Final Project in NT522 course: Machine Learning in Cybersecurity

FraudGNN-RL: A Graph Neural Network with Reinforcement Learning for Adaptive Financial Fraud Detection

Group members: *Trung Do-Quang (23521673), Khang Nguyen-Dinh (23520694), Phuoc Hoang-Bao(23521231)*

Group ID: G16, Project ID: CK20

Introduction

Financial fraud has become a pressing global issue, with estimated annual losses surpassing $5.127 trillion worldwide. This research tackles key challenges in financial fraud detection—namely adaptivity, context-awareness, and privacy preservation—by proposing and evaluating a novel method called FraudGNN-RL. The method is built upon the synergistic integration of Graph Neural Networks (GNNs), Reinforcement Learning (RL), and Federated Learning (FL). Each component is strategically chosen to address specific limitations of traditional methods while leveraging the complex and distributed nature of financial transaction data.

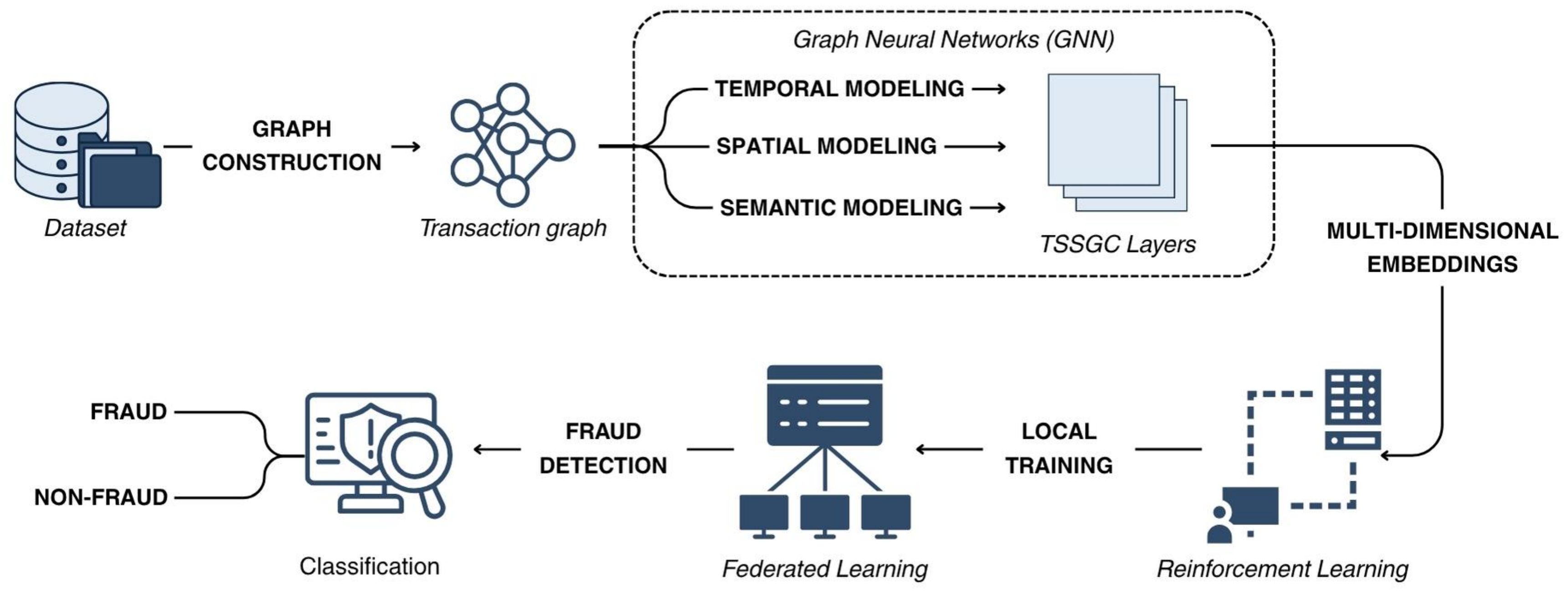

Fig.1 *Our method in project*

Methodology

FraudGNN-RL addresses adaptability and privacy through three components:

- Temporal-Spatial-Semantic Graph Convolution (TSSGC): A GNN-based architecture capturing temporal dynamics, spatial relationships, and semantic information of financial transactions.

- Reinforcement Learning (RL): Employs a Deep Q-Network (DQN) with Normalized Advantage Functions (NAF) to learn an optimal policy, where:

- **State s_t**: The current graph embedding produced by TSSGC layers.
- **Action a_t** : The fraud detection threshold and feature importance weights.
- **Reward r_t** : A combination of detection accuracy and false positive rate.

This enables dynamic adaptation to evolving fraud patterns using a replay buffer and target network.

- Federated Learning (FL): Utilizes FedAvg for privacy-preserving collaborative training across institutions, ensuring data security while enhancing model robustness.

Experiments and Results

Experiments were conducted on the Credit Card Fraud 2023 dataset (550,000 transactions, fraud rate of 0.172%), which exhibits significant class **imbalance** with only 0.172% fraudulent transactions. Model performance was evaluated using F1-score, ROC-AUC, ROC-PR, and Recall to address this imbalance.

Two scenarios were analyzed:

- Evaluation with Related Methods: Compared FraudGNN-RL against traditional methods (e.g., GCN).

- Evaluation on GNN Models: Compared FraudGNN-RL with other GNN variants (e.g., GAT, GraphSAGE).

All results are summarized in the ***Table 2.***

Dataset	Transactions	Features	Fraud Rate	Time Span	Key Characteristics
Credit Card 2023	> 550,000	31	Imbalanced	2023	Anonymized features, temporal patterns

Table 1. *Details of Credit Card Fraud Detection Dataset*

Method	Credit Card 2023			
	F1	AUC-ROC	AUC-PR	Recall@5%
GAT	0.8805	0.9701	0.8621	0.8948
GCN	**0.8957**	**0.9744**	**0.8720**	**0.9151**

Table 2. *Performance Comparison of GNN Methods on Credit Card Fraud 2023 Dataset*

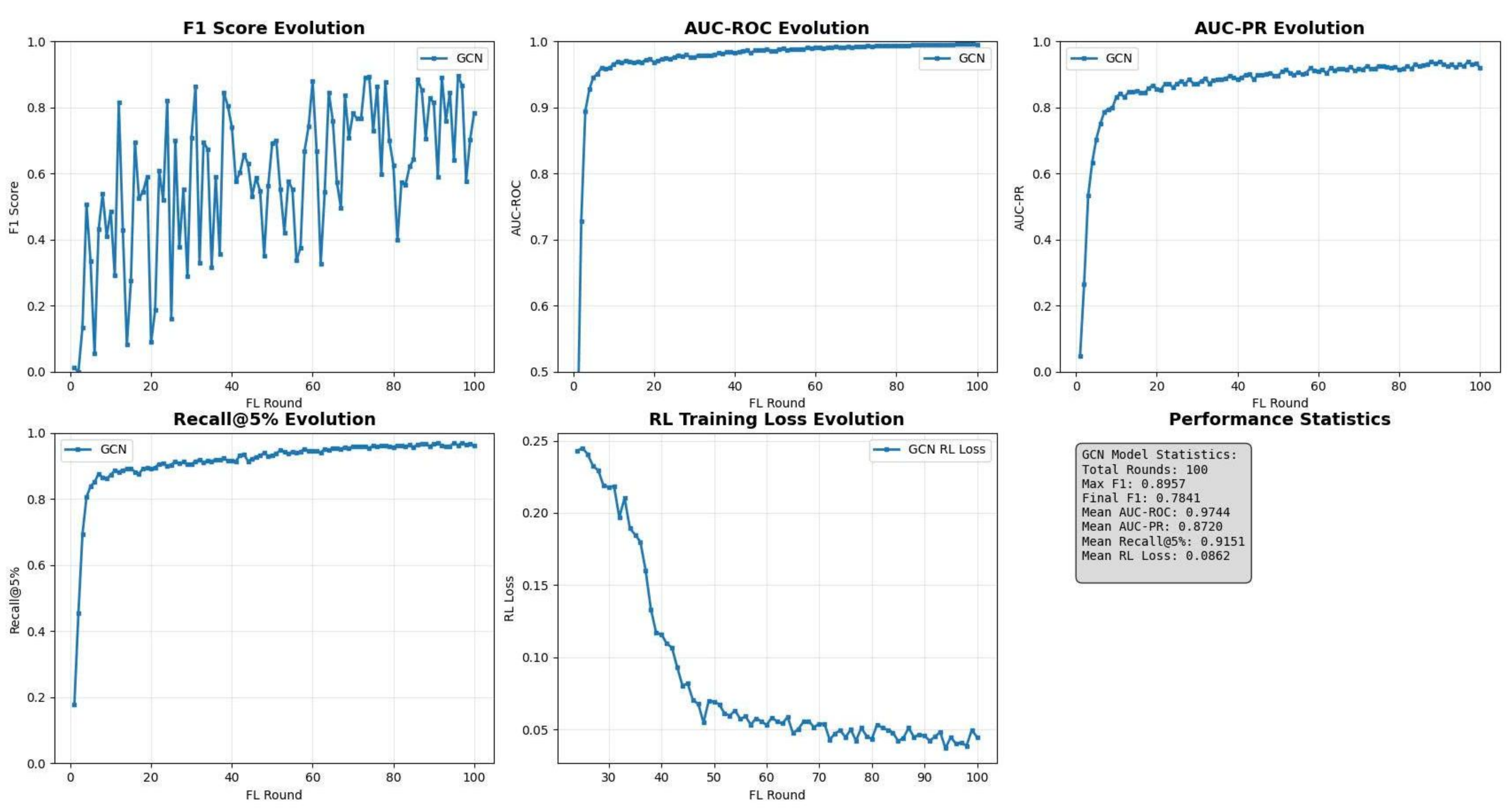

Fig.2 *GCN Model Performance Analysis – Credit Card Fraud Detection*

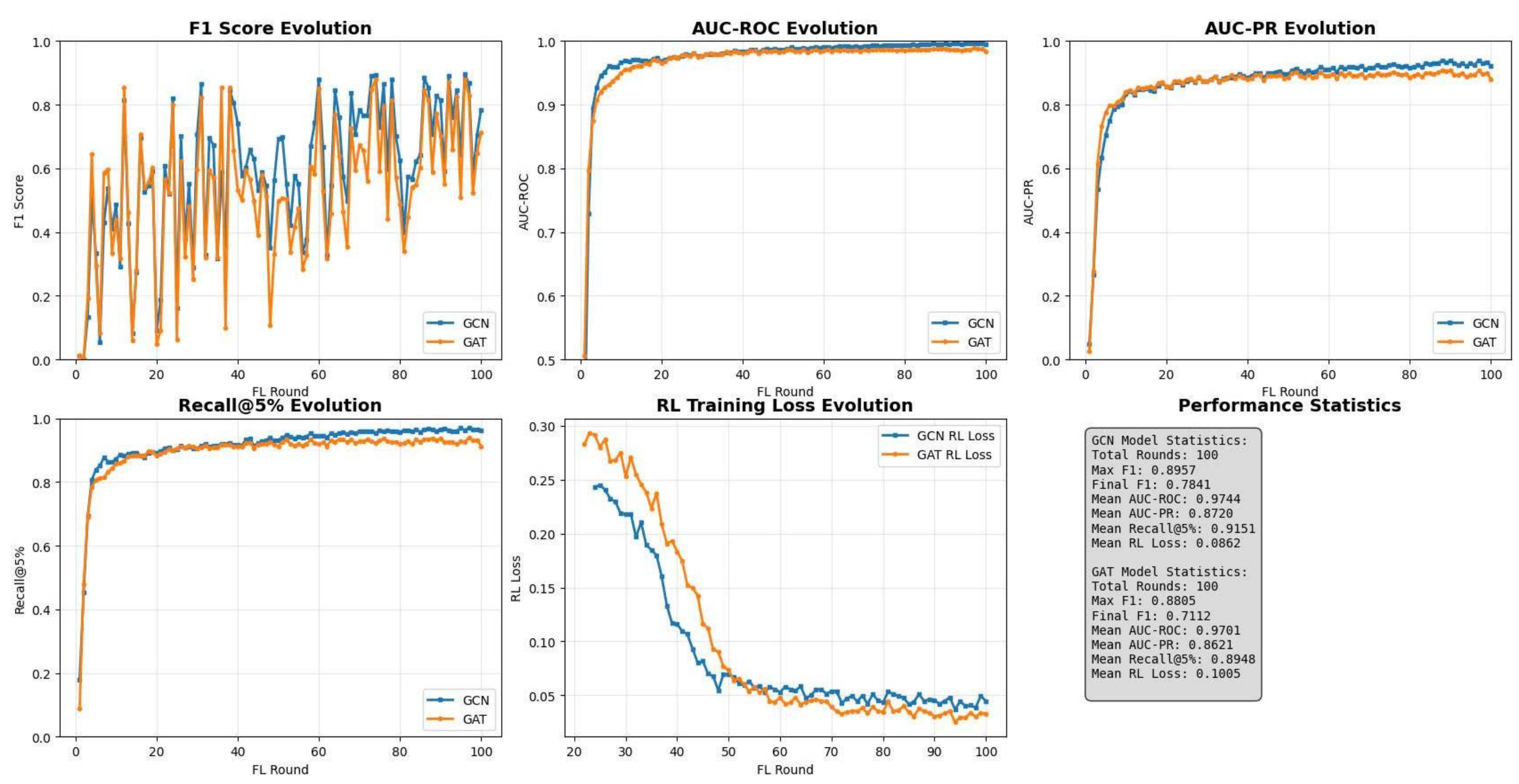

Fig.3 *GCN vs GAT Model Performance Analysis – Credit Card Fraud Detection*