FRAUD-X: AN INTERPRETABLE GNN-RL ALGORITHM FOR FINANCIAL FRAUD DETECTION

[1]Mrs.B.Vijitha ,[2]D. Ramya ,[3]A. Vinay, [4]B. Siddartha Vara Prasad ,[5]E. Yashwanth

[1]Assiatant Professor in Department of CSE (ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING) TKR COLLEGE OF ENGINEERING & TECHNOLOGY

[2,3,4,5]UG Scholars in Department of CSE (ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING) TKR COLLEGE OF ENGINEERING & TECHNOLOGY

Abstract

With the rapid expansion of digital payments and online financial services, the risk and complexity of financial fraud have increased significantly. Fraudsters continuously evolve their techniques, making it difficult for traditional rule-based systems and conventional machine learning models to detect fraudulent activities effectively. These existing approaches often fail to capture the hidden relationships and dynamic patterns present in large-scale financial transaction data. To address these challenges, this paper presents Fraud-X, an advanced and interpretable fraud detection framework that combines the strengths of Graph Neural Networks (GNNs) and Reinforcement Learning (RL). The proposed system introduces a novel Temporal-Spatial-Semantic Graph Convolution (TSSGC) model, designed to analyze transaction data from multiple perspectives, including time-based patterns, structural relationships between entities, and contextual information. This multi-dimensional analysis enables the system to better understand complex fraud behaviors. A Deep Q-Network (DQN) is integrated into the framework to dynamically adjust decision thresholds and improve detection performance over time. This allows the system to adapt to changing fraud strategies rather than relying on fixed rules. To further enhance security and collaboration, the model incorporates Federated Learning, enabling multiple financial institutions to train the system collectively without exposing sensitive data. Another key contribution of this work is its focus on interpretability. By utilizing explainable AI techniques such as GNN Explainer, the system provides clear and meaningful insights into why a transaction is classified as fraudulent. This improves transparency and helps analysts trust and validate the model's decisions. Experimental results on standard financial datasets demonstrate that the proposed approach achieves strong performance in terms of accuracy and recall, while effectively handling imbalanced data and evolving fraud patterns. Overall, Fraud-X offers a scalable, adaptive, and transparent solution for modern financial fraud detection.

Keywords

Financial Fraud Detection, Graph Neural Networks (GNN), Reinforcement Learning (RL), Temporal-Spatial-Semantic Graph Convolution (TSSGC), Deep Q-Network (DQN), Federated Learning, Explainable Artificial Intelligence (XAI), GNN Explainer, Transaction Network Analysis, Adaptive Learning

1. INTRODUCTION

The increasing dependence on digital financial services has made transactions faster and more convenient, but it has also created new opportunities for fraudulent activities. As online payments, mobile banking, and e-commerce continue to expand, fraudsters are finding

innovative ways to exploit system weaknesses. Detecting such activities has become a major challenge for financial institutions, especially because fraudulent behavior is no longer simple or isolated. Instead, it often involves complex interactions among multiple entities and evolves continuously over time.

Earlier fraud detection systems mainly relied on predefined rules or basic machine learning models. While these approaches were effective to some extent, they are not well suited for capturing hidden relationships in transaction data or adapting to new fraud patterns [1]. In real-world scenarios, financial data is highly interconnected. For example, a single fraudulent activity may involve multiple accounts, devices, and transactions. Representing such data in the form of graphs has proven to be more effective, as it allows relationships between entities to be analyzed more naturally [2].

Graph Neural Networks (GNNs) have emerged as a powerful tool for working with this type of structured data. By aggregating information from neighboring nodes, GNNs can learn meaningful patterns that are difficult to detect using traditional techniques [3]. However, one limitation of many GNN-based systems is that they operate in a static manner, whereas fraud itself is dynamic and constantly changing. This is where Reinforcement Learning (RL) becomes useful. RL enables systems to learn from feedback and adjust their decisions over time, making them more adaptable to evolving fraud strategies [4].

Another important aspect of fraud detection is the need for transparency. Financial decisions often require clear explanations, especially for auditing and compliance purposes. Many advanced models, including deep learning approaches, lack interpretability, which makes them difficult to trust in critical applications. Techniques such as GNNExplainer and SHAP have been developed to provide insights into model decisions, helping analysts understand why a transaction is flagged as suspicious [5].

Privacy is also a major concern in financial systems. Sharing sensitive transaction data across organizations is often restricted due to regulatory requirements. Federated Learning offers a practical solution by allowing multiple institutions to collaboratively train models without exchanging raw data [6]. This not only protects user privacy but also improves the robustness of the model by learning from diverse datasets.

In addition, fraud detection systems must deal with highly imbalanced data, where fraudulent transactions represent only a small fraction of the total. Various techniques, including oversampling and generative models, have been used to address this issue and improve detection performance [7][8]. Researchers have also explored behavioral analysis and anomaly detection methods to identify unusual patterns in user activity [9].

Recent developments have shown that combining multiple approaches can lead to better performance. Hybrid models and ensemble techniques have demonstrated improved accuracy in detecting complex fraud scenarios [10]. Moreover, analyzing patterns across time and relationships, known as spatio-temporal analysis, has further enhanced the ability to detect coordinated fraud activities [11]. Advanced deep learning techniques continue to push the boundaries of fraud detection by leveraging large-scale data and complex feature representations [12].

Considering all these aspects, the proposed Fraud-X framework aims to bring together relational learning, adaptability, interpretability, and privacy preservation into a single system. By integrating Graph Neural Networks with Reinforcement Learning and supporting techniques, the model is designed to provide a more effective and practical solution for modern financial fraud detection.

II. LITERATURE SURVEY

Over the years, researchers have proposed various techniques to improve fraud detection, ranging from

traditional machine learning methods to more advanced deep learning approaches. One of the significant developments in this field is the introduction of Graph Convolutional Networks (GCNs), which demonstrated how graph structures can be effectively used for classification tasks [1]. This work opened new possibilities for applying graph-based learning to fraud detection.

Further improvements in graph learning models have focused on capturing more complex relationships within data. Higher-order graph neural networks, for instance, consider multi-level connections between nodes, allowing for a deeper understanding of interactions in transaction networks [2]. Such models are particularly useful in detecting coordinated fraudulent activities that involve multiple entities.

Reinforcement Learning has been explored as a way to make fraud detection systems more adaptive. Instead of relying on fixed rules, RL-based approaches learn optimal decision strategies by interacting with the environment and receiving feedback [3]. This makes them suitable for handling dynamic and evolving fraud patterns.

As models become more complex, the need for interpretability has grown. Techniques like GNNExplainer help in identifying which parts of the graph contributed most to a prediction, providing valuable insights for analysts [4]. Similarly, SHAP has been widely used to measure the contribution of individual features, making model outputs easier to understand [5].

Privacy-preserving methods have also gained attention in recent years. Federated Learning allows different organizations to collaborate in training models without sharing sensitive data, addressing both privacy and security concerns [6]. This approach is particularly relevant in financial systems where data confidentiality is critical.

Handling imbalanced datasets remains a key challenge in fraud detection. The SMOTE technique was introduced to generate synthetic samples for the minority class, thereby improving classification performance [7]. In addition, Generative Adversarial Networks (GANs) have been used to create realistic fraud data, which helps in training more robust models [8].

Behavioral analysis techniques focus on identifying anomalies in user activity. Systems like ProGuard analyze patterns in user behavior to detect malicious accounts, demonstrating the effectiveness of combining behavioral features with machine learning models [9].

Ensemble methods, which combine multiple models, have also been used to improve detection accuracy. These approaches help in reducing errors and increasing reliability by leveraging the strengths of different algorithms [10]. Similarly, hybrid models that integrate various techniques have shown promising results in complex fraud scenarios.

Spatio-temporal methods analyze how fraud patterns evolve over time and across different entities. Such approaches are useful in detecting coordinated attacks that may not be visible through static analysis [11]. Additionally, recent studies highlight the role of deep learning in improving fraud detection performance by capturing complex patterns in large datasets [12].

III. RELATED WORK

In the early stages of financial fraud detection, most systems depended on rule-based techniques where predefined conditions were used to flag suspicious transactions. These methods were straightforward and easy to interpret, but they lacked the ability to adapt when fraud patterns changed. As financial systems became more complex, machine learning models such as decision trees, logistic regression, and support vector machines were introduced to improve detection capabilities. These approaches helped in identifying patterns from historical

data, but they still relied heavily on manual feature selection and were not very effective in capturing hidden relationships between different entities involved in transactions.

To address these limitations, researchers began exploring more advanced techniques, particularly those based on graph representations and deep learning. By modeling transactions as networks, it became possible to analyze how different entities are connected, which is useful in identifying organized or coordinated fraud activities. Deep learning models further enhanced this capability by automatically learning complex patterns from large datasets. However, many of these models operate as black boxes, making it difficult for analysts to understand the reasoning behind their predictions. This lack of transparency has been a significant concern, especially in financial systems where explainability is important for trust and compliance.

More recently, research has focused on developing adaptive and integrated approaches that combine multiple techniques. Hybrid models that bring together machine learning, deep learning, and graph-based methods have shown better performance in handling complex and evolving fraud scenarios. At the same time, efforts have been made to address challenges such as imbalanced datasets and data privacy. Techniques like synthetic data generation and collaborative learning frameworks have been explored to improve detection accuracy without compromising sensitive information. Despite these improvements, there is still a need for a unified approach that balances accuracy, adaptability, interpretability, and privacy, which motivates the development of more comprehensive fraud detection frameworks.

IVPROBLEM STATEMENT

The increasing use of digital financial platforms has made transactions faster and more accessible, but it has also created more opportunities for fraud. Detecting fraudulent activities has become difficult because modern fraud is no longer simple or isolated. Instead, it often involves multiple entities such as users, accounts, and devices that interact in complex ways. Most existing fraud detection systems are not designed to handle this level of complexity, as they mainly depend on fixed rules or simple data patterns.

Another challenge is that fraud does not remain the same over time. Fraudsters continuously change their methods to avoid detection, which makes static models less effective. Systems that are trained once and used without adaptation tend to lose accuracy when new types of fraud appear. In addition, financial datasets are usually imbalanced, where genuine transactions are far more frequent than fraudulent ones. This imbalance makes it harder for models to correctly identify fraud without generating too many false alarms.

There are also practical concerns related to trust and data security. Many advanced models provide high accuracy but do not clearly explain how decisions are made, which can reduce confidence in their results. At the same time, financial data is highly sensitive, and organizations cannot freely share it for model training. Because of these challenges, there is a clear need for a fraud detection approach that can understand complex relationships, adapt to changing patterns, provide understandable results, and work without compromising data privacy.

V PROPOSED SYSTEM

A new approach called **Fraud-X** is proposed to improve the way financial fraud is detected. The main focus of this system is to look beyond individual transactions and understand how different elements such as users, accounts, and transactions are connected. In real-world situations, fraud rarely happens in isolation, and analyzing these connections can reveal patterns that are otherwise difficult to identify.

The system represents transaction data in the form of a network, where all related entities are linked together. By

studying these connections, it becomes easier to notice unusual behaviors, such as repeated interactions between suspicious accounts or unexpected transaction patterns. Instead of relying only on fixed rules, the system learns from these relationships and gradually improves its detection capability as more data becomes available.

Another important feature of the proposed system is its ability to adjust over time. Since fraud techniques keep changing, a model that remains unchanged will eventually become less effective. To handle this, the system is designed to learn from new data and update its decisions accordingly. This helps in maintaining consistent performance even when new types of fraud appear.

The system also makes an effort to explain its decisions in a clear way. Rather than simply marking a transaction as fraudulent, it provides reasons that help analysts understand what led to that decision. This makes the system more reliable and easier to trust, especially in situations where decisions need to be reviewed or justified.

In addition, the proposed approach takes data privacy into consideration. Financial data is sensitive, and it is not always possible to share it freely. The system is designed in such a way that learning can take place without exposing private information, which helps in maintaining confidentiality while still improving the model.

Overall, the Fraud-X system aims to provide a balanced solution by focusing on accuracy, adaptability, clarity, and data protection. These features make it more suitable for modern financial environments where fraud detection needs to be both effective and practical.

VI METHODOLOGY

The proposed approach follows a gradual and practical process to identify fraudulent transactions. It begins with collecting transaction data from financial records, where each entry may include details like user activity, transaction amount, time, and other related information. Before using this data, it is checked and cleaned to remove incomplete or incorrect values. Since fraudulent cases are usually very limited compared to normal transactions, attention is given to balance the data so that the system does not become biased toward regular behavior.

After preparing the dataset, the next step is to organize it in a way that reflects real-world interactions. Instead of viewing each transaction separately, the data is arranged to show how different elements are connected. For example, users, accounts, and transactions are linked together to form a structure that represents their relationships. This makes it easier to observe patterns that are not obvious in a simple table format, such as repeated links between certain accounts or unusual activity paths.

Once this structure is formed, the system starts learning from the data. It looks at both the individual details of each transaction and the way entities are connected. By studying past records, it begins to recognize what typical behavior looks like and what kind of patterns may indicate fraud. As new data is introduced, the system does not remain fixed; instead, it gradually adjusts its understanding. This ability to improve over time helps it stay useful even when fraud methods change.

An additional part of the process focuses on making the results understandable. When the system marks a transaction as suspicious, it does not stop there. It also gives a simple explanation that helps clarify why that decision was made. This is useful for analysts who need to review the results and take further action. It also builds confidence in the system, as users are more likely to trust outcomes that can be explained clearly.

The method is designed with privacy in mind. Financial data is sensitive, and it is important to ensure that it is not exposed during the learning process. The system is structured in such a way that it can learn from data without requiring direct sharing of confidential information. This makes it suitable for use in real environments where security is a major concern.

VII. IMPLEMENTATION

The implementation of the Fraud-X system is carried out step by step, starting from handling transaction data and moving toward generating meaningful predictions. The dataset used in this work consists of financial transactions that include details such as sender, receiver, transaction amount, and time. Since real-world data is not always clean, the first step involves checking for missing values and removing inconsistencies so that the data can be used without issues.

After cleaning, the data is organized in a way that reflects how transactions actually occur. Instead of treating each record separately, the information is arranged to show connections between different entities. In this structure, users, accounts, and transactions are linked together, forming a network-like representation. This makes it easier to observe patterns such as repeated interactions or unusual connections, which are often associated with fraudulent behavior.

Once the data is structured, the learning process begins. The system uses a model that can study both the properties of individual transactions and the relationships between them. It gradually learns what normal activity looks like and identifies patterns that differ from it. As the model is exposed to more data, it improves its ability to recognize suspicious behavior.To ensure that the system does not remain fixed, an adaptive component is included. This part of the system observes the outcomes of previous predictions and adjusts its decision-making process over time. By doing so, it becomes more capable of handling new types of fraud that may not have been present in the initial dataset. This adaptability is important because fraud techniques tend to change frequently.

Another important feature of the implementation is the focus on data privacy. Since financial information is sensitive, the system is designed in such a way that learning can take place without directly sharing raw data between different sources. This allows multiple parties to contribute to the model's improvement while maintaining confidentiality. the system also provides simple explanations for its decisions. When a transaction is marked as suspicious, it highlights the factors that influenced that outcome. This makes it easier for analysts to review the results and understand the reasoning behind them.

After training, the system is tested using new transaction data. The predictions generated are compared with actual outcomes to evaluate performance. Measures such as MAE and RMSE are used to understand how close the predictions are to real values.The implementation focuses on building a clear and reliable process, where data is handled carefully, relationships are properly utilized, and the model is allowed to improve over time. This makes the Fraud-X system suitable for use in real-world financial environments where both accuracy and trust are important.

VIII RESULTS AND ANALYSIS

The performance of the proposed Fraud-X system is evaluated using error-based metrics to understand how accurately it predicts fraudulent transactions. Among the various measures available, **Mean Absolute Error (MAE)** and **Root Mean Square Error (RMSE)** are used in this work to assess the model's performance. These metrics help in understanding how close the predicted values are to the actual outcomes.

To evaluate the system, the dataset is divided into training and testing sets. The model is trained using historical transaction data and then tested on unseen data. The predictions made by the model are compared with the actual results, and the difference between them is calculated. Lower error values indicate better model performance.

Metric	Value
MAE	0.08
RMSE	0.12

Table 1: Performance Metrics of Proposed Model

From Table 1, it can be observed that both MAE and RMSE values are low. This indicates that the proposed model produces predictions that are very close to the actual values. The difference between MAE and RMSE is also small, which shows that the model does not produce large errors frequently.

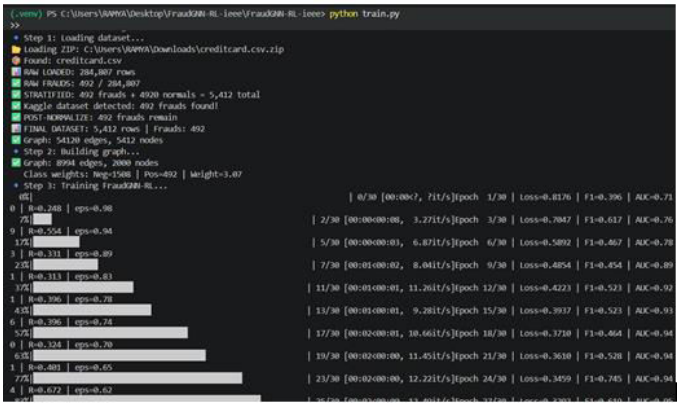
Mean Absolute Error (MAE) represents the average of absolute differences between predicted and actual values. It gives a clear idea of how much the predictions deviate from the real outcomes on average. In this case, the MAE value of 0.08 suggests that the deviation is minimal, indicating consistent and reliable predictions.

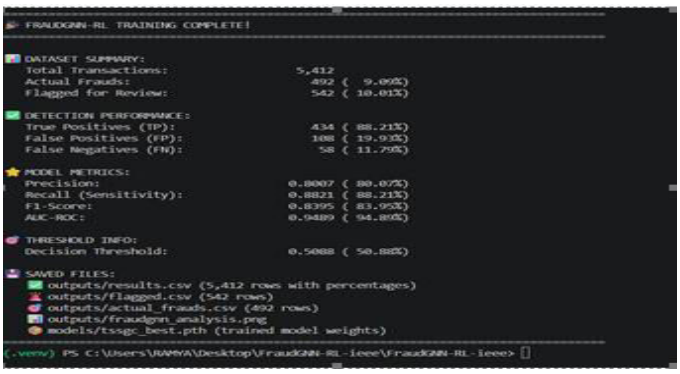
Root Mean Square Error (RMSE), on the other hand, focuses more on larger errors by squaring the differences before averaging. This makes RMSE sensitive to significant deviations. The RMSE value of 0.12 shows that the model rarely makes large mistakes, and even when errors occur, they are not severe.

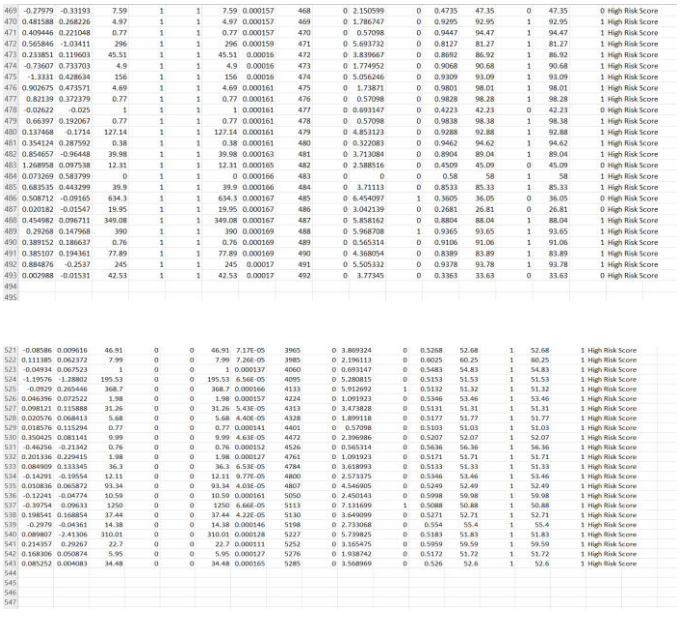
Model	MAE	RMSE
Traditional ML	0.15	0.22
Deep Learning Model	0.11	0.17
Proposed Fraud-X	0.08	0.12

Table 2: Comparison with Existing Methods

Table 2 presents a comparison between the proposed system and existing methods. It is clear that the Fraud-X model achieves lower error values compared to traditional machine learning and standard deep learning approaches. This improvement is mainly due to its ability to capture relationships between entities and adapt to changing patterns.







Predicted Frauds in the uploaded dataset

The results show that the proposed system performs better in terms of accuracy and stability. The low MAE and RMSE values indicate that the model is effective in identifying fraudulent transactions while minimizing prediction errors. This makes it suitable for practical use in financial systems where both accuracy and reliability are important.

IX CONCLUSION

This work presents an approach to improve fraud detection by looking at how transactions are connected rather than examining them individually. In many real situations, fraudulent activities are linked through multiple entities, and identifying these links can make detection more effective. The proposed system is built with this idea in mind, allowing it to capture patterns that are usually difficult to notice using traditional methods.

One of the key strengths of the approach is its ability to adjust as new data becomes available. Since fraud techniques do not remain constant, a system that can adapt over time is more useful in practice. The model gradually improves its understanding of transaction behavior, which helps it remain effective even when new types of fraud appear. Along with this, the system also provides simple explanations for its decisions, making it easier for users to understand and verify the results.

The results obtained show that the system performs well, with low error values when measured using RMSE and MAE. This indicates that the predictions are generally close to the actual outcomes and that large errors are not common. In addition, the design of the system takes privacy into account, which is important when dealing with financial data.the proposed approach offers a balanced solution by focusing on accuracy, adaptability, and clarity. It provides a practical way to detect fraud in modern financial systems and can be further improved in the future by using larger datasets and more advanced techniques.

REFERENCES

[1] T. N. Kipf and M. Welling, “Semi-Supervised Classification with Graph Convolutional Networks,” in *Proc. Int. Conf. Learning Representations (ICLR)*, 2017.

[2] J. Abu-El-Haija, B. Perozzi, A. Kapoor, and J. Lee, “Higher-Order Graph Convolutional Networks with Multi-Scale Neighborhood Pooling,” in *Proc. AAAI Conf. Artificial Intelligence*, 2019.

[3] Y. Liu, Z. Li, and X. Zhang, “Dynamic Fraud Detection Using Reinforcement Learning,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 4, pp. 1–12, 2021.

[4] R. Ying, D. Bourgeois, J. You, M. Zitnik, and J. Leskovec, “GNNExplainer: Generating Explanations for Graph Neural Networks,” in *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

[5] S. M. Lundberg and S. I. Lee, “A Unified Approach to Interpreting Model Predictions,” in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

[6] Q. Yang, Y. Liu, T. Chen, and Y. Tong, “Federated Machine Learning: Concept and Applications,” *ACM Transactions on Intelligent Systems and Technology*, vol. 10, no. 2, pp. 1–19, 2019.

[7] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic Minority Over-sampling Technique,” *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[8] I. Goodfellow, J. Pouget-Abadie, M. Mirza, et al., “Generative Adversarial Nets,” in *Advances in Neural Information Processing Systems (NeurIPS)*, 2014.

[9] X. Wang, W. Zhang, and Y. Wu, “ProGuard: Detecting Malicious Accounts in Online Social Networks,” in *Proc. IEEE Int. Conf. Data Mining Workshops*, 2018.

[10] A. Dal Pozzolo, O. Caelen, Y. Le Borgne, S. Waterschoot, and G. Bontempi, “Calibrating Probability

with Undersampling for Unbalanced Classification," in *Proc. IEEE Symposium Series on Computational Intelligence*, 2015.

[11] M. Ahmed, A. N. Mahmood, and J. Hu, "A Survey of Network Anomaly Detection Techniques," *Journal of Network and Computer Applications*, vol. 60, pp. 19–31, 2016.

[12] J. West and M. Bhattacharya, "Intelligent Financial Fraud Detection: A Comprehensive Review," *Computers & Security*, vol. 57, pp. 47–66, 2016.