Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16245
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Graph-Based Hybrid GNN-Transformer for Imbalanced Credit Card Fraud Detection

Muhammad Bayu WIjaya Putra [1)*], Rinto Priambodo[2)]
[1)]Faculty of Technology and Design, [2)]Universitas Pembangunan Jaya, South Tangerang, Indonesia
[1)]muhammad.bayuwijaya@student.upj.ac.id, [2)] rinto.priambodo@upj.ac.id,

Submitted: May 19, 2026 | **Accepted**: May 24, 2026 | **Published**: July 5, 2026

Abstract: Credit card fraud detection faces two major challenges: severe class imbalance and the limited ability of conventional feature-based models to capture relational patterns among transactions. This study proposes a graph-based Hybrid GNN-Transformer architecture for imbalanced credit card fraud detection by integrating transaction-level relational learning through k-nearest neighbor graph construction and feature-interaction learning through multi-head self-attention. The novelty of this study lies in combining graph-based transaction modeling and Transformer-based feature interaction within a unified architecture. Using the selected graph configuration $k = 3$ and validation-based threshold tuning, the proposed model achieved 79.71% precision, 74.32% recall, 76.92% F1-score, 96.06% ROC-AUC, and 68.65% PR-AUC. Compared with Logistic Regression, Random Forest, and Gradient Boosting baselines, the hybrid model showed competitive fraud detection sensitivity, although the baseline classifiers still achieved stronger overall F1-score and PR-AUC. Ablation results show that the hybrid architecture improves minority-class detection compared with single-branch variants by combining relational transaction information from the GNN branch and feature-interaction information from the Transformer branch. These findings indicate that graph-based hybrid representation learning is a promising direction for imbalanced fraud detection, while further optimization is still required to improve precision-recall balance and competitiveness against strong feature-based baselines.

Keywords: credit card fraud detection; graph neural network; hybrid GNN-Transformer; imbalanced data; graph-based transaction modeling

INTRODUCTION

The rapid growth of digital payment systems has increased the volume and complexity of electronic financial transactions, including credit card transactions. Although this development provides convenience for consumers and financial institutions, it also creates greater opportunities for fraudulent activities. Payment card fraud remains a significant global problem, as fraud losses on global brand cards reached USD 30.74 billion in 2023, increasing from USD 30.40 billion in 2022 (The Nilson Report, 2024). In addition, the Federal Trade Commission reported that consumers lost more than USD 12.5 billion to fraud in 2024, representing a 25% increase over the previous year (Federal Trade Commission, 2025). These figures indicate that fraud detection is not only a technical classification problem, but also an important issue for financial security, customer trust, and operational risk management.

Detecting credit card fraud is challenging because fraudulent transactions usually represent only a very small proportion of total transaction data. Fraud patterns also continue to evolve as fraudsters adapt their strategies to transaction monitoring mechanisms, making static rule-based or purely feature-based detection approaches less sufficient for identifying emerging fraudulent behavior. This severe class imbalance causes conventional classification models to be biased toward the majority class, where normal transactions dominate the learning process. As a result, high accuracy does not necessarily indicate strong fraud detection capability. In practical fraud detection scenarios, false negatives may allow fraudulent transactions to pass undetected, while false positives may incorrectly block legitimate transactions and reduce user trust. Therefore, fraud detection models must be evaluated using metrics that are sensitive to minority-class performance, such as precision, recall, F1-score, ROC-AUC, and PR-AUC (Ali et al., 2022; Baisholan et al., 2025).

Existing machine learning approaches such as Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, and Gradient Boosting have been widely used for credit card fraud detection. These methods are useful as baseline classifiers because they can learn discriminative patterns from transaction features. However,

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

most conventional models treat each transaction as an independent data point and do not explicitly model relationships among transactions. In real transaction environments, fraudulent behavior may appear not only from the attributes of a single transaction, but also from similarities, connections, or behavioral patterns shared across multiple transactions. This limitation creates a need for relational learning approaches that can represent transactions as interconnected entities rather than isolated records (Afriyie et al., 2023; Syahbani et al., 2025; Hernandez Aros et al., 2024).

Deep learning methods provide stronger representation learning capability for complex and nonlinear patterns. However, models that focus only on tabular features or sequential patterns may still be limited in capturing relational structures among transactions. Graph Neural Networks are relevant for this problem because they can model transactions as nodes and represent relationships among transactions as edges. Through graph-based learning, the model can capture relational patterns that may be difficult to identify using ordinary feature-based models. Nevertheless, GNN-based models may not fully capture internal feature interactions within each transaction, especially when fraud patterns depend on complex combinations among transaction attributes (Cherif et al., 2024; Motie & Raahemi, 2024; Cheng et al., 2025).

Transformer-based models offer another advantage through the self-attention mechanism, which enables the model to learn interactions among features with different levels of importance. However, Transformer models applied to tabular transaction data may still lack explicit relational modeling among transactions. Therefore, using either GNN or Transformer alone may not be sufficient to represent both relational transaction patterns and internal feature interactions. A hybrid approach is needed to combine the strengths of both architectures: GNN for relational transaction modeling and Transformer for feature-interaction learning (Chen et al., 2025; Aitha & Pandiri, 2025; Olaniyan et al., 2025).

Based on these limitations, this study proposes a graph-based Hybrid GNN-Transformer model for imbalanced credit card fraud detection. In the proposed approach, transactions are represented as graph nodes using k-nearest neighbor similarity, allowing the GNN branch to learn relational patterns among transactions. At the same time, the Transformer branch learns internal feature interactions through an attention mechanism. The outputs of both branches are then integrated to support fraud classification under highly imbalanced data conditions. The novelty of this study lies in the integration of graph-based transaction modeling and Transformer-based feature interaction learning within a unified hybrid architecture for credit card fraud detection.

This study also implements the trained model into a web-based fraud prediction prototype using FastAPI and an interactive dashboard. However, the system implementation is positioned as a prediction prototype rather than a fully deployed production-level streaming system, because production-level aspects such as latency benchmarking, throughput testing, and concurrent request handling are not the main focus of this study. Therefore, the main contribution of this research is methodological, namely the development of a graph-based Hybrid GNN-Transformer architecture for imbalanced fraud detection, supported by a practical prototype for transaction prediction and monitoring.

LITERATURE REVIEW

Previous studies on credit card fraud detection can be categorized into conventional machine learning, deep learning, graph-based learning, Transformer-based learning, and imbalance handling approaches. Conventional machine learning methods such as Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, and Gradient Boosting are widely used because they are efficient and useful as baseline classifiers. However, these methods generally process transactions as independent feature vectors and do not explicitly capture relationships among transactions. This limitation is important because fraudulent behavior may appear not only from individual transaction attributes, but also from similarities and hidden relationships among multiple transactions (Ali et al., 2022; Afriyie et al., 2023; Syahbani et al., 2025).

Deep learning approaches have been applied to improve representation learning in fraud detection. Models such as CNN, LSTM, autoencoder, and attention-based architectures can capture nonlinear patterns that are difficult for conventional machine learning models to represent. However, models that mainly process transaction data as tabular or sequential input may still be limited in representing relational structures among transactions. Sequential models can learn temporal dependencies, but they do not always capture transaction similarity or network-based fraud behavior. Therefore, deep learning improves feature representation but does not completely solve the relational modeling limitation in fraud detection (Tarissa & Dewayanto, 2024; Habibpour et al., 2023; Chen et al., 2025).

Graph Neural Networks provide a relational learning approach by representing transactions as nodes and relationships among transactions as edges. This structure allows the model to aggregate information from neighboring transactions and capture relational patterns that are difficult to identify using ordinary feature-based models. Nevertheless, GNN-based approaches also have limitations. Their performance depends on the quality of graph construction, including edge definition, number of neighbors, graph sparsity, and graph connectivity. In addition, GNNs may face computational complexity and scalability issues when applied to large-scale transaction

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

networks. Deeper GNN architectures may also suffer from over-smoothing, where node representations become too similar and reduce the model's ability to distinguish fraudulent and legitimate transactions (Cherif et al., 2024; Motie & Raahemi, 2024; Cheng et al., 2025).

Transformer-based models are useful for learning feature interactions through the self-attention mechanism. In fraud detection, attention can help the model assign different importance to transaction features and identify complex feature combinations related to fraudulent behavior. However, Transformer models also have limitations when applied to tabular transaction data because tabular features do not have a natural sequential order like text data. In addition, Transformer models may require sufficient data and computational resources to learn stable attention patterns. When used independently, Transformer models can capture internal feature interactions, but they may not explicitly model relationships among transactions. Therefore, combining GNN and Transformer is relevant because the two architectures address different but complementary aspects of fraud detection: relational transaction modeling and feature-interaction learning (Chen et al., 2025; Aitha & Pandiri, 2025; Olaniyan et al., 2025).Class imbalance remains another critical issue in credit card fraud detection because fraud transactions usually represent only a very small proportion of the dataset. This condition can cause models to be biased toward normal transactions and produce high accuracy while failing to detect fraud cases. Oversampling methods such as SMOTE are commonly used to increase minority-class representation and improve fraud detection sensitivity. However, oversampling may also introduce synthetic patterns that do not fully represent real fraud behavior if not applied carefully. Therefore, imbalance handling must be combined with appropriate evaluation metrics such as precision, recall, F1-score, ROC-AUC, and PR-AUC rather than relying only on accuracy (Amelia et al., 2022; Baisholan et al., 2025; Siagian et al., 2025).

Table 1. **Summary of Previous Studies**

Researcher	Methodological Focus	Strength	Limitation / Gap
Ali et al. (2022)	Machine learning-based financial fraud detection review	Provides broad overview of fraud detection methods	Does not focus on hybrid graph-attention architecture
Afriyie et al. (2023)	Supervised machine learning for credit card fraud detection	Useful as baseline classifiers	Treats transactions mainly as independent feature vectors
Syahbani et al. (2025)	Comparative study of fraud detection algorithms	Shows performance differences among algorithms	Does not integrate graph learning or attention mechanisms
Cherif et al. (2024)	GNN for credit card fraud detection	Captures relational transaction patterns	Does not deeply model internal feature interactions using Transformer
Motie & Raahemi (2024)	Review of GNN for financial fraud detection	Highlights the potential of relational learning	Discusses GNN challenges such as scalability and graph construction
Cheng et al. (2025)	Review of GNN in financial fraud detection	Strengthens the relevance of graph-based fraud detection	Does not specifically focus on hybrid GNN-Transformer for tabular fraud data
Chen et al. (2025)	Deep learning innovation in financial fraud detection	Discusses attention and modern deep learning models	Transformer-based methods may still lack explicit relational modeling
Aitha & Pandiri (2025)	Hybrid GNN-Transformer for financial fraud detection	Shows the potential of integrating GNN and Transformer	Practical implementation and imbalance-specific evaluation remain limited
Olaniyan et al. (2025)	Graph-temporal contrastive Transformer	Combines graph and Transformer-based behavior modeling	Does not specifically focus on credit card fraud prediction prototype
Baisholan et al. (2025)	Credit card fraud detection under class imbalance	Emphasizes imbalance as a major challenge	Does not propose hybrid relational and attention-based modeling

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16245
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Research Gap and Novelty Positioning

Based on the comparison of previous studies, several research gaps can be identified. First, conventional machine learning models are effective as baseline classifiers but generally process transactions as independent feature vectors and do not explicitly capture relational patterns among transactions. Second, deep learning models improve representation learning, but many approaches still focus on tabular or sequential patterns and do not fully utilize relationships among transactions. Third, GNN-based methods are suitable for relational fraud detection, but they depend strongly on graph construction quality and may face computational complexity, scalability, and over-smoothing issues. Fourth, Transformer-based methods can learn feature interactions through attention mechanisms, but they do not explicitly represent transaction relationships when used independently.

Therefore, this study positions its novelty in the development of a graph-based Hybrid GNN-Transformer model for imbalanced credit card fraud detection. The proposed model integrates two complementary representations: relational representations from the GNN branch and feature-interaction representations from the Transformer branch. Transaction relationships are constructed using k-nearest neighbor similarity, while the Transformer branch learns internal interactions among transaction features. This integration is designed to address the limitations of feature-based models, graph-only models, and Transformer-only models by combining relational transaction modeling and feature-interaction learning in a unified architecture. In addition, this study supports the proposed model with a web-based prediction prototype, while positioning the system as a practical prototype rather than a production-level streaming deployment.

METHOD

This study uses a quantitative approach with an experimental method to develop and evaluate a fraud detection system for credit card transactions. The focus of this study is to build a hybrid Graph Neural Network (GNN) and Transformer model that can utilize relationships among transactions and interactions among features in a single unified architecture. This approach is chosen because fraud detection in modern transaction data requires not only individual feature modeling, but also relational patterns among transactions and adaptive capability for imbalanced data (Ali et al., 2022; Motie & Raahemi, 2024; Chen et al., 2025).

Dataset and Data Split

The dataset used in this study is the public Credit Card Fraud Detection dataset, consisting of 284,807 transactions, including 492 fraud transactions and 284,315 non-fraud transactions. The features include Time, Amount, V1–V28, and the Class label, where Class = 0 represents normal transactions and Class = 1 represents fraud transactions. Although this dataset has been widely used, it remains relevant because it represents a real-world-like credit card fraud detection scenario with severe class imbalance, where fraud transactions account for only approximately 0.17% of the total data. This imbalance makes the dataset suitable for evaluating model sensitivity toward rare fraud cases (Ali et al., 2022; Baisholan et al., 2025).

The dataset was selected because it provides a reproducible benchmark, reflects the rarity of fraud cases in financial transaction data, and contains numerical transaction features suitable for similarity-based graph construction. In this study, each transaction is represented as a graph node, while relationships among transactions are formed based on feature similarity. Therefore, the dataset is appropriate for evaluating the proposed graph-based Hybrid GNN-Transformer model, which combines relational transaction modeling and feature-interaction learning (Cherif et al., 2024; Motie & Raahemi, 2024; Cheng et al., 2025).The dataset was divided into training, validation, and testing sets using a stratified split with a proportion of 70%, 15%, and 15%, respectively. Stratification was applied to preserve the fraud and non-fraud class distribution in each subset. To reduce data leakage, preprocessing parameters were fitted only on the training data and then applied to the validation and testing data. Oversampling was also applied only to the training set. However, because the dataset is static and does not represent continuous production transactions, this study does not directly evaluate long-term data drift or evolving fraud behavior. Therefore, generalization to future fraud patterns remains a limitation for future research.

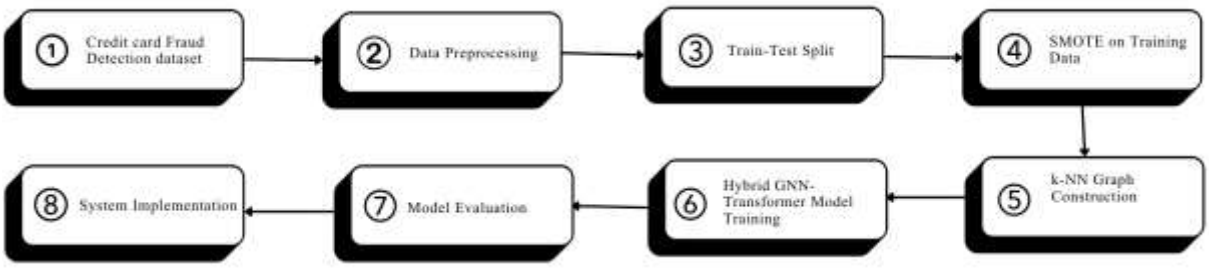

Figure 1. Research Flow

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Data Preprocessing

The preprocessing stage was conducted to prepare the transaction features and reduce the risk of data leakage. The Time and Amount features were normalized using StandardScaler because their scales differ from the anonymized V1–V28 features. The scaler was fitted only on the training data and then applied to the validation and testing data to ensure that no information from evaluation data influenced the training process.

To address the severe class imbalance, this study applied SMOTE only to the training data. SMOTE was selected because it increases minority-class representation by generating synthetic fraud samples from existing minority instances, helping the model learn fraud-related patterns more effectively. Validation and testing data were kept in their original distribution to ensure objective evaluation and avoid leakage from oversampling.

However, SMOTE also has limitations. Synthetic fraud samples may not fully represent real fraudulent behavior and may introduce artificial patterns, especially when fraud patterns are sparse or complex. This risk is referred to as fraud synthesis distortion. Therefore, SMOTE is used only as an imbalance-handling strategy, while the main novelty of this study lies in the graph-based Hybrid GNN-Transformer architecture. Other methods such as ADASYN, Borderline-SMOTE, and GAN-based augmentation were not used in this study. ADASYN may amplify noisy minority samples, Borderline-SMOTE may increase class overlap near the decision boundary, and GAN-based augmentation requires higher computational cost and additional validation to ensure realistic synthetic fraud samples. Therefore, SMOTE was selected as a simpler and more controlled oversampling method for this study (Amelia et al., 2022; Siagian et al., 2025; Baisholan et al., 2025).

Graph Construction

After preprocessing, transaction data were transformed into a graph structure to support relational transaction modeling. Each transaction was represented as a node, while edges were constructed using the k-nearest neighbors approach based on feature similarity in the normalized feature space. This representation enables transactions with similar numerical patterns to be connected, allowing the GNN branch to learn neighborhood-based relational information instead of relying only on individual transaction attributes. Graph-based modeling is relevant for fraud detection because fraudulent behavior may appear not only in isolated transaction features, but also in relational patterns among similar transactions (Cherif et al., 2024; Motie & Raahemi, 2024; Cheng et al., 2025).

The graph was constructed using a similarity-based binary connectivity approach, where edges indicate connections between each transaction and its nearest neighbors. Explicit edge weights were not used in the main implementation to maintain a simple and controlled graph representation. However, edge definition and graph construction quality remain important factors in GNN-based fraud detection, and future work may incorporate distance-based or similarity-based edge weighting to refine relational representation (Motie & Raahemi, 2024; Cheng et al., 2025).

The number of neighbors, k, directly affects graph sparsity, density, connectivity, and computational cost. A smaller k produces a sparser graph but may limit relational information, while a larger k increases connectivity but may introduce noisy relationships from less similar transactions. Therefore, graph sensitivity analysis was conducted using $k = 3$, $k = 4$, and $k = 5$. Based on this analysis, $k = 3$ was selected as the final graph configuration because it provided strong recall and PR-AUC while maintaining a sparse graph structure. The final graph consisted of 30,000 nodes, 90,000 edges, an average degree of 3.0, and a graph density of 0.000100, indicating a sparse graph with sufficient local neighborhood information for relational learning. This graph configuration supports the Hybrid GNN-Transformer architecture by providing relational transaction representations through the GNN branch and feature-interaction representations through the Transformer branch (Aitha & Pandiri, 2025; Olaniyan et al., 2025).

Hybrid GNN-Transformer Architecture

The proposed model uses a Hybrid GNN-Transformer architecture consisting of two parallel branches: a GNN branch and a Transformer branch. This design is used to capture two complementary types of information in credit card fraud detection. The GNN branch learns relational patterns among transactions through the graph structure, while the Transformer branch learns internal feature interactions within each transaction through self-attention. This hybrid strategy is relevant because fraud patterns may emerge from both transaction relationships and complex feature combinations (Chen et al., 2025; Aitha & Pandiri, 2025; Olaniyan et al., 2025).

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16245
e-ISSN : 2541-2019
p-ISSN : 2541-044X

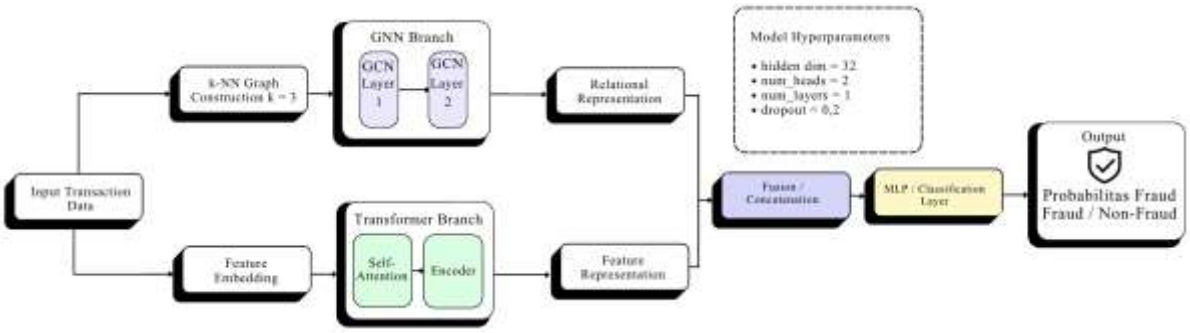

Figure 2. Proposed Graph-Based Hybrid GNN-Transformer Architecture for Credit Card Fraud Detection

Let $X \in \mathbb{R}^{N \times F}$represent the transaction feature matrix, where $N$is the number of transaction nodes and $F = 30$represents the input features consisting of Time, Amount, and V1–V28. In the GNN branch, $X$and the graph edge index are processed using two Graph Convolutional Network layers. The first GCN layer maps the input features into a hidden representation with dimension $D = 32$, followed by ReLU activation and dropout. The second GCN layer further refines the node representation and produces a relational embedding $H_{GNN} \in \mathbb{R}^{N \times 32}$. This branch allows each transaction node to aggregate information from neighboring transactions and capture similarity-based relational patterns, which are important in graph-based fraud detection (Cherif et al., 2024; Motie & Raahemi, 2024; Cheng et al., 2025).

In the Transformer branch, each transaction feature is treated as a feature token. The input tensor is reshaped from $X \in \mathbb{R}^{N \times F}$into $X' \in \mathbb{R}^{N \times F \times 1}$, where each scalar feature is projected into a 32-dimensional embedding space. This produces a feature embedding tensor $E \in \mathbb{R}^{N \times F \times 32}$. The embedded features are then processed using a Transformer encoder with one encoder layer, two attention heads, and dropout of 0.2. The self-attention mechanism enables the model to learn interactions among transaction features and assign different levels of importance to feature combinations that may be relevant for fraud classification.
The attention mechanism in the Transformer branch is formulated as follows:

$$Attention(Q, K, V) = softmax((QK\text{^}T)/\sqrt{d})V$$

where Q, K, and Vrepresent query, key, and value matrices, while $d_k$is the key dimension. In this study, the attention mechanism is used to model relationships among transaction features rather than relationships among transaction nodes. After the Transformer encoder, mean pooling is applied across the feature dimension to obtain a feature-interaction embedding $H_{Trans} \in \mathbb{R}^{N \times 32}$.
The outputs of the two branches are combined using concatenation:

$$\mathrm{HFusion} = [\mathrm{HGNN} \parallel \mathrm{HTrans}]$$

where $H_{Fusion} \in \mathbb{R}^{N \times 64}$. Concatenation was selected because it preserves information from both branches without forcing early compression. The GNN output represents relational transaction information, while the Transformer output represents internal feature-interaction information. By concatenating these representations, the classifier can use both sources of information simultaneously. The fused representation is then passed into a multilayer perceptron classifier consisting of fully connected layers with dimensions 64 → 128 → 64 → 2 to produce the final fraud and non-fraud class logits.

From a computational perspective, the GNN branch depends on the number of nodes and edges in the transaction graph. With the final graph configuration $k = 3$, the training graph consists of 30,000 nodes and 90,000 edges, which keeps the graph sparse and reduces message-passing cost. The Transformer branch depends on the number of features and embedding dimension. Since this study uses 30 features, a hidden dimension of 32, two attention heads, and one encoder layer, the Transformer component remains computationally manageable. Therefore, the architecture is designed to balance relational learning capability, feature-interaction modeling, and computational efficiency.

Training and Evaluation

The model was trained using Cross-Entropy loss and Adam optimizer with a learning rate of 0.001 and weight decay of 1×10^{-5}. Training was conducted for a maximum of 70 epochs with early stopping based on validation PR-AUC to reduce overfitting and improve model selection under class imbalance. During training, F1-score,

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

ROC-AUC, and PR-AUC were monitored because accuracy alone is less reliable for highly imbalanced fraud detection (Ali et al., 2022; Syahbani et al., 2025; Baisholan et al., 2025).

Final evaluation was performed on the testing set using accuracy, precision, recall, F1-score, ROC-AUC, and PR-AUC, with greater emphasis on recall, F1-score, and PR-AUC. Threshold tuning was conducted on the validation set to obtain a better balance between fraud detection sensitivity and false-positive control. The selected threshold was then applied to the testing set, and threshold sensitivity was analyzed using precision, recall, F1-score, false positives, and false negatives.

To strengthen the evaluation, comparison experiments were conducted using Logistic Regression, Random Forest, and Gradient Boosting as baseline classifiers. These models were evaluated using the same data split and evaluation metrics as the proposed model to assess its performance against conventional feature-based approaches (Afriyie et al., 2023; Syahbani et al., 2025). In addition, ablation experiments were conducted using GNN-only, Transformer-only, and Hybrid GNN-Transformer configurations to evaluate the contribution of relational transaction modeling and feature-interaction learning (Chen et al., 2025; Aitha & Pandiri, 2025; Olaniyan et al., 2025).

Robustness analysis was performed using different random seeds to observe performance stability across different initialization settings. Formal cross-validation and statistical significance testing were not conducted due to the computational cost of repeatedly training graph-based deep learning models. Therefore, this limitation is addressed as future work.

System Implementation

The trained model was implemented into a web-based fraud prediction prototype using FastAPI as the backend and an interactive dashboard as the user interface. The system workflow consists of transaction input processing, feature scaling using the same preprocessing parameters from the training stage, model inference, fraud probability generation, and prediction result visualization. The dashboard displays the transaction input, predicted class, fraud probability, risk level, and transaction monitoring information.

In this study, the implementation is positioned as a prototype-level prediction system rather than a production-level real-time fraud detection system. Production deployment requires further evaluation, including streaming transaction integration, concurrent request handling, API throughput, deployment security, and monitoring reliability. Therefore, this study evaluates only prototype-level inference feasibility through latency measurement, while full production benchmarking is left for future work. This implementation demonstrates how the proposed model can be integrated into an application workflow to support fraud prediction and transaction monitoring (Hernandez Aros et al., 2024; Chen et al., 2025).

RESULT

This section presents the experimental results of the proposed graph-based Hybrid GNN-Transformer model and its web-based fraud prediction prototype. The evaluation includes training and validation performance, final test results, baseline comparison, ablation study, threshold sensitivity, graph sensitivity, robustness analysis, confusion matrix, and prototype-level inference performance.

Training Performance

During the training process, the proposed model was optimized for a maximum of 70 epochs with early stopping based on validation PR-AUC. As shown in Figure 3, the training loss decreases consistently across epochs, indicating that the optimization process converges in a stable manner. The curve does not show large fluctuations or sudden increases, suggesting that the model can gradually adjust its parameters without unstable learning behavior. Training stopped at epoch 33 after validation PR-AUC no longer improved, while the best validation PR-AUC was obtained at epoch 23 with a value of 0.5649.

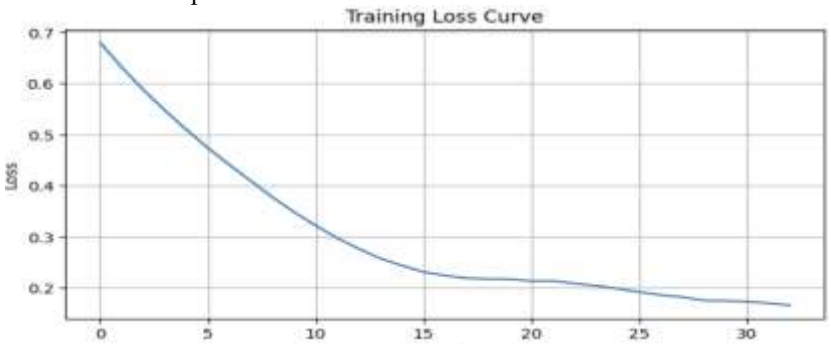

Figure 3. Training Loss Curve

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16245
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Figure 3 indicates that the model can learn data representations progressively while maintaining stable convergence. However, training loss alone is not sufficient to confirm generalization performance or the absence of overfitting. Therefore, the evaluation is further supported by validation metrics, final test performance, threshold sensitivity analysis, and robustness analysis in the following subsections. This interpretation is important because performance variance in imbalanced fraud detection cannot be assessed only from a single loss curve.

Validation Performance

Model performance on the validation data was monitored using F1-score, PR-AUC, and ROC-AUC. As shown in Figure 4, ROC-AUC increases and remains relatively stable at a high level, indicating that the model has a strong ability to distinguish fraud and non-fraud transactions. PR-AUC also improves during training and reaches its best value at epoch 23, showing that the model gradually learns minority-class fraud patterns under imbalanced data conditions. Meanwhile, validation F1-score remains relatively low because it is calculated using the default threshold of 0.5, which is not always optimal for highly imbalanced fraud detection.

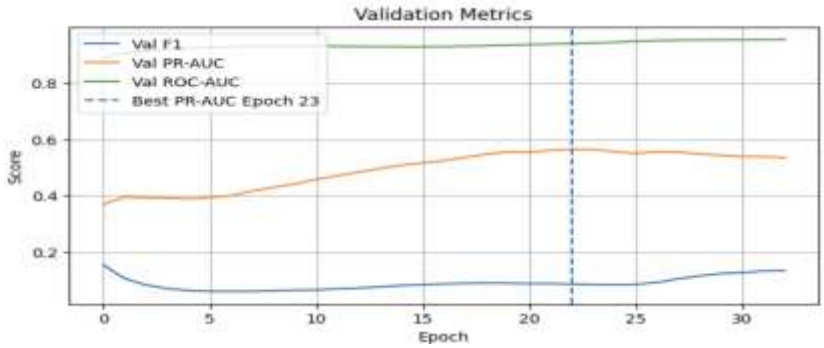

Figure 4. Model Performance on Validation Data

The validation results are not interpreted as standalone evidence of model superiority. Instead, they are used to guide model selection and are further supported by final test performance, baseline comparison, ablation study, and threshold sensitivity analysis in the following subsections. This comparative evaluation is necessary because high validation ROC-AUC alone does not guarantee optimal fraud detection performance, especially when precision-recall trade-offs and minority-class detection capability are more important than overall accuracy.

Test Performance

The final test evaluation was conducted using the selected graph configuration $k = 3$and the validation-tuned threshold. Since the dataset is highly imbalanced, accuracy is reported only as a complementary metric, while the main interpretation focuses on recall, F1-score, and PR-AUC. As shown in Table 2, the proposed model achieved 74.32% recall, 76.92% F1-score, and 68.65% PR-AUC. These results indicate that the model can detect a substantial proportion of fraud transactions while maintaining a reasonable balance between fraud detection sensitivity and prediction reliability.

The selected threshold of 0.994542 reflects a trade-off between recall and precision. At the default threshold of 0.5, the model achieved higher recall but produced many false positives, indicating that many legitimate transactions were incorrectly flagged as fraud. After validation-based threshold tuning, precision increased substantially while recall decreased moderately. This trade-off is important in fraud detection because false positives may disturb legitimate users, while false negatives may allow fraudulent transactions to pass undetected. Therefore, the selected threshold was used to balance fraud sensitivity and false-positive control rather than to maximize accuracy.

Table 2. Evaluation Results of the Hybrid GNN-Transformer Model

Metric	**Value**
Accuracy	99.92%
Precision	79.71%
Recall	74.32%
F1-score	76.92%
ROC-AUC	96.06%
PR-AUC	68.65%
Selected threshold	0.994542

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16245
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Comparative Evaluation

To evaluate the proposed model more comprehensively, comparison experiments were conducted using conventional machine learning baselines and ablation variants. The baseline comparison includes Logistic Regression, Random Forest, and Gradient Boosting, while the ablation study compares GNN-only, Transformer-only, and Hybrid GNN-Transformer configurations.

Table 3. Baseline Comparison Results

Model	Threshold	Accuracy	Precision	Recall	F1-score	ROC-AUC	PR-AUC
Logistic Regression	1.000000	99.95%	90.63%	78.38%	84.06%	96.54%	79.33%
Random Forest	0.910000	99.94%	88.52%	72.97%	80.00%	97.28%	78.82%
Gradient Boosting	0.982166	99.94%	91.53%	72.97%	81.20%	96.72%	75.03%
Proposed Hybrid GNN-Transformer	0.994542	99.92%	79.71%	74.32%	76.92%	96.06%	68.65%

As shown in Table 3, the conventional baseline models achieved stronger overall F1-score and PR-AUC than the proposed Hybrid GNN-Transformer. Logistic Regression produced the highest F1-score and PR-AUC, while Random Forest achieved the highest ROC-AUC. The proposed hybrid model achieved competitive recall, outperforming Random Forest and Gradient Boosting in fraud detection sensitivity, but its lower precision and PR-AUC indicate that further optimization is still needed. This result shows that the hybrid architecture can capture useful relational and feature-interaction information, although strong feature-based classifiers remain highly competitive on this dataset.

Table 4. Ablation Study Results

Model Variant	Threshold	Accuracy	Precision	Recall	F1-score	ROC-AUC	PR-AUC
GNN-only	0.946174	99.93%	85.00%	68.92%	76.12%	96.85%	69.80%
Transformer-only	0.544301	99.78%	35.53%	36.49%	36.00%	93.95%	23.38%
Hybrid GNN-Transformer	0.982952	99.93%	79.17%	77.03%	78.08%	95.41%	70.06%

Table 4 shows that the Hybrid GNN-Transformer achieved the highest recall and F1-score among the ablation variants. The GNN-only model produced higher precision than the hybrid model, indicating that relational transaction modeling is useful for reducing false positives. However, the hybrid model achieved better recall and F1-score, suggesting that combining GNN-based relational learning with Transformer-based feature interaction provides a more balanced fraud detection capability. In contrast, the Transformer-only model showed the weakest performance, indicating that self-attention over tabular features alone is insufficient without relational transaction information. These findings support the contribution of the hybrid architecture, although the precision-recall balance still requires further improvement.

Sensitivity and Robustness Analysis

Sensitivity and robustness analyses were conducted to evaluate the effect of threshold selection, graph configuration, and random initialization on model performance.

Table 5. Threshold Sensitivity Analysis

Threshold	Accuracy	Precision	Recall	F1-score	FP	FN
0.500000	96.95%	4.58%	83.78%	8.69%	1,291	12
0.700000	98.25%	7.76%	83.78%	14.20%	737	12
0.900000	99.65%	31.00%	83.78%	45.26%	138	12
0.950000	99.82%	48.06%	83.78%	61.08%	67	12
0.994542	99.92%	79.71%	74.32%	76.92%	14	19

Table 5 shows that threshold selection strongly affects the trade-off between precision, recall, false positives, and false negatives. Lower thresholds produced higher recall but generated many false positives, while the validation-tuned threshold of 0.994542 substantially reduced false positives to 14 and produced the highest F1-score, although recall decreased to 74.32%. Therefore, the selected threshold provides a more balanced decision boundary for fraud detection.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16245
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Table 6. Graph Sensitivity Analysis

k	Nodes	Edges	Average Degree	Graph Density	Precision	Recall	F1-score	ROC-AUC	PR-AUC
3	30,000	90,000	3.0	0.000100	80.00%	75.68%	77.78%	95.18%	71.56%
4	30,000	120,000	4.0	0.000133	82.09%	74.32%	78.01%	94.09%	68.89%
5	30,000	150,000	5.0	0.000167	79.37%	67.57%	72.99%	94.86%	69.12%

T
able 6 indicates that increasing graph density does not always improve fraud detection performance. The configuration $k = 3$achieved the highest recall and PR-AUC, showing that a sparser graph can provide sufficient local neighborhood information for minority-class detection. Although $k = 4$achieved slightly higher F1-score, $k = 3$was selected as the final graph configuration because it produced stronger PR-AUC, higher recall, fewer edges, and lower computational cost.

Table 7. Robustness Analysis

Seed	Accuracy	Precision	Recall	F1-score	ROC-AUC	PR-AUC
42	99.86%	77.61%	70.27%	73.76%	95.41%	69.50%
123	99.92%	77.03%	77.03%	77.03%	94.09%	70.81%
2025	99.91%	78.26%	72.97%	75.52%	96.03%	70.93%
Mean	99.92%	77.63%	73.42%	75.44%	95.18%	70.41%
Std	0.0036%	0.62%	3.40%	1.64%	0.99%	0.80%

Table 7 shows that the model produced relatively stable results across different random seeds. The low standard deviation of PR-AUC indicates that minority-class detection performance remained consistent, while recall and F1-score also stayed within a reasonable range. These findings suggest that the proposed Hybrid GNN-Transformer has acceptable robustness under the tested initialization settings, although broader validation on additional datasets is still needed.

Confusion Matrix

The confusion matrix was used to analyze the classification errors of the proposed Hybrid GNN-Transformer model in more practical terms. As shown in Table 8, the model correctly classified 42,634 non-fraud transactions and 55 fraud transactions. Meanwhile, 14 non-fraud transactions were incorrectly classified as fraud, and 19 fraud transactions were not detected.

Table 8. Confusion Matrix of the Hybrid GNN-Transformer Model

Actual / Predicted	Non-Fraud	Fraud
Non-Fraud	42,634	14
Fraud	19	55

The 14 false positives indicate legitimate transactions that were incorrectly flagged as fraud. In a real financial system, this type of error may cause unnecessary transaction blocking, customer inconvenience, and additional manual review. However, the number of false positives is relatively low after threshold tuning, indicating that the selected threshold helps control unnecessary fraud alerts. The 19 false negatives represent fraud transactions that were classified as non-fraud. This type of error has a higher financial risk because undetected fraud may result in direct monetary loss and delayed fraud response. Therefore, although the model achieves a reasonable balance between precision and recall, reducing false negatives remains an important priority for future improvement. Overall, the confusion matrix shows that the model can detect most fraud cases while maintaining a low false-positive count, but further optimization is still needed to improve fraud coverage without increasing excessive false alerts.

System Implementation and Inference Performance

The trained Hybrid GNN-Transformer model was integrated into a web-based fraud prediction prototype using FastAPI as the backend and an interactive dashboard as the user interface. The dashboard is used to demonstrate how transaction input, predicted class, fraud probability, risk level, and suspicious transaction monitoring can be presented to users. Therefore, the dashboard is positioned as a prototype interface for model interaction rather than as the main scientific contribution of this study.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16245
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Figure 5. Web-Based Fraud Prediction Prototype Dashboard

To evaluate the inference capability of the prototype, latency measurement was conducted using repeated inference runs on the test data. This evaluation provides an initial indication of whether the trained model can perform prediction efficiently in a prototype environment. However, the result should not be interpreted as full production-level real-time performance because concurrent API requests, streaming transaction integration, database latency, network delay, deployment security, and large-scale monitoring were not evaluated.

Table 9. Prototype Inference Performance

Metric	Value
Number of runs	30
Test transactions	42,722
Average batch inference time	4.1286 s
Minimum batch inference time	3.4696 s
Maximum batch inference time	7.6298 s
Standard deviation	1.0333 s
Average time per transaction	0.0966 ms
Estimated throughput	10,347.74 transactions/s

The latency measurement shows that the proposed model can perform batch inference efficiently in a prototype environment, with an average inference time of approximately 0.0966 ms per transaction. This supports the feasibility of integrating the model into a web-based fraud prediction workflow. Nevertheless, the current implementation remains limited to prototype-level inference. Future work should include API stress testing, concurrent request evaluation, streaming-based deployment, and scalability testing before the system can be considered suitable for production-level fraud monitoring.

DISCUSSIONS

The experimental results show that the proposed Hybrid GNN-Transformer provides competitive fraud detection capability, although it does not outperform all conventional baseline models. Logistic Regression, Random Forest, and Gradient Boosting still achieved stronger overall F1-score and PR-AUC, indicating that feature-based classifiers remain highly competitive on the public credit card fraud dataset. However, the hybrid model achieved balanced performance after threshold tuning, with 79.71% precision, 74.32% recall, 76.92% F1-score, and 68.65% PR-AUC. This suggests that the proposed model is useful for exploring relational and feature-interaction representations, but further optimization is still needed to improve competitiveness against strong baselines.

The behavior of the hybrid model can be explained by the complementary roles of its two branches. The GNN branch captures relational patterns by aggregating information from neighboring transactions in the k-NN graph, allowing the model to identify similarity-based fraud patterns that may be missed when transactions are treated independently. This supports previous findings that graph-based learning is relevant for financial fraud detection because fraud behavior can appear through relational transaction patterns (Cherif et al., 2024; Motie & Raahemi, 2024; Cheng et al., 2025). Meanwhile, the Transformer branch learns internal feature interactions through self-attention, helping the model identify important feature combinations. The ablation results support this interpretation because the Hybrid GNN-Transformer achieved higher recall and F1-score than the Transformer-only model, showing that self-attention alone is insufficient without relational graph information. This finding is consistent with studies on hybrid graph-Transformer architectures for fraud detection (Aitha & Pandiri, 2025; Olaniyan et al., 2025).

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

The threshold and confusion matrix analyses show that the model's prediction behavior depends strongly on the decision threshold. At the default threshold, recall was high but false positives increased substantially. After validation-based threshold tuning, false positives decreased to 14, while false negatives became 19. From a business perspective, false positives may cause unnecessary transaction blocking and manual review, while false negatives are more financially risky because undetected fraud can lead to direct monetary loss. Therefore, the selected threshold reflects a trade-off between fraud sensitivity and false-positive control rather than an attempt to maximize accuracy. Graph sensitivity analysis also shows that graph construction affects model performance. The configuration $k = 3$achieved the highest recall and PR-AUC, while larger values of $k$increased graph density but did not consistently improve minority-class detection. This indicates that denser graphs may introduce less relevant neighborhood relationships, while a sparse graph can preserve useful local transaction similarities with lower computational cost. This confirms the importance of graph construction quality in GNN-based fraud detection (Motie & Raahemi, 2024; Cheng et al., 2025).

Several limitations remain. First, the model was evaluated on one public dataset, so its generalization to other transaction environments, evolving fraud patterns, and data drift still requires further testing. Second, the graph was built using binary feature-similarity edges, while distance-based or weighted edges were not explored. Third, the GNN branch depends on graph size and edge density, while the Transformer branch adds computational cost through attention-based feature interaction. Although prototype inference performance was feasible, the system has not been tested under production-level conditions such as concurrent API requests, streaming integration, database latency, and large-scale deployment. Future research should explore richer graph construction, edge weighting, broader datasets, computational optimization, and production-oriented scalability testing.

CONCLUSION

This study proposed a graph-based Hybrid GNN-Transformer model for imbalanced credit card fraud detection by combining relational transaction modeling and feature-interaction learning in a unified architecture. The final model used a k-nearest neighbor graph with $k = 3$ and achieved 79.71% precision, 74.32% recall, 76.92% F1-score, 96.06% ROC-AUC, and 68.65% PR-AUC after validation-based threshold tuning.The main scientific contribution of this study is the integration of graph-based relational learning and Transformer-based feature interaction for fraud detection. Methodologically, this study contributes through baseline comparison, ablation study, threshold sensitivity analysis, graph sensitivity analysis, and robustness testing. Practically, the trained model was implemented into a web-based fraud prediction prototype and evaluated through prototype-level inference latency.However, the proposed model did not outperform all conventional baseline classifiers in F1-score and PR-AUC, indicating that further optimization is still needed. Future research should explore weighted graph construction, larger and more diverse datasets, computational optimization, and production-level evaluation involving concurrent API requests, streaming transactions, and scalability testing.

ACKNOWLEDGMENT

The authors would like to express their gratitude to all parties who provided support during the research, system development, and preparation of this article. The authors also thank the academic institution that provided a learning environment and supporting facilities so that this study could be completed properly.

REFERENCES

Afriyie, J. K., Tawiah, K., Pels, W. A., Addai-Henne, S., Dwamena, H. A., Owiredu, E. O., Ayeh, S. A., & Eshun, J. (2023). A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions. *Decision Analytics Journal, 6*, 100163. https://doi.org/10.1016/j.dajour.2023.100163

Aitha, A. R., & Pandiri, L. (2025). A hybrid GNN-transformer framework for explainable and scalable financial fraud detection. In *Proceedings of the 1st International Conference on Intelligent Methods and Advanced Computer Scientific Innovations (IMACSI 2025)* (Vol. 1, pp. 424–432). SCITEPRESS. https://doi.org/10.5220/0014159100004932

Ali, A., Abd Razak, S., Othman, S. H., Eisa, T. A. E., Al-Dhaqm, A., Nasser, M., Elhassan, T., Elshafie, H., & Saif, A. (2022). Financial fraud detection based on machine learning: A systematic literature review. *Applied Sciences, 12*(19), 9637. https://doi.org/10.3390/app12199637

Amelia, T. S., Hasibuan, M. N. S., & Pane, R. (2022). Comparative analysis of resampling techniques on machine learning algorithm. *Sinkron: Jurnal dan Penelitian Teknik Informatika, 7*(2), 628–634. https://doi.org/10.33395/sinkron.v7i2.11427

Baisholan, N., Dietz, J. E., Gnatyuk, S., Turdalyuly, M., Matson, E. T., & Baisholanova, K. (2025). A systematic review of machine learning in credit card fraud detection under original class imbalance. *Computers, 14*(10), 437. https://doi.org/10.3390/computers14100437

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Chen, Y., Zhao, C., Xu, Y., Nie, C., & Zhang, Y. (2025). Deep learning in financial fraud detection: Innovations, challenges, and applications. *Data Science and Management*. Advance online publication. https://doi.org/10.1016/j.dsm.2025.08.002

Cheng, D., Zou, Y., Xiang, S., & Jiang, C. (2025). Graph neural networks for financial fraud detection: A review. *Frontiers of Computer Science, 19*(9), 199609. https://doi.org/10.1007/s11704-024-40474-y

Cherif, A., Ammar, H., Kalkatawi, M., Alshehri, S., & Imine, A. (2024). Encoder–decoder graph neural network for credit card fraud detection. *Journal of King Saud University – Computer and Information Sciences, 36*, 102003. https://doi.org/10.1016/j.jksuci.2024.102003

Federal Trade Commission. (2025, March 10). *New FTC data show a big jump in reported losses to fraud to $12.5 billion in 2024*. https://www.ftc.gov/news-events/news/press-releases/2025/03/new-ftc-data-show-big-jump-reported-losses-fraud-125-billion-2024

Habibpour, M., Gharoun, H., Mehdipour, M., Tajally, A. R., Asgharnezhad, H., Shamsi, A., Khosravi, A., & Nahavandi, S. (2023). Uncertainty-aware credit card fraud detection using deep learning. *Engineering Applications of Artificial Intelligence, 123*, 106248. https://doi.org/10.1016/j.engappai.2023.106248

Hernandez Aros, L., Bustamante Molano, L. X., Gutierrez-Portela, F., Moreno Hernandez, J. J., & Rodríguez Barrero, M. S. (2024). Financial fraud detection through the application of machine learning techniques: A literature review. *Humanities and Social Sciences Communications, 11*, 1130. https://doi.org/10.1057/s41599-024-03606-0

Motie, S., & Raahemi, B. (2024). Financial fraud detection using graph neural networks: A systematic review. *Expert Systems with Applications, 240*, 122156. https://doi.org/10.1016/j.eswa.2023.122156

Olaniyan, J., Olaniyan, D., Obagbuwa, I. C., & Ngafeeson, M. (2025). Graph-temporal contrastive transformer for financial fraud detection using transaction behavior modeling. *Algorithms, 18*(12), 770. https://doi.org/10.3390/a18120770

Siagian, N. A., Sipayung, S. P., Rikki, A., & Marbun, N. (2025). Integrating SMOTE with XGBoost for robust classification on imbalanced datasets: A dual-domain evaluation. *Sinkron: Jurnal dan Penelitian Teknik Informatika, 9*(3), 1094–1107. https://doi.org/10.33395/sinkron.v9i3.15029

Syahbani, A. M., Firdaus, W., & Musodo, K. A. (2025). A comparative study of data mining algorithms for fraud detection in financial transactions. *Sinkron: Jurnal dan Penelitian Teknik Informatika, 9*(2), 814–821. https://doi.org/10.33395/sinkron.v9i2.14645

Tarissa, B. V., & Dewayanto, T. (2024). Penerapan *machine learning* dan *deep learning* pada peningkatan deteksi *credit card fraud*: *A systematic literature review*. *Diponegoro Journal of Accounting, 13*(3), 1–15.

The Nilson Report. (2024). *Card fraud losses worldwide in 2023* (No. 1276). https://nilsonreport.com/articles/card-fraud-losses-worldwide-in-2023/

*name of corresponding author