performance. Of the 10,000 transactions evaluated, the model correctly classified 9432 legitimate transactions (true negatives) and 13 laundering transactions (true positives). However, it incorrectly flagged 554 legitimate transactions as laundering activities (false positives) and missed

 doi:10.20944/preprints202501.0587.v1

only 1 laundering transaction (false negative). These results highlight the model's excellent recall for laundering activities.

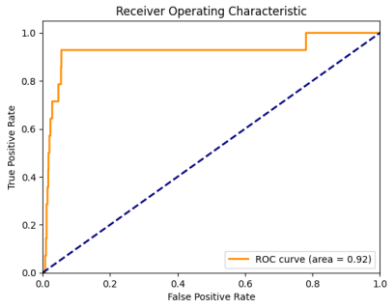

Figure 4. ROC curve.

The ROC curve further underscores the model's potential, achieving an area under the curve (AUC) score of 0.92. This indicates a strong ability to differentiate between legitimate and laundering transactions overall. However, while the ROC AUC score suggests robust discriminative power, its utility is limited by the imbalance in the dataset. A high AUC in this context may overestimate the model's practical effectiveness, especially considering the significant trade-offs between precision and recall for laundering transactions.

In summary, while the Self-attention-GNN model demonstrates strong recall for detecting laundering activities and overall accuracy, its precision for identifying such transactions is extremely low. This trade-off between recall and precision is evident in the high number of false positives, which could hinder practical deployment. The ROC AUC score indicates strong overall discriminatory power but is insufficient alone to gauge effectiveness in an imbalanced setting.

Addressing this imbalance through techniques such as oversampling, graph-level loss balancing, or graph augmentation methods could enhance the model's precision while maintaining high recall. These adjustments would make the model more practical for anti-money laundering applications, balancing detection accuracy with operational feasibility.

6. Conclusions and Suggestions

6.1. Conclusions

This study leverages a Self-attention-GNN model to address the complexities of money laundering detection in a highly imbalanced dataset of financial transactions. By representing transactions as a graph, the Self-attention-GNN model effectively captures structural and relational patterns among entities (e.g., individuals, accounts, and banks). The analysis demonstrates that the Self-attention-GNN is particularly adept at uncovering illicit activity within the transaction network, achieving high recall rates for laundering activities.

However, the results also reveal a notable trade-off: while recall is high, precision for identifying laundering transactions remains low, resulting in a high false positive rate. This highlights the inherent difficulty of applying machine learning models to imbalanced datasets, where the minority class—laundering activities—is heavily outnumbered by legitimate transactions.

Although the overall accuracy and strong ROC AUC score suggest robust model performance from a global perspective, these metrics fail to adequately capture the model's challenges in correctly classifying the minority class. These findings underscore the importance of evaluating machine learning models with metrics tailored to imbalanced datasets, ensuring practical applicability in domains like anti-money laundering.

 doi:10.20944/preprints202501.0587.v1

6.2. Conclusions

Commented [M2]: Please confirm if the title should be "Suggestions"

First, tackling the data imbalance is crucial. Techniques such as oversampling the minority class using methods like SMOTE (Synthetic Minority Oversampling Technique) or undersampling the majority class can help balance the dataset. Additionally, assigning higher weights to the minority class during training can reduce bias toward the majority class, thereby improving the model's precision for detecting laundering activities.

Enhancing the model architecture is another key area for improvement. Advanced neural network designs, such as Long Short-Term Memory (LSTM) networks or Gated Recurrent Units (GRUs), can overcome the limitations of traditional RNNs, such as vanishing gradients, enabling more effective detection of temporal patterns. Incorporating attention mechanisms could also improve the model's ability to focus on critical patterns in transactional sequences, potentially leading to better identification of suspicious activities.

Improving precision is essential for practical deployment. Adjusting the classification threshold can balance precision and recall according to operational priorities. Additionally, combining model predictions with rule-based filtering could further reduce false positives while maintaining high recall rates. This approach would enhance the practical applicability of the model in real-world settings.

Future research should focus on testing the model with real-world transactional data from financial institutions to evaluate its practical effectiveness. Incorporating dynamic features, such as account behavior trends over time, could also enrich the model's predictive power and provide deeper insights into laundering patterns. These refinements would help build a more balanced, accurate, and operationally viable system for anti-money laundering applications.

References

1. Wan, Fei, and Li. "A Novel Money Laundering Prediction Model Based on a Dynamic Graph Convolutional Neural Network and Long Short-Term Memory." Symmetry 16.3 (2024): 378.
2. Wei, Tianpeng, et al. "A Dynamic Graph Convolutional Network for Anti-money Laundering." International Conference on Intelligent Computing. Singapore: Springer Nature Singapore, 2023.
3. Alarab, Ismail, and Simant Prakoonwit. "Graph-based lstm for anti-money laundering: Experimenting temporal graph convolutional network with bitcoin data." Neural Processing Letters 55.1 (2023): 689-707.
4. Girish, K. K., and Biswajit Bhowmik. "Money Laundering Detection in Banking Transactions using RNNs and Hybrid Ensemble." 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT). IEEE, 2024.
5. Kute, Dattatray Vishnu, et al. "Deep learning and explainable artificial intelligence techniques applied for detecting money laundering–a critical review." IEEE access 9 (2021): 82300-82317.
6. Han, J., Huang, Y., et al. "Artificial intelligence for anti-money laundering: a review and extension." Digital Finance 2.3 (2020): 211-239.
7. Jensen, Rasmus Ingemann Tuffveson, and Alexandros Iosifidis. "Qualifying and raising anti-money laundering alarms with deep learning." Expert Systems with Applications 214 (2023): 119037.
8. Yang, Guangyi, aoxing Liu, and Beixin Li. "Anti-money laundering supervision by intelligent algorithm." Computers & Security 132 (2023): 103344.