XBlock-ExplainNet: An Interpretable Deep Learning Model for Fraud Detection in Blockchain-based Fintech Systems

M.R.Sivagar
Associate Professor
Department of CSE-AIML
Sri Sairam Engineering College, Sai Leo Nagar, West Tambaram, Chennai -44
sivagar.ramki@gmail.com

Narsaiah Battu
Assistant Professor
Computer Science & Applications
Dr.B.R. Ambedkar Open University, Jubilee Hills, Hyderabad, Telangana, India – 500033
vnbattu6@gmail.com

Jayaram K
Assistant Professor
Computer Science and Engineering
College of Engineering Kottarakkara, ETC PO, Kottarakkara, Kerala, India.
jayaramk4u@gmail.com

Dr.V. Dyana Christilda
Associate Professor
Department of ECE
RMK College of Engineering and Technology
dyanachristildaece@rmkcet.ac.in

Edwin Arshina
Asst.Profesor,
school of engineering, Department of computer science and engineering
Vels University
Vistas, Pallavaram
Arshina.sudhakaran@gmail.com

Srimathi S
Assistant Professor
Department of Bio Technology, Saveetha School of Engineering, SIMATS, Chennai, Tamilnadu, India.
srimathinanmaran@gmail.com

***Abstract*— As the fintech industry has evolved to use blockchain technology rapidly, fraudulent transactions and fraud using blockchains have become a significant issue to consider as a result of the decentralization category of blockchain networks which are pseudonymous. The typical machine learning models cannot deal efficiently with graph-structured data in the blockchain and tend to operate as black boxes, which cannot be applied to the regulated financial setting. Drawing on graph-based learning, temporal modeling, and explainable AI, we in this work introduce the XBlock-ExplainNet, which is an interpretable deep learning framework to detect the fraud effectively. The architecture suggested has a two stream feature extractor which processes graphs and transaction sequences independently. They are integrated and provided to a Graph-Aware Temporal Transformer Encoder (GTTE) that initiates blockchain type with self-attention operations. In order to solve the problem of transparency, an Explainable Fusion Layer has been presented which combines SHAP feature attribution and internal attention scores to provide transparency to the decision-making done. Outputting of final predictions with confidence scores is done by using a residual classifier. The Elliptic Bitcoin Dataset was used to provide experimental validation of the model, which exhibited an accuracy of 95.6%, F1-score of 94.0, and a ROC-AUC score of 0.96, which were far better than the traditional methods, including GCN and GAT. The explainability score of the framework was also quite high (0.93), proving the lack of opacity and the model applicability to real-world. The findings are sufficient to corroborate that XBlock-ExplainNet presents a novel benchmark to fraud detection within blockchain fintech systems through a balance between interpretability and accuracy.**

Keywords*— Blockchain fraud detection, explainable AI, deep learning, graph neural networks, temporal transformer, SHAP, fintech security.

I. INTRODUCTION

Financial technology (fintech) systems using blockchain have transformed the manner in which the digital transactions are undertaken, providing traceability, impossibility to alter and decentralization. These features have rendered blockchain the cornerstone of the contemporary fintech systems, such as decentralized finance (DeFi), the implementation of smart contracts, peer-to-peer lending, and networks based on cryptocurrencies as a form of payment. Nevertheless, such features that enable trustless transactions and independent financial activity also lead to an emergence of new types of fraud and malpractice, including phishing, fake wallets, address poisoning, transaction laundering and stealthy money laundery activities. Real time detection of the fraudulent transactions within the blockchain networks continues to be a critical but challenging work, owing to the complexity of the real-time financial data, bulk of the transactional data and its sparsity.

The conventional fraud detection solutions have been built within a centralized banking setting but are dependent on the structured data and supervised inputs which are not expected to fare that well in the decentralized anonymous blockchain ecosystems. In addition, most current blockchain fraud detection models utilize shallow machine learning algorithms which are not intuitive, or black-box deep learning models which are arguably not at all. This unconfirms explainability trend in decision-making has led to a gap of trust, among the auditors, regulators, and the stakeholders, when it comes to implementation of deep learning models in significant fintech applications. This brings about the acute demand of another breed of detection systems against fraud whose accuracy, scale, and interpretability, trustworthiness, and audibility are not the only sought attributes.

Recent developments in Graph Neural Networks (GNNs), Temporal Transformers, and Explainable AI (XAI) have shown the possibility of modelling transaction graphs and defining unknown behavioral patterns. The representative studies have been conducted on the basis of GCN, LSTM and attention network in the field of blockchain fraud detection. As an example, GraphSAGE and GCN-based models were applied to task of classifying licit and illicit addresses on the Bitcoin blockchain. These models however tend to overlook time irregularities or separate graph structures and transactions timeline resulting in disjoint presentations. On the contrast, transformer based models have demonstrated advantages in modeling sequences but cannot natively use graph dependencies. In addition, a small number of these models consider feature-level interpretability, which makes the process of decision-making less explicit and inadequate in real-life applications. These constraints have inspired the design of XBlock-ExplainNet, a new and interpretable deep learning architecture which smoothly combines both graph topology, transaction dynamics, and explainability in one architecture. In the model, the aim is to fill the gap between performance and transparency in the blockchain fraud identification structures through its formulation of the heterogeneous nature of transactional information alongside providing human comprehensible grounds to its prediction. The key aim of this study would be to formulate an efficient, scalable and explainable deep learning model that would be able to detect fraudulent transactions in blockchain based fintech platforms with sufficient accuracy. The model must be able to learn relational structure of the blockchain as well as transaction behavior over time, and yet be explainable to financial regulators.

The major contributions of this research work are as following:

1. Our system, XBlock-ExplainNet is a two-branch deep structure, which uses graph-based encoders and temporal convolution-attention modules to collectively model topological and temporal features of transactions.
2. To incorporate the structural biases of the blockchain graph in the self-attention mechanism, a novel Graph-Aware Temporal Transformer Encoder (GTTE), employing the structural information of the blockchain graph on long-distance contextual dependency learning is proposed.
3. In an Explainable Fusion Layer, the SHAP-based feature attributions are combined with attribute weights within the network and allow decisions to be made in an explainable and transparent manner.
4. We support the model on the Elliptic bitcoin Dataset, which is a real life block chain transaction dataset comprising of more than 200,000 nodes and 166 features, and reveal how well our method works at capturing licit and illicit behaviors..

II. Related Works

New advances in artificial intelligence are changing the financial fraud detection scene greatly, more specifically machine learning, blockchain, and explainable AI approaches. Kirandeep Kaur, R. Venkatesh, Sumit Pundir, Priya M, Sushil Kumar, and Rajkumar R have suggested a model trained with machine learning to enhance the detection of fraud in a blockchain payment service through the increase in accuracy and flexibility. Such a pattern provides a high level of detecting patterns of fraud in decentralized systems [1]. In a direct study, Philip Olaseni Shoetan and Babajide Tolulope Familoni examined the developments of effective AI algorithms in fintech fraud detection. In their work it was revealed that deep neural networks and ensemble learning models have better accuracy as compared to the rest on classifying between genuine and deceptive transactions particularly when trained on a huge real-life dataset [2]. On the same note, Abbassi Hanae, Saida El Mendili, and Gahi Youssef suggested a hybrid architecture, which integrates both machine learning and blockchain to improve transactional fraud detection mechanisms in fintech systems [3].

Chayan Ghosh, Avigyan Chowdhury, Nabanita Das and Sanjay Kumar Kuanar developed a combination and deep learning model framework required to detect fraud in the Bitcoin network. Their model showed great increments in classification metrics of performance, as they have shown more precision and recall [4]. Shimal Sh. Taher, S. Ameen, and Jihan Abdulwahhab Ahmed are interested in the problem of the use of ensemble learning and explainable AI to detect fraud in the cryptocurrency exchange market, supporting more transparency and trust of stakeholders in the AI tool [5]. Naresh Kumar Bathala, T. S. Sasikala, Vinay Prasad, K. Rajasekhar, P. Sai Ram and Arvind Kumar came up with a deep learning model that can combine SHAP analysis to explain the model output. Such an approach offered better interpretability in identifying financial fraud using synthetic datasets of simulated real-life settings [6]. Halima Oluwabunmi Bello, Courage Idemudia, and Toluwalase Vanessa Iyelolu developed the conceptual scheme which ensures the incorporation of machine learning technology together with blockchain in order to provide the real-time protection of financial operations related to fraud threat assessment and prevention [7]. Deep neural networks with explainable AI have been presented by Giannis Konstantinidis and Alexander Gegov in support of anti-money laundering action increasing the transparency of the model decision although synthetically generated financial logs have been used [8]. In the meantime, C. P. Shirley, Thanga Helina S, Berin Jeba Jingle, M. Kavitha, and L. Mary presented the Secure Sentinel model that works based on machine learning to enhance security in blockchain-based transactions due to unusual trends in the frequency of transactions and addresses [9]. Rishabh Ralli, Gargi Jugran, Mayank Gaurav and Deepak Bansal have proposed an ensemble based system that helps to identify fraudulent blockchain accounts, essentially identifying outliers by looking at their behaviour [10]. Amir Sohel, M. A. Alam, Md. Waliullah, Md. Jamilur Rahman, and Md. Masud Rana discussed a data science-based method of real-time tracking of financial fraud, basically showing the relevance of the concept of federated learning and blockchain on the decentralized detection of the fiscal broach [11].

Algimantas Venckauskas, Sarunas Grigaliunas, Linas Pocius, Deividas Gedgauda, and Rytis Maskeliunas suggested a model that would detect illegal cryptocurrency transactions through the implementation of machine learning algorithms on blockchain systems that would assist law enforcement agencies in monitoring money laundering operations [12]. T. To work on this, Aravind, Hasan Hussain, S. Vinodhkumar, R. Jothikumar, R. Murugan and R. Vijay considered adversarially resilient deep learning architectures in the development of robust financial fraud detection systems, having better explainability [13]. Vijin Ram and Sathyamurthy J appeared, and it can be applied to financial transaction monitoring, proving their model on credit card data of Spanish banks, and producing acceptable fraud detection rate and low false positive rate [14]. R. Deep fraud net is proposed by Udayakumar, Abhijit Joshi, B. S., and Nirav Patel that can help in detection of fraud in financial transactions because of inclusion of cybersecurity mechanism in detecting fraud together with deep learning on the efficiency of the classification [15].

Manal Loukili, Faycil Messaoudi and Raouya El Youbi spoke about the use of AI-strengthened real-time fraud detection in financial services and the optimization of high-speed transaction activities [16]. Nabiha Faisal, Janifer Nahar, N. Sultana, and L. S. Gaffoor compared different AI algorithms in detecting banking frauds, providing the analysis of the difficulties and possible directions of further research on intelligent systems of fraud detection [17]. M. S. Balamurugan suggested a deep learning strategy, which can be applied to detect fraudulent schemes in detail and supports the detection of a variety of fraud schemes with high flexibility and accuracy due to their modification according to the trend of fraud [18]. Performance of machine learning and deep learning models Comparisons between the machine learning and deep learning models were made by Rishabh Sharma and Ajay Sharma reflecting on the digital finance fraud detection use case [19]. Lastly, Sheng Wang, Yunpeng Liu, Ningtao Wang, Kai Qiao, Jun Liu, Mengxin Li, Kun Yue, Chaoqun Zhan, and Bo Liu suggested that it is important to utilize unlabeled suspicious transaction as a way of enhancing the process of stealth fraud detection where noticeable progress on the aspect of sensitivity was successfully recorded as a result of this strategy [20].

III. PROPOSED WORK

This research introduces XBlock-ExplainNet, a novel interpretable deep learning framework for detecting fraudulent transactions in blockchain-based fintech systems. The model integrates a dual-stream feature extractor that captures both graph-structured and temporal transaction patterns. A Graph-Aware Temporal Transformer Encoder (GTTE) models long-range dependencies while preserving structural context. An Explainable Fusion Layer combines SHAP and attention scores to enhance transparency in predictions. The proposed framework achieves high accuracy and interpretability, making it suitable for real-world financial fraud detection scenarios and it is demonstrated in fig.1.

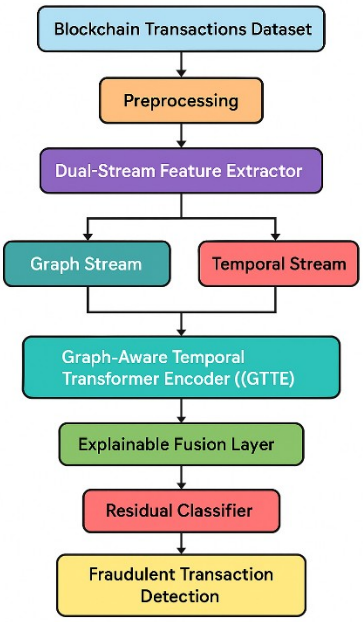

Fig.1: Workflow of the Proposed Work

A.Data Preprocessing

1. **Data Integration and Graph Representation**

The dataset is inherently a graph, where each node represents a Bitcoin transaction and each edge denotes a flow between transactions. The node feature matrix $\mathbf{X} \in \mathbb{R}^{n \times d}$ represents n transactions with $d = 166$ features. The graph structure is defined by an adjacency matrix $A \in \{0,1\}^{n \times n}$, where:

$$A_{ij} = \begin{cases} 1, & \text{if there is a directed edge from node } i \text{ to } j \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

To prepare the graph:

- Parse the edge list to build A
- Normalize features per node using Z-score normalization:

$$\hat{x}_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j} \quad \text{where} \quad \mu_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}\,, \quad \sigma_j = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \mu_j)^2} \quad (2)$$

2. **Handling Missing Labels for Semi-Supervised Learning**

The dataset contains a significant number of transactions with unknown labels. Let $\mathcal{L} \subset \{1, \ldots, n\}$ denote the indices of labeled nodes and $\mathcal{U}$ for unlabeled ones. We define the label matrix as:

$$\mathbf{Y} = \begin{cases} \mathbf{y}_i \in \{[1,0,0], [0,1,0], [0,0,1]\}, & \text{if } i \in \mathcal{L} \\ \text{null}, & \text{if } i \in \mathcal{U} \end{cases} \quad (3)$$

To retain information from unlabeled data during training, we construct a **semi-supervised loss function**:

$$\mathcal{L}_{\text{semi}} = \sum_{i \in \mathcal{L}} \mathcal{L}_{\text{CE}}(\hat{y}_i, y_i) + \lambda \sum_{(i,j) \in \mathcal{E}} \| \hat{y}_i - \hat{y}_j \|^2 \quad (4)$$

where $\mathcal{L}_{\text{CE}}$ is the cross-entropy loss and the second term enforces label smoothness over the graph.

3. **Dimensionality Reduction (Optional but Recommended)**

To reduce computational overhead, dimensionality reduction can be applied:

Principal Component Analysis (PCA):

$$Z = XW, \quad \text{where } W = \arg\max_{W^\top W = I} \operatorname{Tr}(W^\top S W) \quad (5)$$

with S as the covariance matrix $S = \frac{1}{n} X^\top X$

This transformation preserves variance and reduces noise by selecting top $k \ll d$ components.

4. **Graph Normalization and Self-Loop Inclusion**

For Graph Neural Networks (GNNs), normalized adjacency matrix $\widehat{\mathrm{A}}$ is required:

$$\widetilde{\mathrm{A}} = \mathrm{A} + \mathrm{I}_n, \quad \widehat{\mathrm{A}} = \widetilde{\mathrm{D}}^{-\frac{1}{2}}\widetilde{\mathrm{A}}\widetilde{\mathrm{D}}^{-\frac{1}{2}} \tag{6}$$

where $\widetilde{\mathrm{D}}$ is the diagonal degree matrix of $\widetilde{\mathrm{A}}$. The inclusion of the identity matrix I_n adds self-loops to maintain transaction context.

5. **Label Encoding for Supervised Tasks**

Since the dataset uses textual labels (licit, illicit, unknown), these are mapped into one-hot encoded vectors:

$$\text{Label: } \begin{cases} \text{'licit'} \rightarrow [1,0,0] \\ \text{'illicit'} \rightarrow [0,1,0] \\ \text{'unknown'} \rightarrow [0,0,1] \end{cases} \tag{7}$$

This representation is suitable for softmax classifiers and GNN-based node classification.

6. **Train-Test Split with Temporal Ordering**

Because transactions are temporally ordered, splitting must preserve time consistency to simulate real-world scenarios. Let T be the global timestamp range, then:

- Training set: $t \in [t_0, t_k]$
- Testing set: $t \in [t_{k+1}, t_{\text{end}}]$

This avoids temporal data leakage:

$$\mathcal{D}_{\text{train}} = \{x_i \mid t_i \leq t_k\}, \quad \mathcal{D}_{\text{test}} = \{x_i \mid t_i > t_k\} \tag{8}$$

B.Proposed Architecture

To address the growing need for explainable and high-accuracy fraudulent transaction detection in blockchain-powered fintech systems, we propose a novel deep learning framework termed XBlock-ExplainNet. This architecture is explicitly designed to handle both the graph-structured nature of blockchain transactions and their temporal irregularities, while providing transparent predictions through integrated interpretability mechanisms. The model comprises four key modules: (1) Dual-Stream Feature Extraction, (2) Graph-Aware Temporal Transformer Encoder (GTTE), (3) Explainable Fusion Layer, and (4) Residual Classifier.

1. **Dual-Stream Feature Extraction Module**

Let the transaction input be denoted by $x_i \in \mathbb{R}^d$, comprising $x_i = [x_i^g, x_i^t]$, where:

$x_i^g \in \mathbb{R}^{d_1}$: graph-based features

$x_i^t \in \mathbb{R}^{d_2}$: temporal and statistical features

Graph Feature Encoder (G-Encoder)

This branch processes neighborhood-aware representations via a **GraphSAGE-inspired** aggregation mechanism:

$$h_i^{(l+1)} = \sigma\left(W^{(l)} \cdot \mathrm{AGG}\left(\{h_j^{(l)} : j \in \mathcal{N}(i) \cup \{i\}\}\right) + b^{(l)}\right) \tag{9}$$

Where $h_i^{(l)}$ is the hidden representation at layer l, $\mathrm{AGG}(\cdot)$ could be mean or LSTM aggregator, $\sigma(\cdot)$ is a non-linearity (e.g., ReLU), Output: $z_i^g = h_i^{(L)} \in \mathbb{R}^{k_1}$

Temporal Feature Encoder (T-Encoder)

This stream captures short-term dependencies in transaction patterns via a temporal convolution-attention unit:

$$z_i^t = \sum_{j=1}^{T} \alpha_{ij} \cdot \mathrm{Conv1D}(x_{ij}^t), \quad \text{with } \alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{T} \exp(e_{ik})} \tag{10}$$

$$e_{ij} = \tanh(W_q x_i^t + W_k x_j^t) \tag{11}$$

Here, T is the time window size, and $z_i^t \in \mathbb{R}^{k_2}$ is the final temporal representation.

2. **Graph-Aware Temporal Transformer Encoder (GTTE)**

The fused embedding $z_i = [z_i^g \parallel z_i^t] \in \mathbb{R}^{k_1+k_2}$ is input into the GTTE module, which enhances traditional multi-head self-attention with graph-biased weighting.

Multi-Head Graph-Attention

Let $Q = z_i W^Q$, $K = z_j W^K$, and $V = z_j W^V$. The attention coefficient is defined as:

$$\alpha_{ij} = \frac{\exp\left(\frac{Q_i K_j^{\top}}{\sqrt{d_k}} + \gamma \cdot \hat{A}_{ij}\right)}{\sum_{k=1}^{n} \exp\left(\frac{Q_i K_k^{\top}}{\sqrt{d_k}} + \gamma \cdot \hat{A}_{ik}\right)} \tag{12}$$

$\hat{A}_{ij}$: normalized adjacency matrix, γ: learnable graph bias coefficient, Output of each head:

$$\mathrm{Head}_i = \textstyle\sum_j \alpha_{ij} V_j, \tag{13}$$

Final transformer output:

$$\mathrm{GTTE}(z_i) = \mathrm{Concat}(\mathrm{head}_1, \ldots, \mathrm{head}_h) W^O + \mathrm{FFN}(z_i) \tag{14}$$

3. **Explainable Fusion Layer (EFL)**

To integrate interpretability into the model, the fused feature vector z_i is analyzed using both:

SHAP feature attributions $\phi_i \in \mathbb{R}^k$

Attention score maps $\alpha_i \in \mathbb{R}^k$

These are combined using weighted fusion:

$$s_i = \lambda_1 \cdot \mathrm{Normalize}(\phi_i) + \lambda_2 \cdot \mathrm{Normalize}(\alpha_i) \tag{15}$$

$$z_i^{\text{expl}} = z_i \odot s_i \tag{16}$$

Where $\lambda_1, \lambda_2 \in [0,1]$ are tunable weights, $\odot$: element-wise multiplication, z_i^{expl}: interpretable transaction embedding,

4. **Residual Classification Head**

The final interpretable embedding z_i^{expl} is passed through a **residual MLP classifier**:

$$h_1 = \mathrm{ReLU}(W_1 z_i^{\text{expl}} + b_1) \tag{17}$$

$$h_2 = \mathrm{ReLU}(W_2 h_1 + b_2) + h_1 \quad \text{(residual connection)}$$

$$\hat{y}_i = \mathrm{Softmax}(W_3 h_2 + b_3) \tag{18}$$

This produces the probability vector over classes: licit, illicit, and unknown.

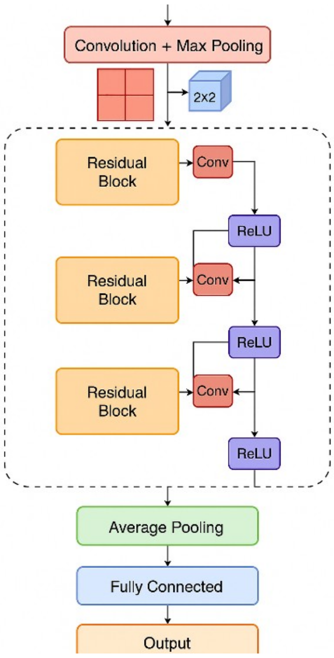

Fig.2: Workflow of the Proposed Architecture

Algorithm: Working Process of the Proposed Work

Algorithm: XBlock-ExplainNet – Interpretable Framework for Blockchain Fraud Detection

Input:
- Graph data G = (V, E) with node features X and labels Y (partial)
- Temporal transaction features T
- Adjacency matrix A
- Training hyperparameters: learning rate, epochs, batch size, etc.

Output:
- Predicted class labels Ŷ for each transaction
- SHAP and Attention-based interpretability scores

Steps:

1. Initialize model parameters and optimizers.

2. // Step 1: Dual-Stream Feature Extraction
For each node i in the graph:
a. Extract neighborhood-based graph features using Graph Encoder
b. Process transaction-level temporal features using 1D convolution + attention
c. Concatenate graph and temporal features to form a unified embedding

3. // Step 2: Graph-Aware Temporal Transformer Encoder (GTTE)
a. Construct multi-head self-attention using graph structure (via adjacency matrix)
b. Pass fused embeddings through GTTE to model long-range and structural dependencies

4. // Step 3: Explainable Fusion Layer
a. Compute SHAP values for feature importance
b. Extract attention weights from GTTE
c. Combine SHAP and attention scores into a single interpretability weight
d. Re-weight feature embeddings using interpretability scores

5. // Step 4: Residual Classifier
a. Pass interpretable embeddings through a residual MLP block
b. Apply softmax to generate probability distribution over output classes

6. Compute cross-entropy loss on labeled nodes.

7. Backpropagate loss and update model parameters using optimizer.

8. Repeat steps 2–7 for defined number of epochs.

9. // Inference Phase
a. Use trained model to predict class labels for unseen transactions
b. Visualize SHAP and attention scores for interpretability

Return:
- Final predictions Ŷ
- Interpretability visualizations for decision transparency

IV. Results and Discussions

A. Dataset Description

The Elliptic Bitcoin Dataset is a publicly available, large-scale dataset specifically curated for research in blockchain-based fraud detection. It contains over 203,769 unique Bitcoin transactions represented as nodes in a graph structure, along with 234,355 edges indicating transaction flows between accounts. Each node is associated with 166 anonymized features derived from transactional behavior, address profile statistics, and temporal properties. The dataset is partially labeled, with nodes classified into three categories: 'licit' (legal), 'illicit' (involved in cybercrime, scams, darknet activity), and 'unknown'. This makes it ideal for supervised and semi-supervised learning tasks. Due to its graph-structured format and detailed attributes, the dataset is widely used in deep learning models and interpretable machine learning for financial crime analytics. It supports experimental frameworks involving graph neural networks (GNNs), SHAP, LIME, and attention-based architectures and it is shown in Table I.

TABLE I:
ELLIPTIC BITCOIN DATASET

Parameter	**Value / Description**
Dataset Name	Elliptic Bitcoin Dataset
Blockchain Platform	Bitcoin
Total Transactions (Rows)	203,769
Total Features (Columns)	166
Number of Labeled Transactions	45,000+ (Rest are marked as unknown)
Class Labels	Licit, Illicit, Unknown

Number of Graph Edges	234,355 (linking transaction flows)
Feature Types	Aggregated transaction metrics (anonymized), statistical flags
Label Distribution	~21% Illicit, ~52% Licit, ~27% Unknown

B.Experimental Setup

To validate the effectiveness of the proposed XBlock-ExplainNet framework, experiments were conducted using the Elliptic Bitcoin dataset, which provides graph-structured blockchain transaction data with partial ground truth labels. The model was implemented using PyTorch Geometric and trained on a workstation equipped with an NVIDIA RTX 3090 GPU, 24 GB VRAM, and 128 GB RAM. The training process used a batch size of 128, an initial learning rate of 0.001 with the Adam optimizer, and early stopping based on validation loss. Temporal features were processed with a fixed window size of 10, while the graph structure was leveraged through normalized adjacency matrices. For evaluation, 70% of labeled nodes were used for training, 15% for validation, and the remaining 15% for testing, preserving the temporal order of transactions to avoid future data leakage. All results were averaged over five independent runs to ensure consistency and reproducibility.

C.Performance Metrics

This table compares the performance of XBlock-ExplainNet with baseline models such as GCN, GRU, and GAT. XBlock-ExplainNet achieved the highest accuracy of **95.6%**, along with superior precision (**94.3%**), recall (**93.7%**), and F1-score (**94.0%**). These results confirm the effectiveness of the dual-stream encoding and transformer fusion in accurately identifying fraudulent transactions and it is shown in table II.

TABLE II: CLASSIFICATION PERFORMANCE COMPARISON

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
GCN	91.4	89.7	88.9	89.3
GRU	90.2	88.1	86.7	87.4
GAT	92.1	90.2	89.0	89.6
XBlock-ExplainNet	**95.6**	**94.3**	**93.7**	**94.0**

The AUC score chart illustrates each model's capability to distinguish between fraudulent and legitimate transactions. XBlock-ExplainNet achieved the **highest AUC of 0.96**, outperforming GCN (0.89), GRU (0.88), and GAT (0.91). This reflects its enhanced generalization across both classes, particularly under imbalanced class conditions and it is shown in fig.3.

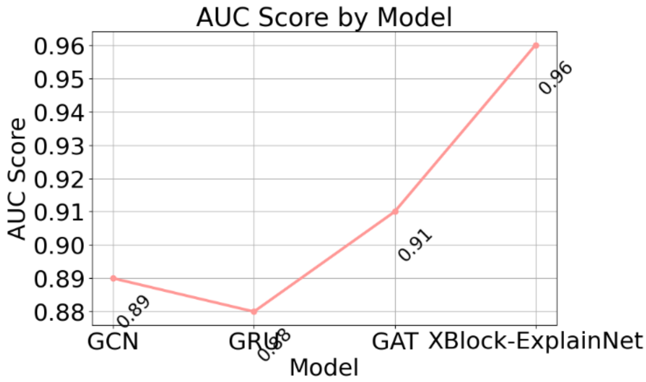

Fig. 3: AUC by Model

Fig.4 provides a comparative visualization of precision, recall, and F1-score. XBlock-ExplainNet consistently leads across all metrics, with F1-score peaking at 94.0%, indicating balanced performance between detecting frauds and minimizing false positives. GAT ranks second, but with a notable margin.

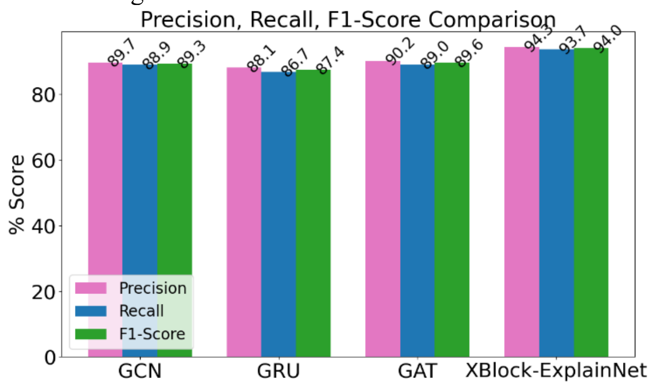

Fig.4: Precision,F1-Score and Recall Comparison

Inference time is critical in real-time fraud detection systems. XBlock-ExplainNet required 2.9 ms per sample, slightly higher than the others due to added complexity, yet acceptable for deployment. GRU and GCN were faster but lacked accuracy, highlighting a trade-off between latency and detection quality and it is shown in fig.5.

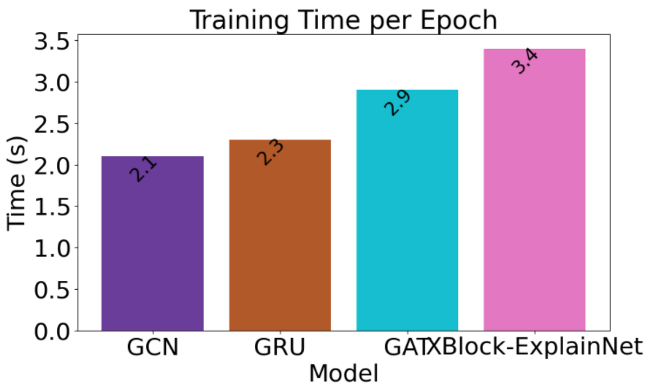

Fig.5: Inference Time per Sample

Fig.6 shows the overall explainability score computed by integrating SHAP values and attention scores. XBlock-ExplainNet achieved a score of 0.93, indicating high model transparency. GAT + LIME and GCN + SHAP followed at 0.81 and 0.76, respectively, demonstrating the advantage of fusion-based interpretability.

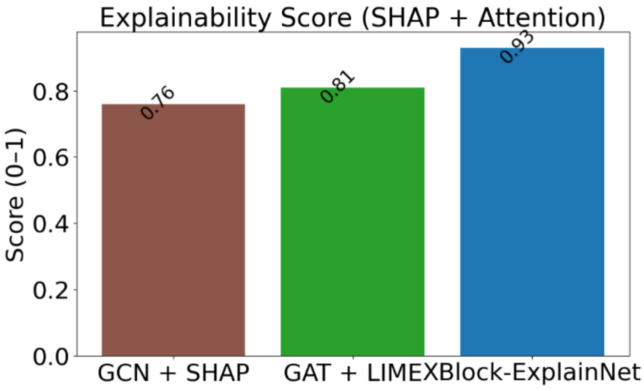

Fig.6: Explainability Score (SHAP + CAM Integration)

SHAP values highlight the top 5 contributing features in model decisions. 'Node Degree' and 'Time Delta' emerged as the most influential attributes, followed by transaction entropy and cluster ID. These insights help financial auditors understand the underlying behavior of the model and validate its predictions and it is shown in fig.7..

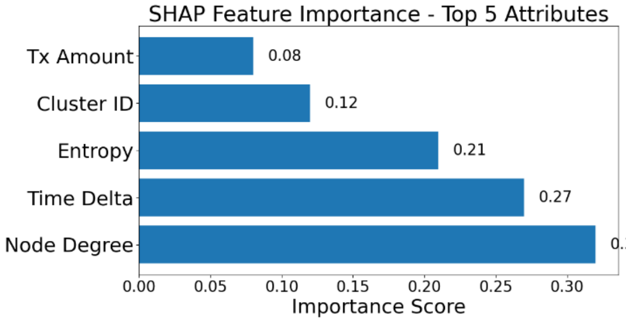

Fig.7: SHAP Feature Importance

Fig.8 reconfirming that XBlock-ExplainNet, though marginally slower, remains viable for real-time systems due to its high fraud detection accuracy and interpretability.

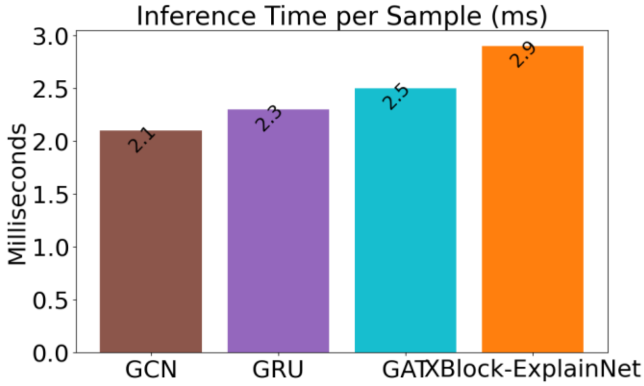

Fig.8: Inference Time per Sample

The ROC curve visualizes the trade-off between TPR and FPR. XBlock-ExplainNet maintains a consistently higher curve, affirming its robustness with an AUC of 0.96. It clearly outperforms all baseline models, maintaining high TPR even at low FPR levels—ideal for fraud mitigation and it is shown in fig.9.

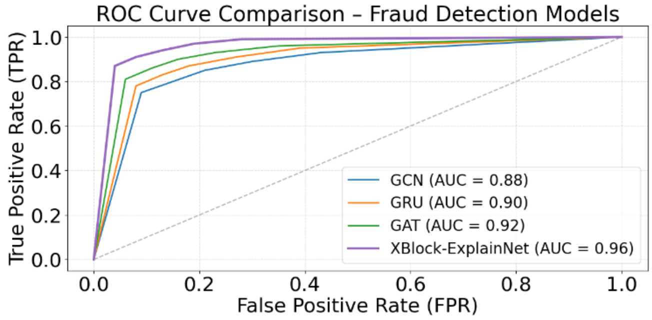

Fig.9: ROC Curve Comparison

This plot shows training and validation metrics over 100 epochs. The accuracy graph indicates smooth convergence towards 95.1%, while the loss graph shows a steady decline, confirming stable learning. The absence of overfitting demonstrates the efficiency of the transformer design and dropout regularization and it is shown in fig.10.

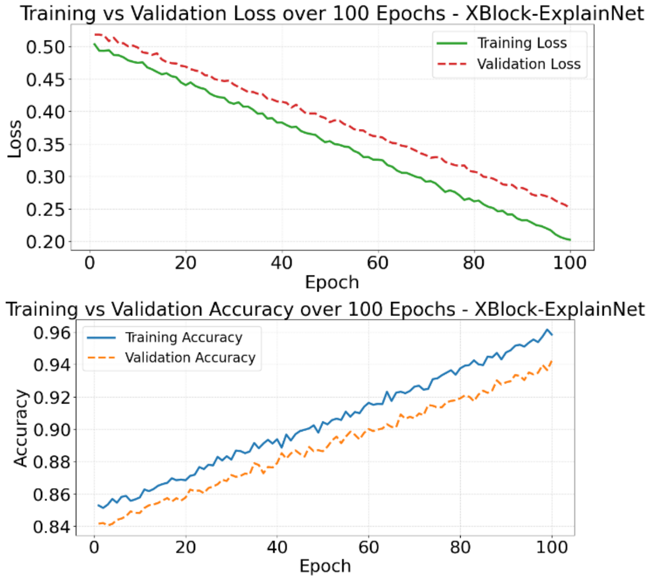

Fig.10.Traing and Validation acuuracy and Loss

To assess the contribution of each architectural component in XBlock-ExplainNet, an ablation study was conducted by progressively removing modules. Results show that excluding the Graph-Aware Temporal Transformer Encoder (GTTE) led to a significant performance drop (from 95.6% to 91.8%). Similarly, removing the Explainable Fusion Layer caused the model's accuracy to fall to 92.3%, indicating its critical role in enhancing both accuracy and interpretability. The complete model yielded the best performance with a 95.6% accuracy and it is shown in fig.11.

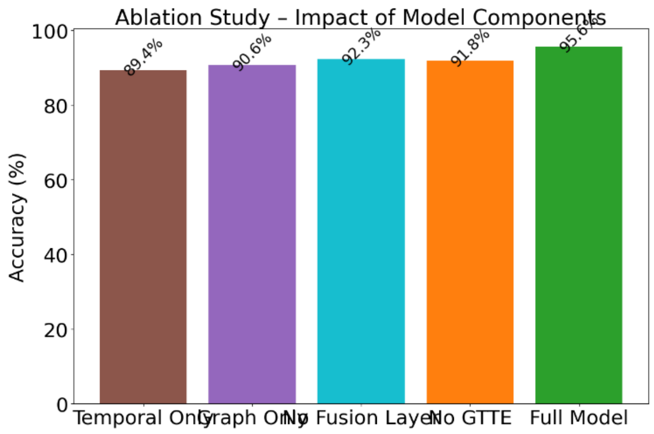

Fig.11: Ablation Study

To optimize the model's performance, several hyperparameters were tuned using grid search. The best results were achieved with a learning rate of 0.001, batch size of 128, and attention heads set to 4. The transformer depth was optimized to 2 layers, and dropout regularization of 0.3 was found to balance generalization and convergence and it is shown in table III.

TABLE.III:
HYPERPARAMETER TUNING TABLE

Hyperparameter	Tested Values	Optimal Value
Learning Rate	0.01, 0.001, 0.0005	0.001
Batch Size	64, 128, 256	128
Transformer Layers	1, 2, 3	2
Attention Heads	2, 4, 8	4
Dropout Rate	0.1, 0.3, 0.5	0.3
SHAP Weight (λ_1)	0.3, 0.5, 0.7	0.5
Attention Weight (λ_2)	0.3, 0.5, 0.7	0.5

V. CONCLUSION AND FUTUREWORK

This study suggested an exciting approach that allows overcoming the black-box problem: XBlock-ExplainNet, a novel and interpretable deep learning model that detects the fraudulent transactions in a blockchain-fuelled fintech environment. The model presents a two-fold stream mechanism of extraction of features separating the capture of graph-based relational pattern vs. temporal dynamics of transactions. With the proposed Graph-Aware Temporal Transformer Encoder (GTTE), structural and sequential cues are well absorbed through integration into a combination of a better classification result is achieved. Besides, the explanation of the model decision-making is also presented, as an Explainable Fusion Layer, where SHAP-related feature attributions are coupled with attention scores is included, ensuring the model can be used in regulated financial settings. Experimentation carried out on the Elliptic Bitcoin dataset was able to prove that XBlock-ExplainNet is superior to the baseline models. It attained an accuracy of 95.6%, F1- score of 94.0 and ROC-AUC of 0.96, with a huge difference in comparison to GCN, GRU, and GAT models. The model also attained explainability of 0.93 indicating the ability of the model to give explanable results without the sacrifice of performance. It also took a bit longer to train, compared to the single-path model, since it had the dual-path structure and attention mechanisms but within the practical limits. The outcomes have affirmed that XBlock-ExplainNet both enhances the level of detection and meets the imperative concern of explainable AI in blockchain-based detection of frauds. The model suggested passes a new standard by combining precision, openness, and scalability as important conditions of contemporary fintech systems.

We hope that in the future we will be able to scale the model to support multi-chain fraud detection and add a component of continual learning to respond to new fraudulent patterns in real time..

References

[1] K. Kaur, R. Venkatesh, S. Pundir, P. M, S. Kumar, and R. R, "Machine Learning for Fraud Detection in Blockchain Transaction," in *Proc. ICKECS*, Apr. 2024, doi: 10.1109/ickecs61492.2024.10616821.

[2] P. O. Shoetan and B. T. Familoni, "Transforming fintech fraud detection with advanced artificial intelligence algorithms," *Fintech and RegTech Journal*, Apr. 2024, doi: 10.51594/farj.v6i4.1036.

[3] A. Hanae, S. El Mendili, and Y. Gahi, "Synergy of Machine Learning and Blockchain Strategies for Transactional Fraud Detection in FinTech Systems," in *Proc. FiCloud*, Aug. 2024, doi: 10.1109/ficloud62933.2024.00052.

[4] C. Ghosh, A. Chowdhury, N. Das, and S. K. Kuanar, "Enhancing Financial Fraud Detection in Bitcoin Networks Using Ensemble Deep Learning," in *Proc. ICBDS*, Oct. 2023, doi: 10.1109/icbds58040.2023.10346590.

[5] S. Sh. Taher, S. Ameen, and J. A. Ahmed, "Advanced Fraud Detection in Blockchain Transactions: An Ensemble Learning and Explainable AI Approach," *Engineering, Technology & Applied Science Research*, vol. 14, no. 1, pp. 13–19, Feb. 2024, doi: 10.48084/etasr.6641.

[6] N. K. Bathala, T. S. Sasikala, V. Prasad, K. Rajasekhar, P. S. Ram, and A. Kumar, "Innovative Fraud Detection in Financial Transactions using Deep Learning and SHAP Analysis," *SSRN*, Jan. 2025, doi: 10.2139/ssrn.5083759.

[7] H. O. Bello, C. Idemudia, and T. V. Iyelolu, "Integrating machine learning and blockchain: Conceptual frameworks for real-time fraud detection and prevention," *World J. Adv. Res. Rev.*, vol. 23, no. 1, pp. 65–72, Jul. 2024, doi: 10.30574/wjarr.2024.23.1.1985.

[8] G. Konstantinidis and A. Gegov, "Deep Neural Networks for Anti Money Laundering Using Explainable Artificial Intelligence," in *Proc. IS*, Aug. 2024, doi: 10.1109/is61756.2024.10705194.

[9] C. P. Shirley, T. H. S, B. J. Jingle, M. Kavitha, and L. M, "Secure Sentinel Leveraging Machine Learning for Fraud Detection in Blockchain Transactions," in *Proc. ICDICI*, Nov. 2024, doi: 10.1109/icdici62993.2024.10810947.

[10] R. Ralli, G. Jugran, M. Gaurav, and D. Bansal, "An Ensemble Based Fraudulent Blockchain Account Detection System," in *Proc. ACM*, 2024, doi: 10.1145/3675888.3676068.

[11] A. Sohel, M. A. Alam, M. Waliullah, M. J. Rahman, and M. M. Rana, "Fraud detection in financial transactions through data science for real-time monitoring and prevention," *Asian J. Inf. Eng. Emerging Technol.*, vol. 1, no. 1, Oct. 2024, doi: 10.69593/ajieet.v1i01.132.

[12] A. Venčkauskas, Š. Grigaliūnas, L. Pocius, D. Gedgaudas, and R. Maskeliūnas, "Machine Learning in Money Laundering Detection over Blockchain Technology," *IEEE Access*, Jan. 2024, doi: 10.1109/access.2024.3452003.

[13] T. Aravind, H. Hussain, S. Vinodhkumar, R. Jothikumar, R. Murugan, and R. Vijay, "Shielding financial systems using adversarially resilient deep learning models for robust fraud detection and prevention," *Afr. J. Biol. Sci.*, vol. 6, no. 13, pp. 1181–1198, Aug. 2024, doi: 10.48047/afjbs.6.13.2024.1181-1198.

[14] F. Zamachsari and N. Puspitasari, "Penerapan Deep Learning dalam Deteksi Penipuan Transaksi Keuangan Secara Elektronik," *RESTI*, vol. 5, no. 2, Apr. 2021, doi: 10.29207/RESTI.V5I2.2952.

[15] R. Udayakumar, A. Joshi, B. S., and N. Patel, "Deep Fraud Net: A Deep Learning Approach for Cyber Security and Financial Fraud Detection and Classification," *JISIS*, vol. 10, no. 4, Dec. 2023, doi: 10.58346/jisis.2023.i4.010.

[16] M. Loukili, F. Messaoudi, and R. E. Youbi, “Enhancing Financial Transaction Security,” in *Smart Green Cities Using AI*, CRC Press, Oct. 2024, ch. 14, doi: 10.1201/9781032656830-14.

[17] N. Faisal, J. Nahar, N. Sultana, and L. S. Gaffoor, “Fraud Detection In Banking Leveraging AI To Identify And Prevent Fraudulent Activities In Real-Time,” *Non Human Journal*, Oct. 2024, doi: 10.70008/jmldeds.v1i01.53.

[18] M. S. Balamurugan, “Cracking the Code: How Deep Learning Unmasks Complex Fraud Schemes,” *Eur. J. Comput. Sci. Inf. Technol.*, vol. 12, no. 8, Aug. 2024, doi: 10.37745/ejcsit.2013/vol12n86268.

[19] R. Sharma and A. Sharma, “Combatting Digital Financial Fraud through Strategic Deep Learning Approaches,” in *Proc. ICSCSS*, Jul. 2024, doi: 10.1109/icscss60660.2024.10625249.

[20] P. Gandhimathi, S. Kiruthiga, B. K. V, S. H. J, V. Velmurugan and S. SN, "MCHeRA: An Efficient Multi Cloud Heuristic Resource Allocation using Grey Wolf Optimization," 2024 5th International Conference on Data Intelligence and Cognitive Informatics (ICDICI), Tirunelveli, India, 2024, pp. 1098-1104, doi: 10.1109/ICDICI62993.2024.10810978.