Class-Weighted Graph Attention Networks for Illicit Bitcoin Transaction Detection and Forensic Screening

Lusindah Erdian Nova[1], Hilman F. Pardede[2]

1 Computer Science Faculty, Universitas Nusa Mandiri, Depok, Jawa Barat, Indonesia

2 Computer Science Faculty, Universitas Nusa Mandiri, Depok, Jawa Barat, Indonesia
1 14240010@nusamandiri.ac.id

2 hilman@nusamandiri.ac.id

Abstract

Detecting illicit cryptocurrency transactions is challenging because blockchain transaction networks are highly imbalanced and suspicious behavior may depend on relational patterns between transactions. This study investigates a class-weighted Graph Attention Network (GAT) for illicit Bitcoin transaction detection and forensic screening using the Elliptic Bitcoin Dataset. Five configurations are evaluated under the same stratified 80:20 labeled-data split: Support Vector Machine (SVM), Graph Convolutional Network (GCN), class-weighted GCN, GAT, and class-weighted GAT. While SVM achieved the highest accuracy (97.36%), precision (93.68%), and illicit-class F1-score (85.25%), class-weighted GAT achieved the highest illicit-class recall (90.43%) and ROC-AUC (0.9809), reducing false negatives to 87. The selected model was subsequently applied to 157,205 unlabeled transactions. It identified 38,105 model-positive transactions, while graph analysis identified 4,027 transactions directly connected to known illicit transactions. Combining both indicators produced a conservative priority set of 1,942 transactions for further examination. These candidates are not treated as confirmed illicit activity because ground-truth labels are unavailable. These findings show that class weighting can make GAT more sensitive to illicit transactions, although this improvement comes with more false positives. In this setting, the model is therefore more appropriate as a screening tool than as a standalone decision mechanism.

Keywords:

Bitcoin; blockchain security; cryptocurrency; graph attention network; graph neural network; illicit transaction detection; class imbalance; digital forensics.

1. Introduction

Bitcoin provides a public record of transaction activity, but the identities associated with its addresses are not directly disclosed. This makes blockchain data valuable for forensic analysis while also creating an attribution challenge. Previous studies have shown that machine-learning methods can support the detection of suspicious cryptocurrency activity when transaction volumes are too large for manual inspection [1] [2] [3] However, transaction risk is not necessarily visible

from individual attributes alone. A transaction that appears ordinary in isolation may become more informative when its relationships with other transactions are considered.

This relational characteristic makes graph-based learning well suited to Bitcoin transaction analysis. In the Elliptic dataset, transactions are represented as nodes and transaction-flow relationships as directed edges. Graph Neural Networks (GNNs) can therefore learn from both transaction features and network structure [4], [5]. The use of GNNs in security-related applications has also received increasing attention because these models can capture relational patterns while raising important privacy and security considerations [6]. Graph Convolutional Networks (GCNs) aggregate information from neighboring nodes, whereas Graph Attention Networks (GATs) learn attention coefficients that allow neighboring transactions to contribute differently to the representation of a target node [7].

Class imbalance is a second challenge. Within the labeled portion of the Elliptic dataset, Illicit transactions are far less frequent than Licit transactions. In this setting, a model may achieve high overall accuracy while still missing a meaningful number of Illicit cases. For forensic screening, these false negatives are especially important because a missed Illicit transaction may not be selected for further examination.

To address this issue, this study combines GAT with class weighting, assigning a larger training penalty to errors on the minority Illicit class. The proposed model is compared with SVM, GCN, class-weighted GCN, and unweighted GAT. Model selection is not based on accuracy alone; Illicit-class recall, false-negative count, F1-score, and ROC-AUC are considered together.

The supervised experiment is also extended to the large Unknown portion of the Elliptic graph. After model selection, predictions for Unknown transactions are combined with direct connectivity to known Illicit transactions. This step does not create new ground-truth labels. It is used only to narrow the set of transactions that may deserve closer forensic examination.

Recent research has applied graph convolution, recurrent graph learning, contrastive learning, attention mechanisms, and temporal graph models to cryptocurrency anti-money-laundering analysis [8], [9], [10], [11], [12], [13], [14], [15]. These studies demonstrate the value of relational information, but class imbalance remains an important practical issue. Strong aggregate performance does not necessarily imply that the minority Illicit class is detected reliably, particularly when false negatives are not examined explicitly.

Two practical issues are examined in this study. The first is whether class weighting helps GAT identify more transactions from the minority Illicit class, with particular attention to reducing false negatives. The second is whether model predictions can be combined with transaction-network information to narrow the large Unknown set for further examination. The model output is not interpreted as a forensic conclusion. Instead, prediction results and graph connectivity are used together as screening indicators to identify transactions that may require further examination.

Accordingly, the study is guided by three research questions: (RQ1) How does class weighting affect the detection of the minority Illicit class in GAT? (RQ2) How does class weighting change

the trade-off between false negatives and false positives? (RQ3) Can model predictions and direct connectivity to known Illicit transactions be combined to produce a smaller and defensible priority set from the Unknown class?

The contributions of this study are threefold. First, it provides a controlled comparison of conventional feature-based classification and graph-based learning by evaluating SVM, GCN, class-weighted GCN, GAT, and class-weighted GAT under the same labeled-data split. Second, it investigates class weighting as an imbalance-aware learning strategy for graph-based illicit-transaction detection, explicitly analyzing its effect on illicit-class recall and the false-negative/false-positive trade-off. Third, it extends the selected graph model to a security-oriented screening procedure for unlabeled transactions by combining model predictions with direct connectivity to known Illicit transactions. The resulting subset is interpreted as an investigation-priority set rather than as automatically assigned ground-truth labels.

2. Materials and Methods

2.1 Dataset and graph representation

Table 1. Dataset composition and experimental split.

Item	Value
Total transactions	203,769
Total edges	234,355
Licit	42,019
Illicit	4,545
Unknown	157,205
Features per transaction	165
Labeled transactions used for supervised experiment	46,564
Training nodes	37,251 (80%)
Testing nodes	9,313 (20%)
Random seed	42

The full Elliptic graph contains 203,769 transaction nodes and 234,355 directed edges. Each node is represented by 165 anonymized features, while each edge captures a directed transaction-flow relationship. The original labels are Licit, Illicit, and Unknown; only Licit and Illicit nodes are used as supervised targets. The labeled set is divided using a stratified 80:20 split to preserve the relative proportions of the Licit and Illicit classes. StandardScaler is fitted on the training features and then applied to the test features, preventing information from the test set from influencing feature scaling. For the graph models, the split indices are converted into training and test masks in PyTorch Geometric.

2.2 Class-Weighted GAT Architecture

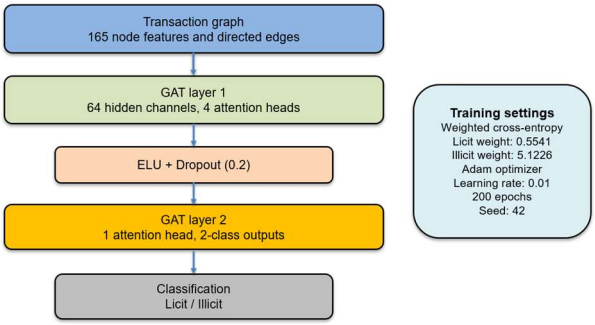


Figure 1. Architecture of the class-weighted GAT model.

The proposed model consists of two GAT layers. The first layer receives 165 transaction features and uses 64 hidden channels with four attention heads. The outputs of the four heads are concatenated to form a 256-dimensional representation. ELU activation and dropout of 0.2 are then applied. The second GAT layer uses a single attention head without concatenation and produces two output logits corresponding to the Licit and Illicit classes.

Table 2. Final configuration of GAT + Class Weight.

Parameter	Setting
Input features	165
Hidden dimension	64
Attention heads	4 → 1
Output classes	2
Activation	ELU
Dropout	0.2
Optimizer	Adam
Learning rate	0.01
Epochs	200
Loss	CrossEntropyLoss (+ Class Weight)
Licit class weight	0.5541
Illicit class weight	5.1225

Adam was used as the optimization algorithm because adaptive optimization methods are widely employed in the training of modern neural networks [16]. Class weights are calculated from the training labels only. The training set contains 33,615 Licit and 3,636 Illicit transactions. Using balanced class weights results in weights of 0.5541 for Licit and 5.1225 for Illicit. Consequently, misclassifying an Illicit transaction incurs a larger loss during optimization, encouraging the model to place greater emphasis on the minority class.

2.3 Evaluation strategy

Performance is evaluated using Accuracy, Precision, Recall, F1-score, ROC-AUC, and confusion-matrix counts. Macro Precision, Macro Recall, Macro F1, Micro Precision, Micro Recall, and

Micro F1 are also reported. Illicit-class Recall is emphasized because it measures the proportion of actual Illicit transactions detected by the model. The false-negative count is reported explicitly because, in a forensic screening context, missed Illicit transactions may not proceed to further examination. Accordingly, model selection emphasizes Illicit-class Recall and false-negative reduction while considering Precision, F1-score, and ROC-AUC to assess the associated performance trade-offs.

2.4 Embedding analysis

Learned transaction representations are projected into two dimensions using t-SNE [17]. The visualization is used for qualitative inspection of class organization in the embedding space and is not treated as evidence of formal class separability. Hidden dimensions of 16, 32, 64, and 128 are also examined for GCN and GAT during model development.

2.5 Unknown-transaction prioritization

After supervised evaluation, the selected class-weighted GAT is applied to all 157,205 Unknown transactions. The predicted class is treated only as a screening signal because these transactions do not have ground-truth labels. A second signal is obtained from graph structure by identifying Unknown transactions with at least one direct edge to a known Illicit transaction. Transactions meeting both criteria are placed in a higher-priority group for further examination.

3. Results

3.1 Main classification results

Table 3. Main performance metrics of all tested models.

Model	Acc. (%)	Prec. (%)	Recall (%)	F1 (%)	ROC-AUC
SVM	97.36	93.68	78.22	85.25	0.9733
GCN	97.14	87.60	82.40	84.92	0.9709
GCN + CW	95.20	70.09	88.67	78.29	0.9734
GAT	97.21	88.77	81.74	85.11	0.9800
GAT + CW	96.77	79.34	90.43	84.52	0.9809

SVM achieved the highest Accuracy, Precision, and F1-score, but it also produced the lowest Illicit-class Recall (78.22%). GCN increased Recall to 82.40%. Adding class weighting further increased GCN Recall to 88.67%, although Precision decreased to 70.09%. Unweighted GAT achieved a ROC-AUC of 0.9800, while GAT with class weighting achieved the highest Recall (90.43%) and ROC-AUC (0.9809). These results show that the preferred model depends on the evaluation objective rather than on a single metric.

3.2 Confusion-matrix and balanced-evaluation analysis

Table 4. Confusion-matrix counts and class-balanced evaluation metrics

Model	TN	FP	FN	TP	Macro P	Macro R	Macro F1	Micro F1
SVM	8356	48	198	711	0.9568	0.8882	0.9190	0.9736
GCN	8298	106	160	749	0.9286	0.9057	0.9167	0.9714
GCN + CW	8060	344	103	806	0.8441	0.9229	0.8780	0.9520
GAT	8310	94	166	743	0.9341	0.9031	0.9178	0.9721
GAT + CW	8190	214	87	822	0.8915	0.9394	0.9136	0.9677

The effect of class weighting is clearly visible in the false-negative count. For GCN, false negatives decreased from 160 to 103. For GAT, false negatives decreased from 166 to 87. At the same time, false positives increased, which explains the lower precision of the weighted models. GAT + Class Weight produced the highest Macro Recall of 0.9394. By contrast, SVM achieved the highest Macro F1-score of 0.9190 and Micro F1-score of 0.9736, while GAT followed closely with a Macro F1-score of 0.9178 and a Micro F1-score of 0.9721.

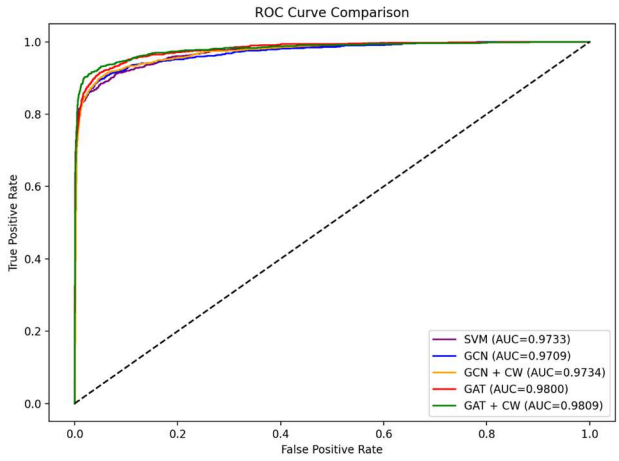


Figure 2. ROC curves for the five evaluated models.

All models achieved ROC-AUC values above 0.97. The highest value was obtained by GAT + Class Weight (0.9809), followed closely by GAT (approximately 0.9800). This result indicates that the increase in minority-class Recall obtained through class weighting was achieved while maintaining a similar level of overall discrimination ability.

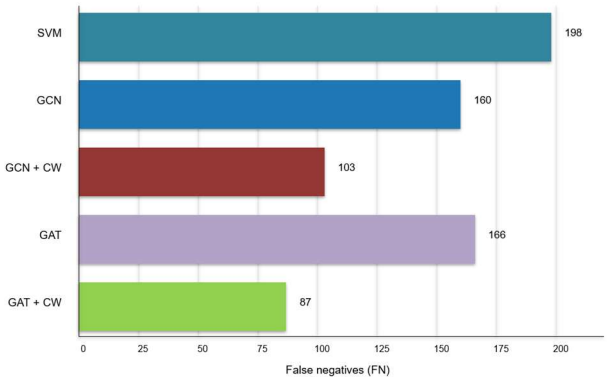


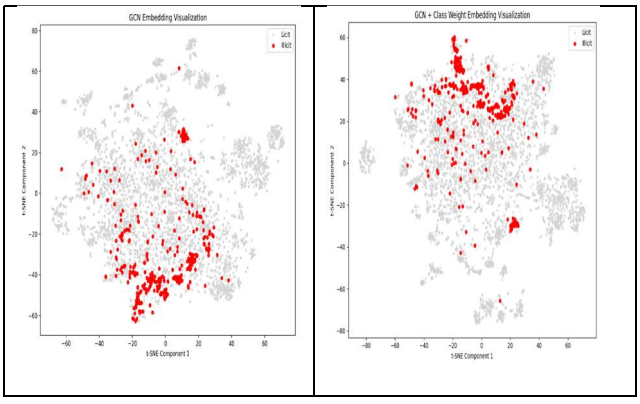
Figure 3. False-negative counts for the five evaluated models.

The weighted models detect a larger share of known Illicit transactions, but this gain is accompanied by more false positives and lower Precision. This trade-off is important for forensic screening because the preferred operating point depends on the relative cost of missed Illicit transactions and additional analyst review.

3.3 Hidden-dimension sensitivity

A sensitivity analysis was performed on the unweighted GCN and GAT models using hidden dimensions of 16, 32, 64, and 128. For GCN, F1-score increased from approximately 0.79 at 16 dimensions to 0.81 at 32 and about 0.835 at 64, with little additional gain at 128. For GAT, F1-score was approximately 0.823 at 16 dimensions, decreased slightly at 32, reached its best value of about 0.832 at 64, and fell to about 0.80 at 128. Based on this sensitivity analysis, 64 hidden dimensions were selected for the final model configuration because this setting produced the best or near-best F1-score for both architectures among the evaluated dimensions. The analysis was conducted on the base GCN and GAT architectures because class weighting changes the loss function rather than the representation size.

3.4 Embedding analysis



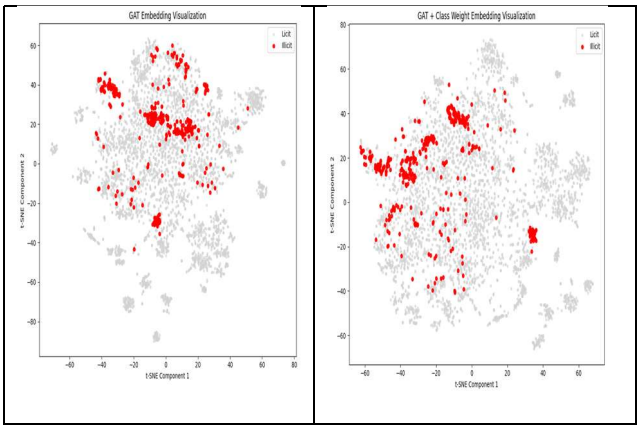


Figure 4. t-SNE projections of the learned representations.

The embedding plots provide visual support for the numerical results. In the unweighted GCN embedding, many Illicit points remain mixed with Licit points. After class weighting is applied, Illicit points form more visible groups. A similar pattern appears for GAT. The class-weighted GAT visualization shows more concentrated regions of Illicit points and fewer Illicit points within visually dense Licit regions. This pattern is consistent with the increase in recall from 81.74% to 90.43% and the decrease in false negatives from 166 to 87.

3.5 Unknown-transaction screening

Table 5. Unknown-Transaction Screening Results

Screening result	Count	Percentage of Unknown
Predicted Licit	119,100	75.76%
Predicted Illicit	38,105	24.24%
Directly connected to known Illicit	4,027	2.56%
Predicted Illicit + directly connected to Illicit	1,942	1.24%

The model predicted 38,105 Unknown transactions as closer to the Illicit class. Because the Unknown group has no ground truth, this number cannot be interpreted as the number of truly illicit transactions. Graph analysis then identified 4,027 Unknown transactions that had at least one direct connection to a known Illicit transaction. When both indicators were combined, the candidate set was reduced to 1,942 transactions, or about 1.24% of all Unknown transactions. As an additional structural check, several Unknown transactions had many direct links to known Illicit nodes. For example, txId 355174807 had 68 direct Illicit connections, txId 339502544 had 63, and txId 139273997 had 59. These counts do not establish illicit status, but they provide an interpretable structural signal for prioritization.

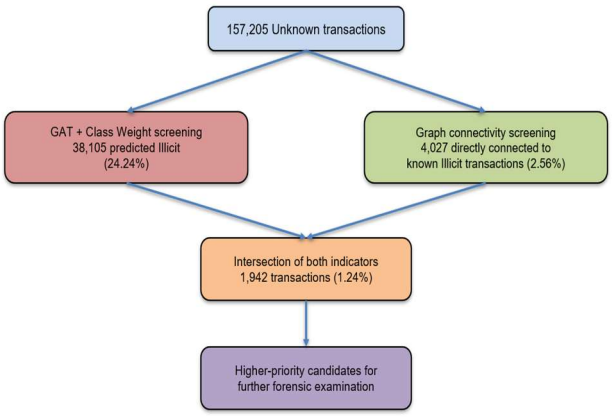


Figure 5. Screening of Unknown transactions using model prediction and direct graph connectivity.

The intersection contains 1,942 transactions, approximately 1.24% of the Unknown group. This smaller subset is supported by both a model-based indicator and a direct structural indicator. It is a screening shortlist rather than a set of confirmed illicit transactions.

4. Discussion

4.1 Model selection and forensic objective

Model selection depends on the operational objective. SVM achieved the strongest accuracy, precision, and Illicit-class F1-score, whereas class-weighted GAT was more sensitive to the minority class. It reduced false negatives from 198 for SVM to 87 and from 166 for unweighted GAT to 87, while reaching 90.43% Illicit-class recall and a ROC-AUC of 0.9809. The improvement was accompanied by more false positives. Class-weighted GAT was therefore selected for Unknown-transaction screening because this study prioritizes missed Illicit cases, not because the model dominates every evaluation metric.

4.2 Effect of class weighting

Class weighting changed the error profile rather than improving every metric. For GAT, Illicit-class recall increased from 81.74% to 90.43% and false negatives decreased from 166 to 87, but false positives increased from 94 to 214 and precision decreased from 88.77% to 79.34%. GCN showed the same direction of change. This pattern is consistent with cost-sensitive learning: assigning a larger penalty to minority-class errors makes the classifier more willing to identify Illicit transactions, at the cost of additional false alarms[18], [19], [20]. In a forensic-screening workflow, this trade-off can be useful when missing an Illicit transaction is considered more consequential than reviewing additional candidates.

4.3 Relation to recent work

Recent Bitcoin anti-money-laundering studies have demonstrated the value of graph convolution, temporal modeling, contrastive learning, attention mechanisms, and semi-supervised strategies [8],[9],[10],[11] ,[12],[13], [14],[15],[21], [22]. Rather than introducing a new GAT architecture,

this study examines how class weighting affects the detection of illicit transactions in an imbalanced Bitcoin transaction graph. The analysis focuses on changes in false negatives and false positives and then applies the selected model to prioritize unlabeled transactions for further screening. Direct numerical comparisons with previous studies are avoided because reported results on the Elliptic dataset can vary with the train-test protocol, temporal design, preprocessing procedure, and metric definition. Therefore, the comparison with previous work focuses on methodological differences rather than a claim of state-of-the-art performance.

4.4 Digital forensic interpretation

The embedding and Unknown-transaction analyses are supporting tools rather than proof of illicit activity. The t-SNE projections provide a qualitative view of the learned representation, while the Unknown analysis combines a model-based signal with direct graph connectivity. Neither signal establishes ownership, intent, or legal attribution. Their practical role is to reduce the search space and help investigators decide which transactions warrant closer examination.

The study outputs can be mapped to the four-stage NIST-oriented forensic workflow used in the thesis. This mapping shows where the model can support an investigation without treating its predictions as final evidence.

Table 6. Mapping of study outputs to the digital forensic workflow.

Forensic stage	Study activity	Output
Collection	Retain transaction features, labels, and the directed edge list.	Elliptic transaction graph for analysis.
Examination	Inspect graph structure and direct links between Unknown and known Illicit transactions.	4,027 Unknown transactions with at least one direct Illicit connection.
Analysis	Apply the selected class-weighted GAT and combine prediction with the structural signal.	38,105 predicted Illicit; 1,942 meet both screening criteria.
Reporting	Report model metrics, uncertainty, and the prioritized Unknown subset.	A review shortlist for further investigation, not confirmed Illicit labels.

5. Limitations

This study has several limitations. First, all experiments use a single public benchmark, so the findings should not be generalized directly to other blockchains or operational anti-money-laundering systems. Second, the reported 80:20 split is stratified rather than time-ordered. Because the Elliptic graph spans discrete time steps, the present evaluation focuses on classification performance under a fixed stratified split and does not evaluate temporal generalization to future transaction periods. Temporal or rolling evaluation may therefore be considered in future studies.

In addition, the reported experiments use a single train-test split with a fixed random seed (Seed 42). Therefore, variability across different random seeds or data partitions was not evaluated, and the reported results should be interpreted within this experimental setting. Third, the Unknown class has no ground-truth labels. Consequently, the 38,105 model-positive Unknown transactions and the 1,942 transactions meeting both screening criteria cannot be used to estimate precision, recall, or accuracy on the Unknown population. Fourth, class weighting reduces false negatives but increases false positives, which may increase analyst workload. Fifth, t-SNE is used only for qualitative visualization and is not statistical evidence of class separability. Finally, direct graph connectivity is a structural screening signal only; it does not establish ownership, intent, identity, or legal attribution.

6. Conclusion and Future Work

This study investigated class-weighted graph attention as an imbalance-aware computational approach for illicit Bitcoin transaction detection and forensic screening. Although SVM achieved the highest accuracy, precision, and illicit-class F1-score, class-weighted GAT achieved the highest illicit-class recall (90.43%) and ROC-AUC (0.9809), reducing false negatives to 87. This result demonstrates a clear trade-off between minority-class sensitivity and false-positive burden rather than universal performance improvement.

The selected model was further used in a two-indicator screening procedure for unlabeled transactions. Among 157,205 Unknown transactions, 38,105 were model-positive, 4,027 had direct connectivity to known Illicit transactions, and 1,942 satisfied both criteria. These transactions constitute a computationally prioritized screening set rather than confirmed illicit activity. Taken together, the results show that GAT predictions and transaction connectivity can be used to narrow the number of Unknown transactions requiring closer examination. The resulting shortlist should be interpreted as a screening output, not as evidence that the selected transactions are illicit.

Future studies may extend the present work by evaluating model stability across different random seeds or data partitions and by adopting temporal or rolling evaluation protocols that preserve the time-step structure of the Elliptic graph. Further research may also investigate external blockchain datasets, probability calibration, alternative imbalance-aware objectives such as focal loss, explainable graph-learning methods, and dynamic graph architectures. The proposed prioritization procedure may additionally be evaluated when later labels or independent investigative evidence become available.

Data and Code Availability

The study uses the publicly available Elliptic Bitcoin Dataset. All experiments were implemented in Python using scikit-learn, PyTorch, PyTorch Geometric, NetworkX, and related scientific-computing libraries. Random seed 42 was used for the reported data split and model training. The preprocessing, training, evaluation, and analysis code developed for this study is available from the corresponding author upon reasonable request.

References

[1] H. F. Atlam, N. Ekuri, M. A. Azad, and H. S. Lallie, "Blockchain Forensics: A Systematic Literature Review of Techniques, Applications, Challenges, and Future Directions," Sep. 01, 2024, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/electronics13173568.

[2] J. Nicholls, A. Kuppa, and N. A. Le-Khac, "The next phase of identifying illicit activity in Bitcoin," *International Journal of Network Management*, vol. 34, no. 5, Sep. 2024, doi: 10.1002/nem.2259.

[3] N. Pocher, M. Zichichi, F. Merizzi, M. Z. Shafiq, and S. Ferretti, "Detecting anomalous cryptocurrency transactions: An AML/CFT application of machine learning-based forensics," *Electronic Markets*, vol. 33, no. 1, Dec. 2023, doi: 10.1007/s12525-023-00654-3.

[4] B. Khemani, S. Patil, K. Kotecha, and S. Tanwar, "A review of graph neural networks: concepts, architectures, techniques, challenges, datasets, applications, and future directions," *J. Big Data*, vol. 11, no. 1, Dec. 2024, doi: 10.1186/s40537-023-00876-4.

[5] V. T. Hoang, H. J. Jeon, E. S. You, Y. Yoon, S. Jung, and O. J. Lee, "Graph Representation Learning and Its Applications: A Survey," Apr. 01, 2023, *MDPI*. doi: 10.3390/s23084168.

[6] F. Guan, T. Zhu, W. Zhou, and K. K. R. Choo, "Graph neural networks: a survey on the links between privacy and security," *Artif. Intell. Rev.*, vol. 57, no. 2, Feb. 2024, doi: 10.1007/s10462-023-10656-4.

[7] A. G. Vrahatis, K. Lazaros, and S. Kotsiantis, "Graph Attention Networks: A Comprehensive Review of Methods and Applications," Sep. 01, 2024, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/fi16090318.

[8] A. Asiri and K. Somasundaram, "Graph convolution network for fraud detection in bitcoin transactions," *Sci. Rep.*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-95672-w.

[9] S. Ouyang, Q. Bai, H. Feng, and B. Hu, "Bitcoin Money Laundering Detection via Subgraph Contrastive Learning," *Entropy*, vol. 26, no. 3, Mar. 2024, doi: 10.3390/e26030211.

[10] I. Alarab and S. Prakoonwit, "Graph-Based LSTM for Anti-money Laundering: Experimenting Temporal Graph Convolutional Network with Bitcoin Data," *Neural Process. Lett.*, vol. 55, no. 1, pp. 689–707, Feb. 2023, doi: 10.1007/s11063-022-10904-8.

[11] I. Alarab and S. Prakoonwit, "Robust recurrent graph convolutional network approach based sequential prediction of illicit transactions in cryptocurrencies," *Multimed. Tools Appl.*, vol. 83, no. 20, pp. 58449–58464, Jun. 2024, doi: 10.1007/s11042-023-17323-4.

[12] Z. Li and E. He, “Graph Neural Network-Based Bitcoin Transaction Tracking Model,” *IEEE Access*, vol. 11, pp. 62109–62120, 2023, doi: 10.1109/ACCESS.2023.3288026.

[13] D. Cheng, Y. Ye, S. Xiang, Z. Ma, Y. Zhang, and C. Jiang, “Anti-Money Laundering by Group-Aware Deep Graph Learning,” *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 12, pp. 12444–12457, Dec. 2023, doi: 10.1109/TKDE.2023.3272396.

[14] F. Wan and P. Li, “A Novel Money Laundering Prediction Model Based on a Dynamic Graph Convolutional Neural Network and Long Short-Term Memory,” *Symmetry (Basel).*, vol. 16, no. 3, Mar. 2024, doi: 10.3390/sym16030378.

[15] M. Li, L. Jia, and X. Q. Su, “Global-local graph attention with cyclic pseudo-labels for bitcoin anti-money laundering detection,” *Sci. Rep.*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-08365-9.

[16] R. Abdulkadirov, P. Lyakhov, and N. Nagornov, “Survey of Optimization Algorithms in Modern Neural Networks,” Jun. 01, 2023, *MDPI*. doi: 10.3390/math11112466.

[17] S. Jung, T. Dagobert, J. M. Morel, and G. Facciolo, “A Review of t-SNE,” *Image Processing On Line*, vol. 14, pp. 250–270, 2024, doi: 10.5201/ipol.2024.528.

[18] W. Chen, K. Yang, Z. Yu, Y. Shi, and C. L. P. Chen, “A survey on imbalanced learning: latest research, applications and future directions,” *Artif. Intell. Rev.*, vol. 57, no. 6, Jun. 2024, doi: 10.1007/s10462-024-10759-6.

[19] L. Xu, H. Zhu, and J. Chen, “Imbalanced graph learning via mixed entropy minimization,” *Sci. Rep.*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-75999-6.

[20] P. Gao, Z. Li, D. Zhou, and L. Zhang, “Reinforced Cost-Sensitive Graph Network for Detecting Fraud Leaders in Telecom Fraud,” *IEEE Access*, vol. 12, pp. 173638–173646, 2024, doi: 10.1109/ACCESS.2024.3448260.

[21] N. Han, R. Zhang, X. Liu, and H. Zhang, “Illicit Bitcoin transaction detection via feature-gated temporal graph learning,” *Sci. Rep.*, May 2026, doi: 10.1038/s41598-026-53783-y.

[22] S. Ferretti, G. D’Angelo, and V. Ghini, “Enhancing Anti-Money Laundering Frameworks: An Application of Graph Neural Networks in Cryptocurrency Transaction Classification,” *IEEE Access*, vol. 13, pp. 50201–50215, 2025, doi: 10.1109/ACCESS.2025.3552240.