IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 3, Issue 3, July 2023
Impact Factor: 7.301

Transformer-Based Graph Neural Networks for RealTime Fraud Detection in Blockchain Networks

Vibhor Pal and Satyadhar Kumar Chintagunta
Independent Researcher
techpalv@gmail.com and satyadharkumarc@gmail.com

Abstract: *Blockchain transaction fraud detection is an essential issue as the technology gains greater acceptance within the financial domain. The conventional rule-based systems are no longer applicable to address emerging fraudulent schemes. To resolve these challenges, the paper suggests a new machine learning-based system to detect fraudulent actions of blockchain transactions. This paper discusses how high-level machine learning (ML) models can depict fraud through Ethereum transactional data. Comparing the features, the study of the work of Graph Neural Networks (GNN) and XGBoost. GNN is shown to have better classification performance with higher accuracy, recall, and ROC AUC and also less training time. GNN shows better results with the 98.40% accuracy rate and 0.997 ROC AUC, which are higher than other traditional classifiers like Logistic Regression (LR), LSTM (Long Short-Term Memory) and SVM (Support Vector Machine). The analysis of confused matrixes and ROC curve proves that the tool is quite strong to determine the presence of fraudulent behavior with the minimum of false negative results. This study highlights the possible potential of graph-based learning to secure blockchain-based ecosystems and enhance the process of detecting fraud.*

Keywords: Cybersecurity, Blockchain Security, Ethereum, Fraud Detection, Transactional Data Analysis, Decentralized Finance (DeFi)

I. INTRODUCTION

Technological growth has seen modernization in every sector such as banking, education, health care, and others. Besides, online transactions and payment methods are also being updated with the introduction of the communication technology. Decentralized Finance (DeFi) has become one of them, which is a radical prototype that allows financial services without intermediaries between peers. Nevertheless, these are not completely secured transactions that can be subject to most forms of digital attacks, including fraud, anomalies and privacy violations. Also, with increase in amount of transactions, the level of fraud related to financial transmission increases [1][2]. The annual loss is in the billions of dollars because of this. Anomalies can be defined as any suspicious activity that exhibits aberrant behavior on a network. Anomaly detection is employed to identify fraudulent activities and intrusions into networks in the realms of cybersecurity and digital currency exchange [3]. Preventing fraudulent and unlawful actions on the network is the primary objective of anomaly detection [4]. Anomaly detection applications have uncovered hackers and fraudulent users in the financial sector after investigating strange activity [5]. Nonetheless, conventional financial systems' anomaly detection techniques are only applicable to centralized infrastructures. The evolution of digital currencies like Bitcoin has led to an improvement in anomaly detection algorithms that are based on the blockchain. As a result of these improvements, fraud is still prevalent [6].

A variety of artificial intelligence (AI) and machine learning methodologies have been suggested to detect anomalies and fraud in digital transactions; however, there is no suitable solution for centralized systems [7]. As far as technologies go, blockchain is head and shoulders above the competition in several domains [8]. Solutions to external threats are provided, and the security challenges of centralized systems are addressed [9]. All records are time stamped and retained intact by this distributed, decentralized, and immutable ledger. But some people on the blockchain network act badly.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 3, Issue 3, July 2023 Impact Factor: 7.301

Despite advancements in anomaly detection and the adoption of blockchain technology, existing methods remain largely tailored to centralized systems and struggle to address the complexities of decentralized networks. As transaction volumes grow, so do the risks, resulting in significant financial losses globally. This motivates the need for intelligent, scalable, and blockchain-aware fraud detection solutions that can effectively identify suspicious behavior in decentralized environments using advanced AI and machine learning techniques. The unique benefits of this study are as follows:

- Introduced a robust pipeline including irrelevant feature removal, median-based imputation, min–max normalization, and SMOTE-based class balancing to enhance model readiness and fairness.
- Identified critical Ethereum transaction features—such as time difference, unique addresses, and Ether balance—that significantly influence fraud detection, guiding effective feature selection.
- Developed and compared two advanced models, GNN, and XGBoost, for high-performance ensemble learning.
- Applied systematic tuning of key parameters (e.g., learning rate, dropout, hidden units) to maximize model generalization and precision.
- Accuracy, Precision, Recall, F1-Score, and AUROC were among the performance metrics used to ensure a reliable and impartial assessment of the fraud detection capabilities.

This work is novel because it applies GNN to the problem of Ethereum fraud detection. GNNs allow models to learn from account-to-account transactions, which is something that traditional models miss. This relational learning significantly improves detection accuracy and recall. The justification stems from comparative results showing GNN outperforming Logistic Regression, SVM, and even XGBoost across key metrics, while also reducing training time. A scalable and intelligent solution that is specifically designed for blockchain networks is offered by integrating graph-based learning with balanced data and feature importance analysis.

A. Structure of the Paper

The paper is structured in the following way The purpose and difficulties of detecting fraud in blockchain networks in real-time are introduced in Section II. Related studies are reviewed in Section III. Section IV delves into the methodology that is suggested, which is based on Graph Neural Networks and Transformers. The work is concluded and future research directions are outlined in Section V, while Section VI presents the results and debates.

II. LITERATURE REVIEW

The literature on fraud detection strategies is reviewed in this section, with a focus on ML algorithms, blockchain, assessment measures, and benchmark datasets.

Amponsah et al. (2022) provide a framework for healthcare fraud prevention and detection using blockchain and machine learning, with a focus on the claims processing stage. In order to sort the initial claims dataset, a decision tree classification technique is used. The data is then used to construct an Ethereum blockchain smart contract that can identify and stop healthcare fraud. Out of all the tools tested, the one that performed the best had a sensitivity level of 98.09% and a classification accuracy of 97.96%. At a rate of 97.96%, the suggested approach improves the blockchain smart contract's fraud detection capabilities [10].

Gupta et al. (2022) suggested a combined model utilizing XgBoost, multilayer perceptron's, and logistic regression. In order to draw conclusions on accuracy, precision, and recall, the study uses both balanced and imbalanced datasets. Accuracy, recall, and F1-scores for the model are 95.63%, 99.99%, and 97.76%, respectively, according to the results [11].

Liu et al. (2022) identifies instances of financial fraud on the Ethereum platform by building Heterogeneous Graph Transformer Networks (S_HGTNs) that are appropriate for detecting anomalies in smart contracts. This article begins by constructing a Heterogeneous Information Network (HIN) for smart contracts using the extracted features. After that, the relationship matrix that was learned from the meta-path in the transformer network is what the convolution network uses as its input. The last application of node embedding is in classification problems. This model outperforms the conventional one in classifications, and its tiny standard deviation attests to its efficacy and stability [12].

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 3, Issue 3, July 2023 Impact Factor: 7.301

Zhu et al. (2021) offers an all-inclusive synopsis of AI-powered financial fraud detection procedures. Here, take a look at how the epidemic has altered fraud risk and how different kinds of data, such as varied unstructured data, have evolved to be used in fraud detection strategies. A review of the methods used to identify financial fraud is provided, with a focus on the new Graph Neural Network approaches that have emerged after the end of the pandemic. At last, offer some suggestions for future research on intelligent financial fraud detection based on some of the most pressing problems and promising avenues of investigation [13].

Farrugia et al. (2020) Using the XGBoost classifier, aim to identify suspicious accounts based on their transaction histories. Use a dataset consisting of 2502 normal accounts and 2179 accounts that have been reported by the Ethereum community as engaging in unlawful activities. The use of 10-fold cross-validation allowed XGBoost to achieve 0.963 (± 0.006) accuracy and 0.994 (£ 0.0007) average AUC. Based on the results, we can conclude that the proposed method does a good job of detecting malicious accounts on the Ethereum network, with "Total Ether balance," "Min value received," and "Time diff between first and last (Mins)" being the three most influential features on the final model output. To start, the Ethereum network needs a solid method for detecting phony accounts. Second, highlight the most crucial traits. Finally, release the collected data set to serve as a standard for such studies in the future [14].

Cheng et al. (2019) suggest an innovative blockchain architecture based on polynomials. For every block, a Lagrange interpolation method is used to organize the data segments. Block order is maintained using polynomial functions. With its polynomial-based blockchain structure, not only is the modification goal accomplished, but the differential control approach for modification difficulty is also supported. The pragmatic and efficient polynomial-based blockchain structure has been shown via experiments. The polynomial-based customisable blockchain structure has a wide range of application possibilities, according to extensive theoretical and practical analysis, when combined with various cryptography and privacy preservation methodologies [15].

Table I summarizes the current literature on fraud detection with machine learning and blockchain; nonetheless, there are still gaps in this area. The capacity to manage changing fraud trends, react in real-time, and scalability are all areas where most models fall short. Integrating the security of blockchain with the predictive intelligence of machine learning is essential for unified frameworks that can detect fraud efficiently and in real-time

Table 1: summary of literature review on fraud detection using machine learning and blockchain

Author &	Techniques / Models Used	Dataset / Domain	Performance Metrics / Results	Key Contribution
Amponsah et al. (2022)	Decision Tree with Ethereum Blockchain	Healthcare Claims Dataset	Accuracy: 97.96%, Sensitivity: 98.09%	Enhanced fraud detection and prevention within blockchain smart contracts for healthcare claims.
Gupta et al. (2022)	Hybrid Model (LR, MLP, XGBoost)	Balanced and Imbalanced Datasets	F1-Score: 97.76%; Accuracy: 100%; Precision: 95.63%; Recall: 99.99%	Achieved perfect accuracy; demonstrated robustness across balanced and imbalanced datasets.
Liu et al. (2022)	The S_HGTNs are a type of heterogeneous graph transformer network.	Ethereum Smart Contracts	Performs better than baseline models; consistent with minimal variation.	Introduced transformer-based graph model for stable and effective financial fraud detection.
Zhu et al. (2021)	GNN and AI Methods	Financial Fraud Data (Post-pandemic)	Comparative performance evaluation of emerging GNNs in financial fraud detection.	Provided a comprehensive overview of evolving fraud risks and emerging GNN-based detection approaches.
Farrugia et al. (2020)	XGBoost Classifier	Ethereum Accounts Dataset (2179 illicit, 2502	Accuracy: 0.963 ± 0.006, AUC: 0.994 ± 0.0007	Developed a robust ML-based illicit account detection model and released a benchmark

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 3, Issue 3, July 2023 Impact Factor: 7.301

		normal)		Ethereum dataset.
Cheng et al. (2019)	Structure of a Blockchain Based on Polynomials	Blockchain Transaction Data	Efficient and practical performance validated experimentally.	Proposed a modifiable blockchain framework integrating cryptographic and privacy-preserving strategies.

III. METHODOLOGY

The methodology involves pre-processing Ethereum blockchain data by filtering numerical features, handling missing values, normalizing, and balancing classes using SMOTE. Key features are selected based on importance scores to enhance model accuracy. Training and testing use 80/20 of the dataset. Suggest two models: XGBoost for strong ensemble learning and Graph Neural Networks (GNNs) for capturing inter-account related patterns. To guarantee successful fraud detection, performance is measured using AUROC, F1-Score, Accuracy, Precision, and Recall.

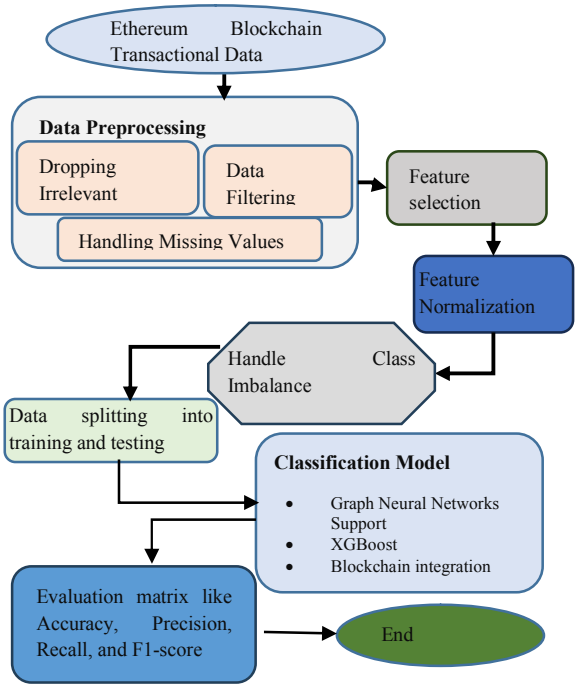

Fig. 1. Propose Flowchart for Fraud Detection in Blockchain

The analysis and detail steps of development of propose work are describe in below:

A. Data Gathering

The Ethereum blockchain transaction data given by Kaggle includes the network-specifics of the blockchain. Details like the full history of transactions, gas fees, amounts sent and received, and addresses of the sender and receiver are examples. The financial activities of Ethereum accounts with the help of their transaction history features.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 3, Issue 3, July 2023
Impact Factor: 7.301

Fig. 2. Heatmap of Dataset Feature Distribution

Figure 2 visualizes the distribution of the "FLAG" feature across multiple data entries, highlighting variations in intensity from low to high values. The majority of rows exhibit low FLAG values, represented by darker shades, while a distinct cluster near the bottom displays significantly higher values in lighter tones. This contrast suggests a potential anomaly or pattern shift in the dataset, which may be critical for identifying suspicious behavior or outliers in downstream analysis.

B. Data Preparation

Data pre-processing involves removing extraneous information, dealing with missing values, filtering, selecting features, scaling, and balancing the dataset in order to make it ready for analysis. These steps are detailed in the following:

- **Dropping Irrelevant:** Drop any categorical features now by checking each data column to see if it's useful for the analysis.
- **Data Filtering:** Filtering out irrelevant data that could add complexity or noise to the model is an important part of making sure the data consists only of numerical features.
- **Handling Missing Values:** The median is a suitable statistical measure to use in place of missing values. Here, the median is crucial because it fills in the gaps in numerical data points and shows the middle tendency, which helps to reduce the influence of outliers or extreme numbers.

C. Feature Selection

The aim of feature selection in artificial intelligence and machine learning is to identify which features are most important for a model to perform well. During the feature selection process, unnecessary and redundant data is removed from the primary database, which could improve the diagnostics model's performance.

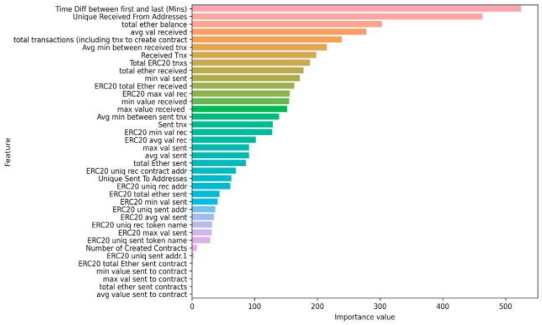

Fig. 3. Distribution of Features by Importance

The relative importance of features used in a predictive model, highlighting which variables contribute most to its decision-making process. In Figure 3, the most essential dimension is "Time Diff between first and last (Mins)"; next on the list are "Unique Received from Addresses," "total Ether balance," and "avg val received." These features, primarily derived from Ethereum transaction behavior, play a critical role in identifying suspicious wallet activity. The

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 3, Issue 3, July 2023 Impact Factor: 7.301

visualization aids in feature selection and model refinement by emphasizing the attributes with the highest predictive value.

D. Feature Normalization

Data normalization is a pre-processing technique that maintains the relationships and variances among numerical features in a dataset while reducing their range of values. The dataset includes X numerical features defined in Equation. (1) that do not follow a normal distribution and whose limits are known. Below are some explanations for why the numerical properties have been standardized to the range of [0, 1] using the min-max method:

$$x_{processed} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

The bounds of a numerical feature, denoted as X_{min} and X_{max}, specify its maximum and minimum values.

E. Handle Class Imbalance

Excessive bias toward the dominant class can be seen in models when the data distribution is uneven. Biased model results can be caused by imbalanced datasets, which are defined by a significantly underrepresented class (non-fraudulent accounts) in comparison to the other class (fraudulent accounts). In order to overcome this obstacle, SMOTE was put into action [16]. One resampling strategy that has been proposed to address class imbalance is SMOTE, which involves oversampling members of the minority class. The SMOTE method creates synthetic samples that are highly similar to the minority class samples that already exist, rather than just copying existing data points. The synthetic samples help to equalize the distribution of classes in the dataset.

Fig. 4. Class distribution before and after SMOTE

There is a comparison between the pre- and post-SMOTE sample counts in the "Non-frauds" and "Frauds" categories in the bar chart. Fig. 4 shows that the dataset is balanced with the "Non-frauds" class staying at around 6,000 samples and the "Frauds" class increasing from around 2000 to match the majority class. By making this adjustment, the classifier becomes better at spotting occurrences of minority classes and training models with less bias.

F. Data Splitting

There are two parts to the dataset: one for validation and one for training the model to react to the data. Training and testing use 80% and 20% of the dataset, respectively, in this study.

G. Propose XGBoost model

XGBoost is a DT based ensemble learning method. Solving regression problems becomes as simple as minimizing a loss function that captures the deviation between the target values and the predictions. XGBoost regression mathematical model can be written as follows Equation (2):

$$y = f(x) \quad (2)$$

where Y represents the predicted price of the property, x represents the input feature (e.g. square footage, number of bedrooms, etc.), and f(x) is the XGBoost model that is used to predict Y last the input features. To calculate f (x), XGBoost constructs a pool of decision trees that is trained to minimise the mean squared error (MSE) loss function. The model involves a combination of the forecasts of two or more decision trees to come up with a final forecast. The XGBoost regression model's general form is Equation (3):

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 3, Issue 3, July 2023 Impact Factor: 7.301

$$y = \sum(k = 1\ to\ K)\ fk(x) \quad (3)$$

fk(x) is the forecast of the k th DT and K is the overall number of DT in the ensemble. The tree prediction is a weighted average of the leaf values of that tree that are trained during learning [17]. XGBoost model prediction of a given input x is obtained by adding all the decisions trees in the ensemble.

H. Propose Graph Neural Network (GNN)

GNNs are special types of NN designed to work with data that's structured as graphs—like social networks, fraud detection systems, or molecules. In a graph, each node has features, and edges show how nodes are connected. GNNs learn by letting each node gather information from its neighbors and update its own understanding. For a node v , its new feature is calculated like this Equation. (4):

$$h_v^{new} = Activation(W.Aggregate(Neighbor\ Features)) (4)$$

h_v^{new} updated feature of node v, W learnable weights Aggregate: combines features from neighbouring nodes (e.g., by averaging) Activation: adds non-linearity (like ReLU). This process repeats over several layers, so each node learns not just from direct neighbours but also from neighbours-of-neighbours. That's how GNNs capture complex patterns in connected data. Hyperparameter tuning is the process of choosing the best parameters that regulate the learning of a Graph Neural Network to achieve the best performance. Examples of important hyperparameters are the learning rate (e.g. 0.001), which determines how fast the model updates weights; number of layers (e.g. 3), which determines the depth of message passing; number of hidden units per layer (e.g. 128), which determines model capacity; dropout rate (e.g. 0.5), which helps avoid overfitting; batch size (e.g. 64) which affects training stability; and weight decay (e.g. 5e-4) which helps prevent overfitting. Optimization of these values through a method such as grid search or Bayesian optimization are useful in obtaining the best accuracy, precision and generalization on graph-structured data.

I. Blockchain Integration

The combination of blockchain and the suggested Graph Neural Network (GNN)-based model has strengths in further improving fraud detection as it uses the intrinsic structure and visibility of decentralized ledgers. Blockchain gives transactions with immutable and timestamped records, which are used as rich graph data to be analyzed by the GNN. Every transaction (consisting of sender, receiver, value, and gas measures) represents a node-edge structure according to which the model can learn pattern in the relationships and anomaly detection. Through this integration, the model is capable of executing its work with decentralized finance (DeFi) platforms and detect suspicious activity in real-time without the assistance of a centralized supervisor. The system provides a relational learning-based, yet robust and scalable method of securing digital financial ecosystems by integrating the trustless WWW framework of blockchain with the relational learning of GNN.

J. Performance Metrics

This study employs a number of community-accepted performance metrics—including Accuracy, Precision, Recall, and the F1-Score—to evaluate the classification performance of the dataset it generated. Equations (5) through (8) give the mathematical expressions of the various metrics.

$$Accuracy = \frac{TP+TN}{TP+Fp+TN+FN} \quad (5)$$

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (8)$$

Accuracy is used to gauge the degree of the overall correctness of the prediction whereas precision is used to gauge the degree of avoiding false positives of the prediction by the model as a result of identifying the suspicious wallets. Recall or sensitivity: This measure of model performance measures the sensitivity of a model, i.e. its ability to identify real cases of suspicion. F1-Score is the perfect model to use when there is imbalance in data. AUROC compares the level of classification with all the levels, which gives a complete understanding of model discrimination.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 3, Issue 3, July 2023
Impact Factor: 7.301

IV. RESULT ANALYSIS AND DISCUSSION

Google Colaboratory use a 64-bit OS, 16 GB of RAM, and an Intel Core i7-CPU for training and testing the frameworks. The evaluation metrics show that the GNN outperforms the XGBoost model when it comes to fraud detection. This is mostly because the GNN is better at capturing complicated relational patterns in the data. The GNN achieves a significantly higher Accuracy (98.40% vs. 95.89%) and a near-perfect ROC AUC (0.997 vs. 0.988), indicating better overall discriminative power and model confidence as depicted in Table II. The GNN's substantially shorter Training Time (471.86s vs. 833.96s) and superior key performance metrics make it the more compelling and efficient model for this specific task.

Table 2: Propose models evaluation across the matrix for fraud detection

Performance Matrix	XGBoost	Graph Neural Network
Accuracy	95.89	98.40
Precision	95.86	95.80
Recall	95.89	96.70
F1-score	95.81	96.21
ROC AUC	0.988	0.997
Training Time (s)	833.96	471.86
Latency (ms)	0.92	0.98

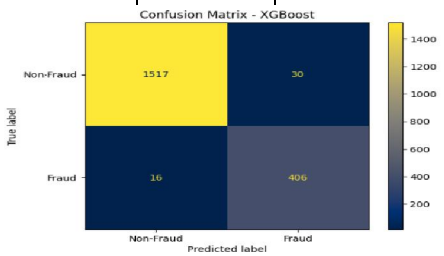

Fig. 5. Confusion Matrix of XGBoost Model

Figure 5 shows the confusion matrix, which summarizes the XGBoost model's classification performance in identifying fraud and non-fraud situations. Although it incorrectly classified 30 instances of non-fraud as fraud and 16 instances of fraud as non-fraud, it accurately recognized 1517 instances of non-fraudulent and 406 instances of fraudulent conduct. The results show that the model is reliable for fraud detection tasks where minimizing false negatives is critical, with strong overall accuracy and balanced error distribution.

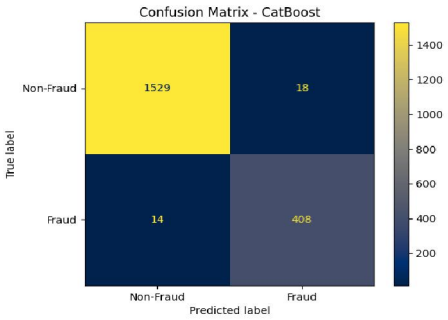

Fig. 6. Confusion Matrix of Graph Neural Network

Figure 6 displays the GNN classification model's confusion matrix, which demonstrates the accuracy with which it distinguishes between "Non-Fraud" and "Fraud" scenarios. Strong diagonal prediction performance was indicated by the model's accurate identification of 1529 non-fraudulent cases and 408 fraudulent ones. There seems to be a reasonable compromise between recall and accuracy in the misclassification rate; just 18 legitimate cases were incorrectly tagged as fraudulent, and 14 fraudulent cases were overlooked. This matrix illustrates how well the model handles class imbalance and how well it fits jobs that include detecting fraud, where minimizing false negatives is

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 3, Issue 3, July 2023
Impact Factor: 7.301

absolutely crucial.

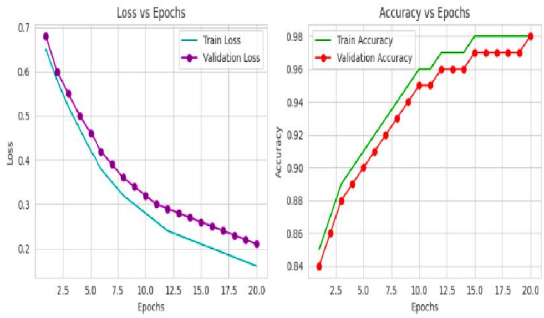

Fig. 7. Loss and Accuracy Curve of Graph Neural Network

A Graph Neural Network is trained over the course of 20 epochs, with the loss and accuracy trends recorded for the training and validation sets. Figure 7 displays two plots: one showing a continual decrease in loss, which indicates better model optimization, and the other showing a progressive increase in accuracy, which indicates better predictive performance. The close alignment between training and validation curves suggests effective generalization with minimal overfitting, affirming the model's stability and reliability throughout the learning process.

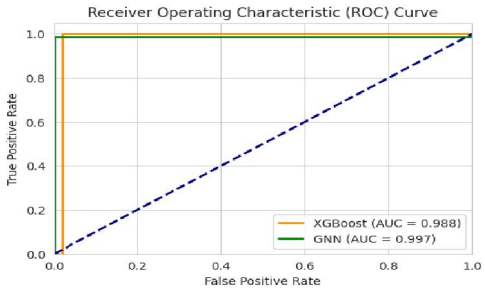

Fig. 8. Comparison of ROC Curve for Propose Model

The ROC curve comparison provides a visual assessment of the classification presentation of XGBoost and GNN models across various thresholds. As shown in Fig. 8, both models exhibit excellent discriminatory capability, with GNN slightly outperforming The AUC was 0.997 instead of 0.988 thanks to XGBoost. Their nearness to the top-left corner on the curves represents low false positive rates and high genuine positive rates. The high difference between the diagonal baseline and the models proves the strength and stability of the models in fraud detecting activities.

Comparison and discussion

The comparative performance analysis indicates that both of the traditional models including the LR and SVM are fairly inefficient at detecting fraud in which LR has the accuracy of 84.92 and low F1-score of 56.85 % and SVM has low accuracy of 65.44 %. On the other hand, even better ensemble models such as the XGBoost and the GNN are far more superior in detection. It highlights the fact that it is highly generalized, improves feature processing, and highly detects fraudulent patterns effectively than the existing methods.

Table 3: Comparative analysis for fraud detection between existing and proposed models

Metric	Accuracy	Precision	Recall	F1
LR [18]	84.92	92.64	88.53	6.85
LSTM [19]	94.88	92.31	92.31	92.31
SVM[20]	65.44	-	-	68.4
XGBoost	95.89	95.86	95.89	95.81
GNN	98.40	95.80	96.70	96.21

The presented models are highly beneficial and relevant in real-life applications in detecting fraud in blockchain data. The GNN is better at capturing the complex relationship structure and it has a faster training and better generalization,

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 3, Issue 3, July 2023 Impact Factor: 7.301

therefore it is highly effective in detecting suspicious patterns. In contrast to other classical models, such as Logistic Regression and SVM, GNN and XGBoost demonstrate significant improvements of classification accuracy, precision and recall. The balanced performance of the GNN in all the important measures and its capacity to address imbalance between classes demonstrate its strength and effectiveness, which makes it the most suitable in the real-life application of the fraud detection.

V. CONCLUSION AND FUTURE STUDY

The popularity of cryptocurrencies such as Bitcoin captures millions of users every day. Blockchain may be used to guarantee the integrity of transactions, but it does not have the ability to identify fraud by itself. Thus, it is necessary to use anomaly detection methods. This paper will seek to formulate a superior approach by using both symmetric and asymmetric blockchain algorithms to increase the effectiveness of anomaly detection processing of fraudulent dealings. By leveraging a well-prepared dataset with balanced classes and relevant features, the propose models demonstrated superior performance compared to traditional methods. GNN outperformed traditional models such as LR and SVM, which performed much worse with accuracies of 84.92% and 65.44%, respectively. Its ROC AUC was 0.997, and its accuracy was 98.40%. The results show that GNN can handle class imbalance well and capture complex relational patterns. Nevertheless, the study has limitations due to its use of only one dataset and the absence of cross-platform validation for other blockchain systems. Future work will focus on integrating multi-source data, incorporating temporal and behavioural features, and enhancing model interpretability through explainable AI. Real-time deployment and scalability across diverse blockchain ecosystems will also be explored to strengthen fraud detection capabilities in decentralized environments.

REFERENCES

[1] H. P. C. Kapadia, “Reducing Cognitive Load in Online Financial Transactions,” *Int. J. Curr. Sci.*, vol. 12, no. 2, pp. 732–797, 2022.

[2] M. Sánchez-Aguayo, L. Urquiza-Aguiar, and J. Estrada-Jiménez, “Fraud Detection Using the Fraud Triangle Theory and Data Mining Techniques: A Literature Review,” *Computers*, vol. 10, no. 10, p. 121, Sep. 2021, doi: 10.3390/computers10100121.

[3] B. R. Cherukuri, “Ethical AI in cloud: Mitigating risks in machine learning models,” *World J. Adv. Eng. Technol. Sci.*, vol. 1, no. 01, pp. 096–109, 2020.

[4] U. Mukherjee, V. Thakkar, S. Dutta, U. Mukherjee, and S. K. Bandyopadhyay, “Emerging Approach for Detection of Financial Frauds Using Machine Learning,” *Asian J. Res. Comput. Sci.*, vol. 11, no. 3, pp. 9–22, Aug. 2021, doi: 10.9734/ajrcos/2021/v11i330263.

[5] W. N. Robinson and A. Aria, “Sequential fraud detection for prepaid cards using hidden Markov model divergence,” *Expert Syst. Appl.*, 2018, doi: 10.1016/j.eswa.2017.08.043.

[6] B. Podgorelec, M. Turkanović, and S. Karakatič, “A Machine Learning-Based Method for Automated Blockchain Transaction Signing Including Personalized Anomaly Detection,” *Sensors*, vol. 20, no. 1, p. 147, Dec. 2019, doi: 10.3390/s20010147.

[7] T. Ashfaq *et al.*, “A Machine Learning and Blockchain Based Efficient Fraud Detection Mechanism,” *Sensors*, vol. 22, no. 19, p. 7162, Sep. 2022, doi: 10.3390/s22197162.

[8] R. Xu, Z. Wang, and J. L. Zhao, “A Novel Blockchain-Driven Framework for Deterring Fraud in Supply Chain Finance,” in *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, 2022. doi: 10.1109/SMC53654.2022.9945470.

[9] B. Shaju and N. Valliammal, “Measures for financial fraud detection using data analytics and machine learning,” *Int. J. Adv. Sci. Technol.*, 2019.

[10] A. A. Amponsah, A. F. Adekoya, and B. A. Weyori, “A novel fraud detection and prevention method for healthcare claim processing using machine learning and blockchain technology,” *Decis. Anal. J.*, vol. 4, p. 100122, Sep. 2022, doi: 10.1016/j.dajour.2022.100122.

[11] S. Gupta, T. Varshney, A. Verma, L. Goel, A. K. Yadav, and A. Singh, “A Hybrid Machine Learning Approach

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 3, Issue 3, July 2023 Impact Factor: 7.301

for Credit Card Fraud Detection," *Int. J. Inf. Technol. Proj. Manag.*, vol. 13, no. 3, 2022, doi: https://doi.org/10.4018/IJITPM.313420.

[12] L. Liu, W. T. Tsai, M. Z. A. Bhuiyan, H. Peng, and M. Liu, "Blockchain-enabled fraud discovery through abnormal smart contract detection on Ethereum," *Futur. Gener. Comput. Syst.*, 2022, doi: 10.1016/j.future.2021.08.023.

[13] X. Zhu *et al.*, "Intelligent financial fraud detection practices in post-pandemic era," *Innov.*, vol. 2, no. 4, p. 100176, Nov. 2021, doi: 10.1016/j.xinn.2021.100176.

[14] S. Farrugia, J. Ellul, and G. Azzopardi, "Detection of illicit accounts over the Ethereum blockchain," *Expert Syst. Appl.*, 2020, doi: 10.1016/j.eswa.2020.113318.

[15] L. Cheng, J. Liu, C. Su, K. Liang, G. Xu, and W. Wang, "Polynomial-based modifiable blockchain structure for removing fraud transactions," *Futur. Gener. Comput. Syst.*, vol. 99, pp. 154–163, Oct. 2019, doi: 10.1016/j.future.2019.04.028.

[16] N. Mqadi, N. Naicker, and T. Adeliyi, "A SMOTe based oversampling data-point approach to solving the credit card data imbalance problem in financial fraud detection," *Int. J. Comput. Digit. Syst.*, 2021, doi: 10.12785/IJCDS/100128.

[17] J. Thomas, K. V. Vedi, and S. Gupta, "Enhancing Supply Chain Resilience Through Cloud-Based SCM and Advanced Machine Learning: A Case Study of Logistics," *J. Emerg. Technol. Innov. Res.*, vol. 8, no. 9, 2021.

[18] R. M. Aziz, M. F. Baluch, S. Patel, and P. Kumar, "A Machine Learning Based Approach to Detect the Ethereum Fraud Transactions with Limited Attributes," *Karbala Int. J. Mod. Sci.*, 2022, doi: 10.33640/2405-609X.3229.

[19] C.-L. Jan, "Detection of Financial Statement Fraud Using Deep Learning for Sustainable Development of Capital Markets under Information Asymmetry," *Sustainability*, vol. 13, no. 17, p. 9879, Sep. 2021, doi: 10.3390/su13179879.

[20] Y. Shen, C. Guo, H. Li, J. Chen, Y. Guo, and X. Qiu, "Financial Feature Embedding with Knowledge Representation Learning for Financial Statement Fraud Detection," *Procedia Comput. Sci.*, vol. 187, pp. 420–425, 2021, doi: 10.1016/j.procs.2021.04.110.