**Detecting Bitcoin Fraud in Blockchain Transactions Using Hybrid Machine Learning Models**

**1. Introduction**

The rise of cryptocurrencies, particularly Bitcoin, has transformed the financial landscape, offering decentralization, security, and transparency. However, this new digital currency paradigm has also opened avenues for illicit activities, including fraud. Detecting fraudulent transactions in a blockchain network is crucial for preserving the integrity and trust of the cryptocurrency ecosystem. This paper explores the use of machine learning (ML) models, specifically Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN), to identify fraudulent Bitcoin transactions by detecting patterns in transaction data that indicate potential fraud.

**2. Related Work**

Numerous studies have explored the application of machine learning techniques in blockchain and cryptocurrency fraud detection. Traditional approaches, including rule-based systems, anomaly detection, and statistical methods, often fall short due to the complexity and high dimensionality of transaction data. Recent advancements in deep learning, particularly LSTM and CNN models, have shown promising results in anomaly detection. LSTM has proven successful in capturing long-range dependencies and temporal patterns, while CNN excels at feature extraction and pattern recognition in structured data, such as transaction amounts and frequencies. This paper builds on previous works by integrating LSTM and CNN into a unified framework for detecting Bitcoin fraud.

**3. Problem Formulation**

Detecting fraud in Bitcoin transactions is framed as a binary classification problem: given a transaction, classify it as either legitimate or fraudulent. The dataset includes features such as transaction amounts, timestamps, sender/receiver addresses, and transaction metadata. One of the challenges in fraud detection is the imbalance between fraudulent and legitimate transactions. The minority class (fraudulent transactions) can cause model training issues, making it essential to apply techniques such as SMOTE to oversample the minority class and improve model performance.

**4. Contribution**

In this paper, we propose a hybrid machine learning model for Bitcoin fraud detection using two state-of-the-art approaches: LSTM and CNN. Our contributions are as follows:

1. Hybrid Model Architecture: We combine LSTM for time-series modeling and CNN for pattern recognition to leverage their respective strengths.
2. Data Preprocessing: We apply MinMax scaling for feature normalization and use SMOTE for generating synthetic data points to balance the dataset.
3. Ensemble Learning: We combine the predictions from LSTM and CNN models using a weighted ensemble approach to improve classification performance.
4. Evaluation and Comparison: We evaluate the model on a real-world Bitcoin transaction dataset and compare its performance against traditional machine learning methods and standalone models.

**5. Blockchain Transactions**

Blockchain technology underpins Bitcoin, ensuring transparency and immutability. Bitcoin transactions consist of several components:

* Transaction Amount: The value being transferred in the transaction.
* Sender/Receiver Address: The cryptographic addresses of the parties involved.
* Timestamp: The date and time when the transaction occurred.
* Transaction Fees: The cost associated with processing the transaction.
* Inputs and Outputs: The Unspent Transaction Outputs (UTXOs) representing the transaction history.

Fraudulent transactions may exhibit suspicious patterns, such as unusually large transfers, irregular frequencies, or manipulated metadata. Detecting these anomalies requires sophisticated modeling techniques capable of learning from complex and imbalanced data.

**6. Proposed Model**

We propose an ensemble model that combines two deep learning techniques: LSTM and CNN.

* LSTM Model: LSTM is ideal for capturing temporal dependencies in transaction data. Our LSTM model consists of two stacked LSTM layers followed by dense layers for classification.
* CNN Model: The CNN model is used for feature extraction and recognizing patterns in the structured data. It includes a Conv1D layer to process the input sequence, followed by flattening and dense layers for classification.

These two models are trained independently, and their predictions are combined using a weighted averaging approach to provide the final classification decision.

**7. Results and Discussion**

We evaluated the proposed LSTM + CNN hybrid model using a real-world Bitcoin transaction dataset and compared its performance against standalone LSTM, CNN, and traditional machine learning models.

**Model Performance:**

* **Accuracy**: 97.55% on the test set, indicating strong overall performance.
* **Precision & Recall**:
  + **Class 0 (Legitimate Transactions)**:
    - **Precision**: 0.97
    - **Recall**: 0.99
    - **Interpretation**: The model has high precision and recall for legitimate transactions, meaning very few legitimate transactions are misclassified as fraudulent.
  + **Class 1 (Fraudulent Transactions)**:
    - **Precision**: 0.98
    - **Recall**: 0.91
    - **Interpretation**: The model has high precision for fraudulent transactions but slightly lower recall, meaning some fraudulent transactions are missed (false negatives), though it still identifies most frauds effectively.
* **F1-Score**:
  + **Class 0**: 0.98, reflecting a strong balance between precision and recall for legitimate transactions.
  + **Class 1**: 0.94, showing good performance in identifying fraudulent transactions despite some false negatives.
* **Overall Evaluation**:
  + **Accuracy**: 98.0% for the entire dataset (2000 samples).
  + **Macro Average**:
    - **Precision**: 0.98
    - **Recall**: 0.95
    - **F1-Score**: 0.96
  + **Weighted Average**:
    - **Precision**: 0.98
    - **Recall**: 0.98
    - **F1-Score**: 0.98

These results indicate that the hybrid model significantly outperforms standalone models in terms of both accuracy and F1-score, especially for the fraudulent transaction class (label 1). The weighted averaging approach for combining LSTM and CNN predictions further boosts performance, particularly in handling the imbalanced dataset.

A blue squares with white text

AI-generated content may be incorrect.

**ROC CURVE**

A line graph with blue and orange lines

AI-generated content may be incorrect.

**Conclusion**

This paper presents an effective hybrid machine learning model for detecting fraudulent Bitcoin transactions using LSTM and CNN. The results indicate that this combined approach is highly successful in both accuracy and class balancing, offering a robust solution for real-time fraud detection in blockchain networks. Future work could explore the integration of additional features, such as network-level data, or even other advanced machine learning techniques to further enhance fraud detection capabilities.