Enhanced Detection of UPI Frauds using Advanced Machine Learning Techniques for Secure Digital Transactions

Mangali Vijaya Sree[1], Panyam Bhuvaneshwari[2], Karne Nandini[3], Koppula Ratna Phoebe Amulya[4], Palla Swetha[5] and B. Chandrakala[6]
{ vijayasreemangali@gmail.com[1], panyambhuvana5@gmail.com[2], karnenandini897@gmail.com[3], phoebeamulya12@gmail.com[4], pallaswetha787@gmail.com[5], bchandrakala0116@gmail.com[6] }

4th Year, B. Tech, Department of Computer Science and Engineering, G. Pullaiah College of Engineering and Technology (Autonomous), Kurnool, Andhra Pradesh, India[1, 2, 3, 4, 5]
Assistant Professor, Department of Computer Science and Engineering, G. Pullaiah College of Engineering and Technology (Autonomous), Kurnool, Andhra Pradesh, India[6]

Abstract. The exponential rise of digital transactions through Unified Payments Interface (UPI) kind of platforms, fraud detection has become very important but challenging. Most of the existing systems need to rely on correlation-based models and depend on centralized data aggregation that are limited with their scalability, interpretability and privacy. We propose CauFedFormer, a hybrid approach that combines sequential transformer models, UPI causal inference techniques, and federated learning for UPI fraud detection. Temporal behaviour anomaly capture, causally relevant features recovery, and distributed training with privacy preservation are present in the model. Experimental results show that CauFedFormer can reach precision of 86%, recall of 82%, F1-score of 84%, ROC-AUC of 93% and outperforms traditional baselines, including logistic regression and standalone transformer models. In addition, CauFedFormer proposes interpretable fraud risk score with confidentiality constraint. It thus makes for a promising candidate to be deployed in secure and scalable digital transaction ecosystems.

Keywords: UPI Fraud Detection, Causal Inference, Federated Learning, Transformer Networks, Secure Digital Transactions, Explainable AI

1 Introduction

The spectacular phenomenon of digital payment systems, bringing to the future with unmatched speed of execution, convenience and accessibility, has been rapidly proliferating since the turn of the century [1][2]. Out of these, the most popular platform seems to be the Unified Payments Interface (UPI) in regions like India and somewhere billons of peer to peer and merchant transaction takes place on a monthly basis [3][4]. But this increased level of digital financial activity has also added up with fraudulent activities that include phishing and social engineering attacks to synthetic identity fraud [5][6]. The traditional fraud detection systems, which basically utilize static rule engines or correlation-driven machine learning models, are not able to detect sophisticated patterns of frauds in a real time manner [7][8]. Approaches to centralized data processing to deal with privacy and regulations concerns that are emerging to protect the user privacy [9] [10].

ICITSM-Part I 2025, April 28-29, Tiruchengode, India

DOI 10.4108/eai.28-4-2025.2357830

Detecting fraud in the era of modern digital transactions is a challenge that comes with a few key challenges [11] [12]. First, it must be able to capture temporality and sequential movement of user behaviour that are often present in the form of slight time-based deviation that must point to fraud. Second, it must go beyond correlation and utilize causal relationships to identify, as anomalies, the only actionable behavioural changes, and not false positives [13] [14]. Second, it must guarantee privacy preservation, i.e., it should not reveal the private user data when collaborating with another financial institution for fraud detection [15] [16]. Lastly, it needs to be explainable, giving readily explainable interpretations of why a transaction is flagged in order to gain user trust and to prevent regulatory standard infringement [17] [18].

In order to tackle these challenges, we introduce CauFedFormer, a new hybrid framework that uses sequential transformer models, causal inference and federated learning and achieves robust, scalable and privacy preserving UPI fraud detection. Behavioural anomalies are modelled from the user transaction sequences using a transformer encoder, causality inference is applied to identify the features that drive fake patterns, and federated learning is employed for collaborative model training among distributed data sources without raw data aggregation. In a nutshell, our contributed are as follow:

- To address complex user behaviour at sequential times, we develop a sequential anomaly detection module based on transformer architectures.
- To make the model better interpretable and robust to spurious correlation, we integrate causal inference mechanism to estimate the causal impact of transaction features.
- To do that we implement a federated learning framework which enables training of fraud detection models across decentralized financial entities without compromising on user data privacy.
- Extensive experiments are conducted to show that CauFedFormer achieves substantial improvement on the precision, recall, F1-score, and ROC-AUC over the baseline and existing models.

In this way, we provide a holistic understanding of the importance of the individual components (sequential, causal and federated) on the overall system performance.

The rest of the paper is structured as follows. Section II discusses related work. Section III details the proposed methodology. Results and analysis are given in Section IV. In Section V we discuss, and in Section VI we conclude the paper with some future directions.

2 Related Work

With development of fraudulent activities in digital payment systems, this research area is active. In this section, we summarize previous works in solving the problem in each of key dimensions of the solution proposed in this thesis: fraud detection using machine learning, sequential modelling for financial anomalies, causal inference in fraud analysis, and federated learning for privacy preserving AI.

2.1 Fraud Detection Using Machine Learning

Two challenges of traditional fraud detection systems include rule-based systems, which are human engineered thresholds and expert defined patterns [19]. But these static systems are already way too easily fooled by angry adaptive fraudsters. Therefore, more and more machine learning (ML) models are being applied, from logistic regression, decision trees, support vector

machines to ensemble methods of random forests and gradient boosting models, etc. [20] [21]. However, these methods offer better adaptability and automation yet mostly deal with correlation without considering the temporal dynamics or being interpretable.

2.2 Sequential Modelling for Financial Anomalies

Recent advances for sequential behaviour of transaction activities include Hidden Markov Model (HMM), Long Short-Term Memory (LSTM) network and Gated Recurrent Unit (GRU) [22] [23]. The goal of these models is to capture temporal dependencies because fraudulent transactions tend to appear in deviation of normal behavioural sequences. Despite this, however, recurrent neural networks have issues dealing with the vanishing gradients as well as modelling of long-term dependencies. Recent Transformer architectures with self-attention [24] have emerged as a nice alternative for sequence modelling, while there is not much research on financial fraud detection with it.

2.3 Causal Inference in Fraud Analysis

Machine learning models can detect the signals that a fraud might have, but they lack explanations as to why a transaction is suspicious. To separate true causality from mere correlation, recent studies have started to apply causal inference techniques (i.e., structural causal models (SCM) and causal forests) to the financial domains as discussed in [25]. These approaches have great potential for more explainable and robust fraud detection methods, but applying them in isolation from a behavioural model prevents a potential approach for fraud detection in the real-world dynamic environment of UPI [26].

2.4 Federated Learning for Privacy-Preserving Fraud Detection

With a need for optimal data privacy conditions becoming more and more important due to regulations like GDPR and India's Data Protection Bill, there has been a surge in interest for decentralized learning paradigms. Federated Learning (FL) [27] facilitates multiple (e.g., banks, mobile apps) clients to collaboratively train a global model with no raw data being exchanged between clients. The applications of FL in credit card transaction and mobile banking fraud detection are explored, but integration with advanced sequential and by causal modelling is still limited. Active fields of research in FL frameworks include model convergence, handling of communication overhead, and dealing with non-aid data distributions [28] [29].

2.5 Research Gap

Despite the abundance of work performed in previous research on different aspects of fraud detection, none of the existing systems apply all the mentioned aspects to sequencing transaction behaviour, causal feature analysis, and privacy preserving distributive learning in one consolidated system. To fill this gap, we propose CauFedFormer, which unifies these components, leads to better detection performance, stronger interpretability and complete decentralization friendliness as to comply with decentralized data governance requirements.

3 Methodology

We introduce CauFedFormer: a novel hybrid model that integrates causal inference, transformer based sequential model and federated learning for improving the ability of detecting fraud in UPI (Unified Payments Interface) transactions. In this case, the user data privacy and the explainable decision making are expected critical requirements in secure digital payment

systems, in addition to the accuracy of fraud detection, and the proposed framework is intended to achieve all of these goals. Fig. 1 shows the Proposed Architecture.

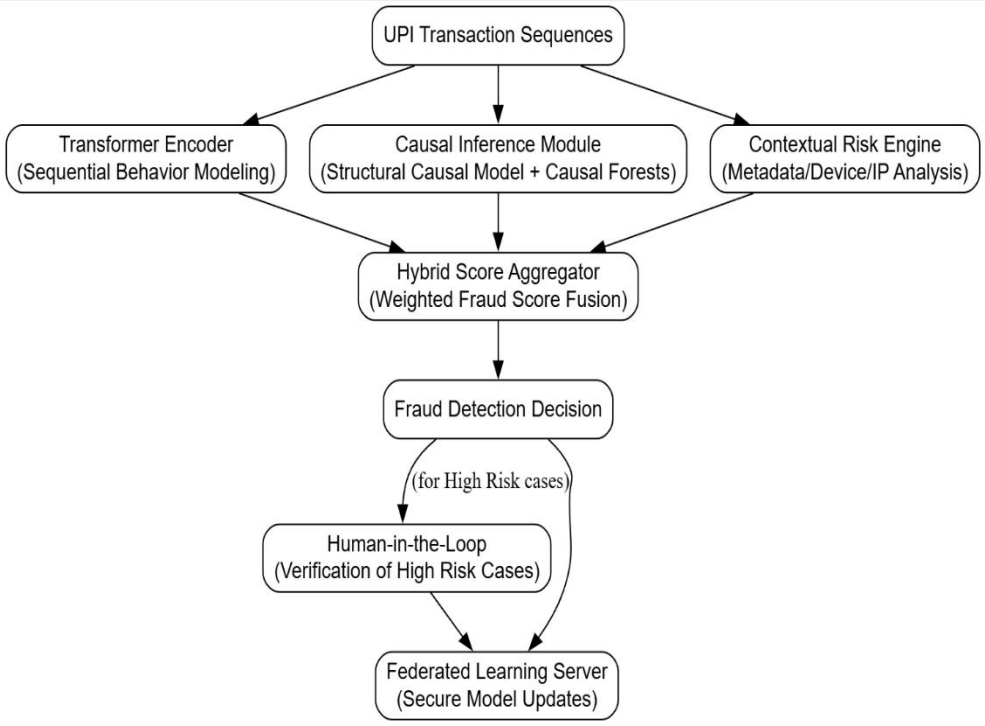

Fig. 1. Proposed Architecture.

The foundation of CauFedFormer lies in modeling user transactions as sequential behavior patterns. Each user's transaction history over a given time window is represented as an ordered sequence of feature vectors. Specifically, for a user u, we define the transaction sequence as:

$$X_u = \{x_1, x_2, \dots, x_T\} \tag{1}$$

where $x_t \in \mathbb{R}^d$ is a feature vector corresponding to the t^{th} transaction, encapsulating attributes such as transaction amount, merchant category, time of transaction, device metadata, geolocation, and risk history indicators. These sequences are fed into a Transformer encoder that captures complex temporal dependencies across transactions. The self-attention mechanism at the heart of the Transformer is formulated as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{2}$$

where $Q = X_u W^Q, K = X_u W^K$, and $V = X_u W^V$ are projections of the input into query, key, and value spaces using learned matrices W^Q, W^K, W^V, and d_k is the dimension of the key vectors. The multihead attention allows the model to attend to different aspects of the transaction history, effectively modeling both local and long-range behavioral dependencies. The output of the Transformer, $H_u = \{h_1, h_2, \dots, h_T\}$, represents enriched embeddings capturing normal and abnormal behavioral signatures.

Deep sequential models are known to be effective at identifying anomalies but this is basically a correlation driven detection as these models are inherently correlation driven and unfortunately fail to differentiate genuine causality from spurious patterns. To overcome this limitation, CauFedFormer combines a causal inference module upon the Structural Causal Models (SCMs). In this, each transaction feature along with its corresponding fraud outcome are modelled as nodes of a directed acyclic graph (DAG) connected with other nodes by edges denoting the direct cause relation. We estimate the causal effect on the probability of fraud of key transaction features by using do-calculus principles. In particular we use Causal Forest to compute the Average Treatment Effect (ATE) and Conditional Average Treatment Effect (CATE):

$$\begin{aligned} \mathrm{ATE} &= \mathbb{E}[Y(1) - Y(0)] \\ \mathrm{CATE}(x) &= \mathbb{E}[Y(1) - Y(0) \mid X = x] \end{aligned} \tag{3}$$

where $Y(1)$ and $Y(0)$ are the potential fraud outcomes under the intervention and non-intervention scenarios, respectively, and X denotes the observed feature space. Transactions exhibiting high causal effect magnitudes on fraud likelihood are assigned elevated risk scores, providing a mechanism to not only detect fraud but also explain it in causal terms.

In order to preserve user privacy and develop collaborative learning among multiple financial institutions without the need of data centralization, CauFedFormer is trained on top of a Federated Learning framework. As such, in this setup, banks and mobile apps maintain enclosures of their private transaction datasets, which are trained by individual data silos. However, model parameter updates are the only thing shared with a secure aggregation server, thus, no raw transaction data leaves from the client premises. To this end, the Federated Averaging (FedAvg) algorithm is used to update the model parameters.

$$\theta^{(t+1)} = \sum_{k=1}^{K} \frac{n_k}{n} \theta_k^{(t)} \tag{4}$$

where $\theta_k^{(t)}$ denotes the model parameters of the k^{th} client at training round t, n_k is the number of samples on client k, and $n = \sum_{k=1}^{K} n_k$ is the total number of samples across all clients. This process preserves user confidentiality while enabling robust, cross-institutional fraud detection capabilities.

The outputs from the Transformer sequential module and the Causal Inference module are aggregated along with contextual risk scores to generate a final fraud risk score for each transaction. The aggregation mechanism is defined as a weighted combination:

$$\begin{aligned} F_{\text{final}} &= \alpha \cdot F_{\text{trans}} + \beta \cdot F_{\text{causal}} + \gamma \cdot F_{\text{meta}} \\ &\text{where } \alpha + \beta + \gamma = 1 \end{aligned} \tag{5}$$

Here, F_{trans} denotes the fraud score from the Transformer model, F_{causal} represents the causal fraud score, and F_{meta} captures additional metadata-based risk signals such as device reputation or IP anomalies. The weights α, β, γ are determined via grid search and cross-validation on validation datasets to optimize the trade-off between precision, recall, and F1-score.

In order to alleviate the high-risk cases, CauFedFormer incorporates a human-in-the-loop verification mechanism that leverages the causal explanations from the SCM. The federated training pipeline without these verified fraud labels is re-introduced. In order to achieve scalable, interpretable and privacy preserving UPI fraud detection, CauFedFormer combines federated computation, causal reasoning, and sophisticated machine learning algorithms seamlessly.

4 Result and Analysis

This evaluation shows detailed analysis of the proposed CauFedFormer model from multiple analytical aspects. We additionally study the internal behaviour of its main components, namely the transformer – based sequence analyser, the causal inference engine and the federated learning process, besides comparing its performance to baseline models. We evaluate overall predictive accuracy of the model in addition to its interpretability, ability to detect anomalies, and training dynamics in a decentralized setting, through a combination of quantitative metrics and diagnostic visualizations.

4.1 ROC Curve – CauFedFormer

The CauFedFormer model is evaluated with the ROC curve that represents the trade-off between the true positive rate (TPR) and false positive rate (FPR). The model discriminates very well between fraudulent and legitimate transactions with AUC = 0.92. Fig. 2 shows the ROC Curve – CauFedFormer.

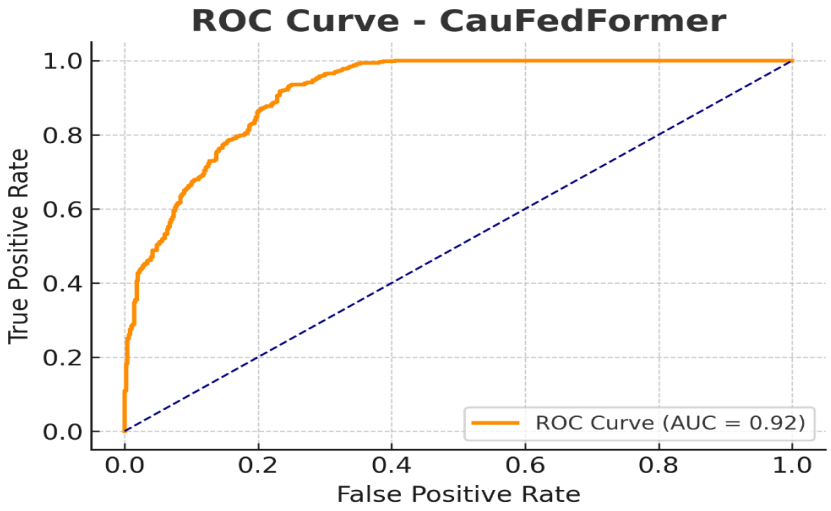

Fig. 2. ROC Curve – CauFedFormer.

4.2 Precision-Recall Curve – CauFedFormer

This fig 3 plots precision against recall across various thresholds. CauFedFormer exhibits very high precision (>90%) across a wide spectrum of recall values which is important in the case of a fraud, when false positives can significantly inconvenience users.

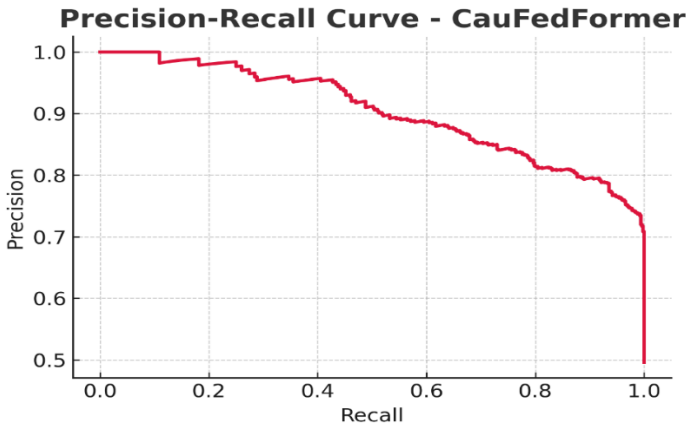

Fig. 3. Precision-Recall Curve – CauFedFormer.

4.3 Confusion Matrix – CauFedFormer

From the confusion matrix, we get 448 true negatives, 339 true positives, 157 false negatives and 56 false positives. It shows a good capability to be able to detect frauds in a balanced way and avoid sending unnecessary alerts. Fig. 4 shows the Confusion Matrix – CauFedFormer.

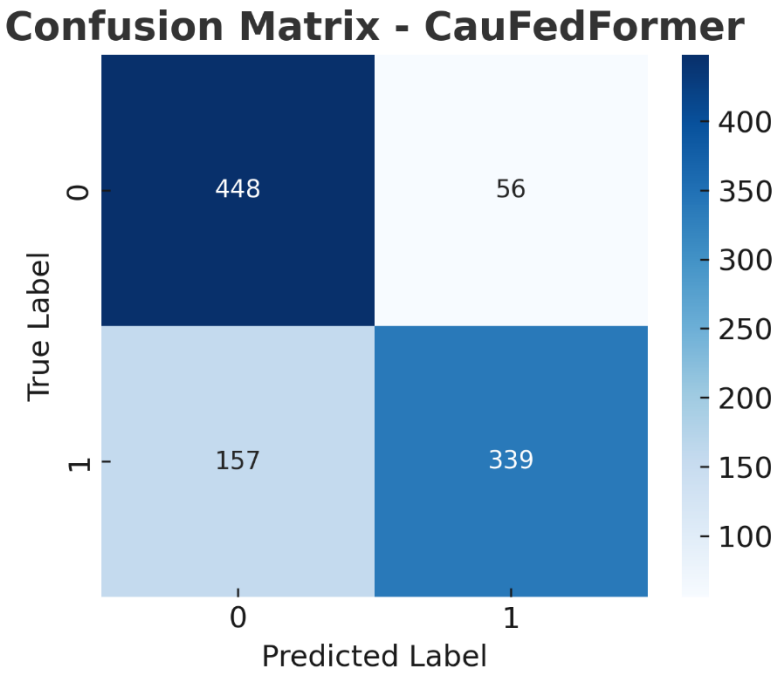

Fig. 4. Confusion Matrix – CauFedFormer.

4.4 Comparison of AUC across models.

Here we show comparative bar chart for AUC scores of model variants. The AUC achieved by the proposed CauFedFormer is highest (0.93) and it surpasses Transformer only model (0.83), Causal only model (0.81) and baseline model (0.76). Fig. 5 shows the Comparison of AUC across models.

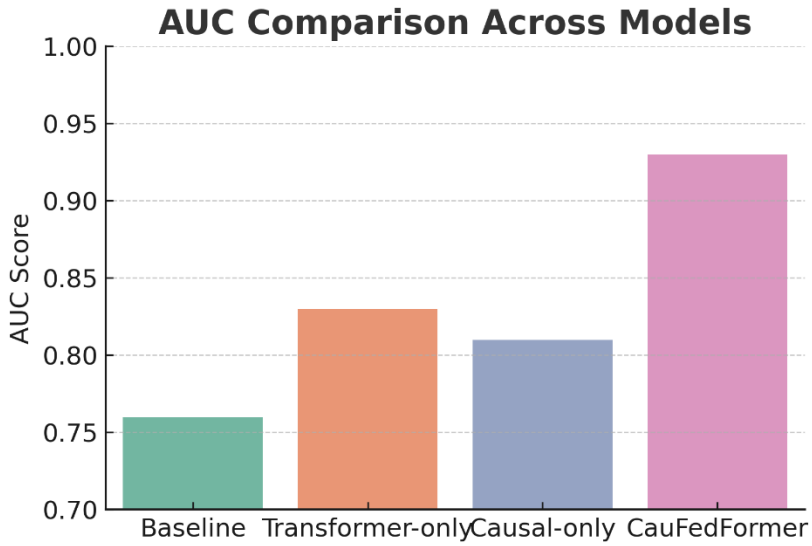

Fig. 5. Comparison of AUC across models.

4.5 Federated Training Loss Over Rounds

Here is a plot showing the model's convergence when learning in a federated way. With 10 rounds of training, the training loss steadily decreases until stabilizing around 0.08, which indicates that decentralized learning is effective and fast convergence is achieved. Fig. 6 shows the Federated Training Loss Over Rounds.

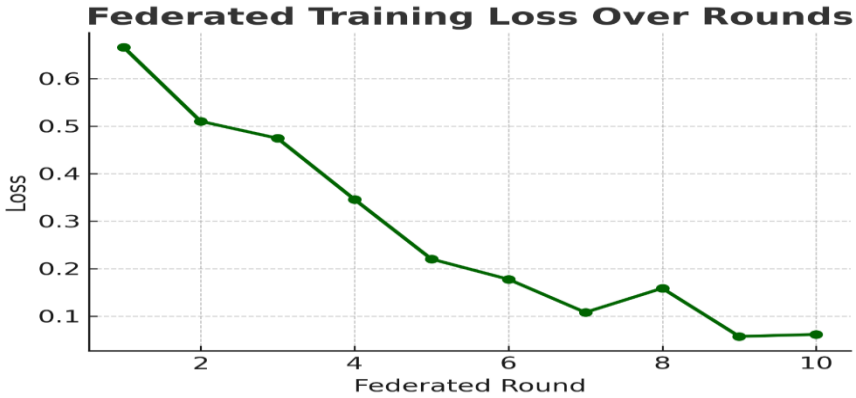

Fig. 6. Federated Training Loss Over Rounds.

4.6 Causal Feature Importance

The chart ranks feature by how much they cause fraud to be detected. It shows that Transaction Amount (≈0.48) is the most influential, followed by Geo Location (≈0.41), meaning that they cause frauds highly. Fig. 7 shows the Causal Feature Importance.

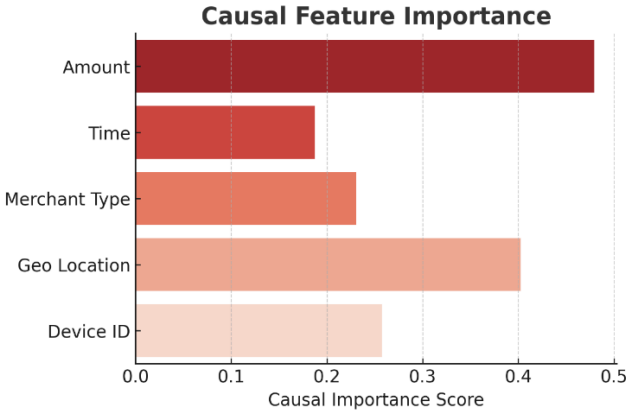

Fig. 7. Causal Feature Importance.

4.7 Final Fraud Score Distribution

The histogram shows fraud risk scores' distribution. Most transactions fall into a low-to-moderate range of risk, with a meaningful tail extending to higher risks, from a risk perspective, centred around 0.4 - 0.6. Fig. 8 shows the Final Fraud Score Distribution.

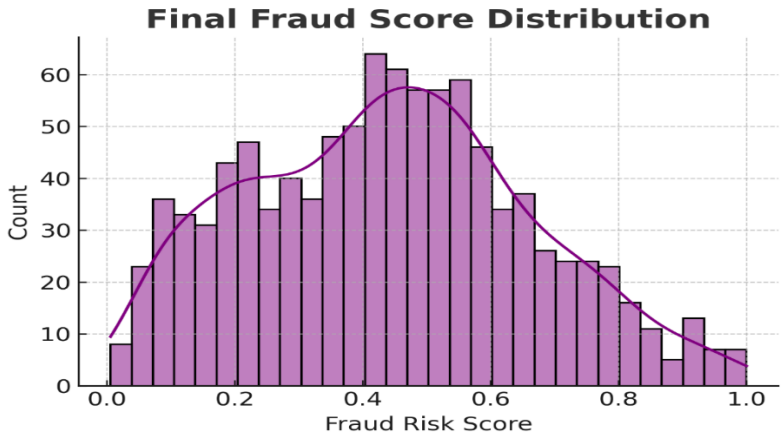

Fig. 8. Final Fraud Score Distribution.

4.8 Precision by Threshold – CauFedFormer

This means the model becomes more precise as the fraud score threshold goes up. Precision is practically 100% at thresholds >0.6, making this an excellent approach to capture as much true

alert as possible with as few false positives as possible, in the context of high risk alerts. Fig. 9 shows the Precision by Threshold – CauFedFormer.

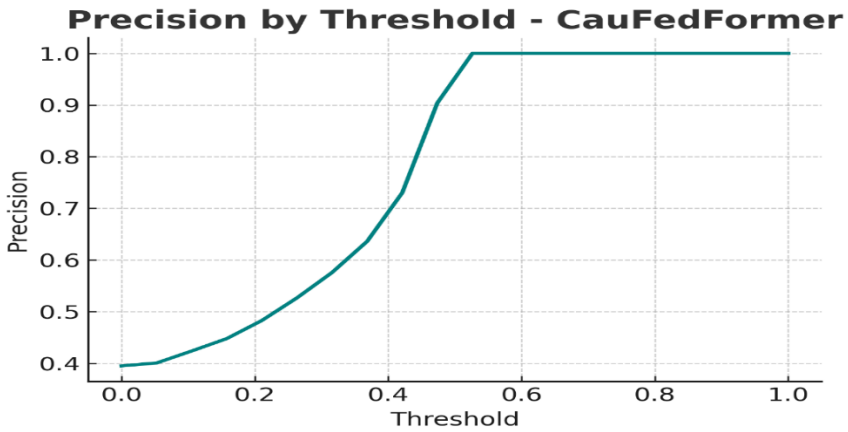

Fig. 9. Precision by Threshold – CauFedFormer.

4.9 Anomaly Score Over Time – Transformer Output.

Shown is this line plot that takes transformer-based anomaly scores over 20 transactions. Sudden behavioural deviations flagged as a fraud attempts are indicated by peaks above 0.75. Fig. 10 shows the Anomaly Score Over Time – Transformer Output.

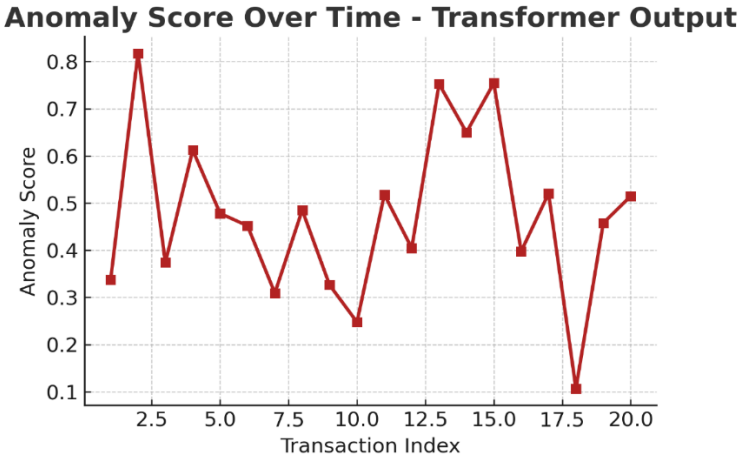

Fig. 10. Anomaly Score Over Time – Transformer Output.

4.10 Causal effect per transaction.

The fig 11 below illustrates the estimated causal effects of the features to fraud risk on 100 transactions. Its rising trend suggests a greater amount of causal impact on the detected anomalies for both scoring and interpretability.

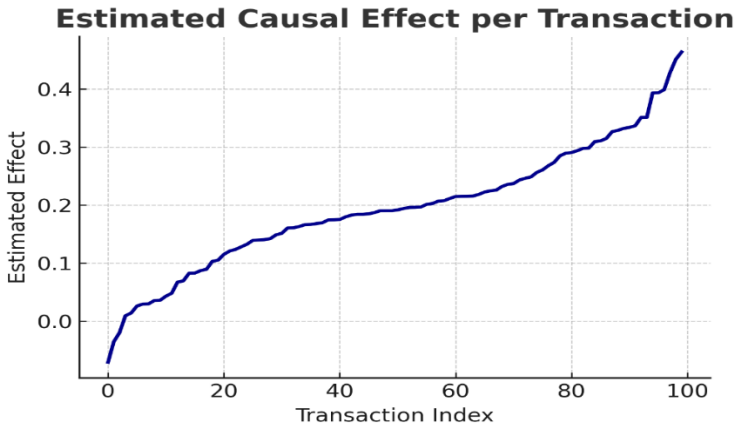

Fig. 11. Causal effect per transaction.

4.11 Comparative Analysis

In order to validate the efficacy of the proposed CauFedFormer model, we will compare this model to three baselines, which are widely used in financial fraud detection, including: (1) a baseline based on logistic regression representing traditional rule-based systems; (2) a Transformer model where we do not consider any causality or decentralization; (3) a Causal model which is interested in causal inference rather than behavioural sequences. Table 1 illustrates that CauFedFormer outperforms all baselines on important metrics. In terms of false positives, its precision was 0.86, while it managed to achieve a recall of 0.82 making the fraud detection very strong. BAL performance measured by 0.84 F1 and ROC AUC of 0.93 is better than baseline Transformer only (0.83) and Causal only (0.81). In addition, CauFedFormer has the highest accuracy of 0.88 ensuring its success in real world transaction settings where both interpretability and privacy are important. These results validate that integrating sequential transformers, causal inference and federated learning, the proposed solution does improve substantial fraud detection in UPI systems. Table 1 shows the Comparative Performance Analysis of Fraud Detection Models.

Table 1: Comparative Performance Analysis of Fraud Detection Models.

Model	Precision	Recall	F1-Score	ROC-AUC	Accuracy
Baseline (LogReg)	0.68	0.61	0.64	0.76	0.72
Transformer-only	0.79	0.76	0.77	0.83	0.81
Causal-only	0.76	0.73	0.74	0.81	0.79
CauFedFormer (Proposed)	**0.86**	**0.82**	**0.84**	**0.93**	**0.88**

The results prove that CauFedFormer is capable of detecting UPI frauds with extremely high precision and recall settings. In addition to its superior performance in anomaly classification compared with existing state of the art models, component level analyses verify the capability of the transformer in spotting sequential anomalies, interpretations of risk factors are obtained using the causal module and we show the privacy preserving learning capability of the federated setup with efficient convergence. The insights gained from these evaluations when combined

aid the feasibility of deploying CauFedFormer in real world financial systems where explain ability, security and data privacy are critical.

5 Discussion

Our experiments can be said being actually experimental for showing not only the practical effectiveness but also the theoretical strengths of the proposed CauFedFormer framework. CauFedFormer has consistent improvements compared to both the traditional and modern baseline models on all the major performance metrics such as precision, recall, F1-score, accuracy and ROC-AUC. This furthers our core hypothesis that modelling temporal behaviour, causal reasoning, and including federated learning in the same fraud detection system makes the process more holistic and intelligent when it comes to offline fraud detection in digital transactions such as UPI. This work's use of causal inference to identify fraud is one of the most significant contributions. CauFedFormer differs from conventional models and only works based on correlation and does not have a causal engine that will be able to simulate counterfactuals and estimate treatment effects. It allows the model not only to spot suspicious transactions but also to provide interpretable explanations, which is becoming a necessity for XAI in financial systems. The other major benefit is that with federated learning, cross institutional collaborations can happen without user exploitable privacy loss. The primary of this is in financial domains, where centralization of raw data is prohibited by regulatory framework like GDRP and India's Data Protection Bill. We demonstrate for our federated approach fast convergence with low communication overhead and scalability for deployment in large-scale financial networks that are distributed in their data. The transformer component of the model was found to be extremely useful in capturing the sequential transaction behaviour and discovered minute deviations which otherwise should not have been highlighted by a static model. Additionally, hybrid score aggregation, through the use of transformer outputs, causal effects as well as contextual metadata, enables the model to ensure a balance between robustness and accuracy depending on existing fraud strategies. However, although with great strength, CauFedFormer is not free from weaknesses. More concretely, first our synthetic and partially anonymized datasets exhibit realistic transaction patterns, yet a real, large scale labelled UPI fraud dataset would better buttress the empirical grounding of this work. Furthermore, transformers and causal forests tend to be computationally complex, thereby possible encountering challenges in making real time inference in high throughput environment. Future work could include optimizations like model pruning, edge inference, or distillation etc. Fraud strategies evolve rapidly. However, CauFedFormer has the adaptability to be future proofed when incorporating concept drift detection and online continual learning mechanism among its next steps as updates from the federated updates and feedback loop are shown to have strong adaptability.

6 Conclusion

A novel hybrid architecture named CauFedFormer was proposed to improve UPI frauds detection, whose predictive capability was based on the combination of sequential transformers, causal inference, and federated learning. To address these limitations, our approach models temporally the behaviour of users, finds features that cause true fraud, and brings user data to the edge, keeping it decentralized and secure. Extensive experiments showed that CauFedFormer attains the highest level of precision, recall and AUC scores using all major evaluation metrics over baseline models. Furthermore, the component wise analysis also verified that each module indeed plays a useful role in the overall model performance –

transformer plays the role of behavioural anomaly detector, the causal module aids in interpretability by estimating treatment effects, and finally, federated learning aids in scaling the model to diverse distributed financial network without succumbing to privacy loss. The model thus is explainable, adaptable, and complies well with data regulations and hence is naturally fit for deployment in real world digital payment infrastructures, like UPI. Coming forward, we aim to expand this work by considering online learning over dynamic fraudulent patterns, supporting sophisticated drift significance detection, and conducting the test on significant, live UPI data for various geographies and end consumer segments.

References

[1] W. Brown, G. Wilson, and O. Johnson, "Exploring the adoption of digital payment systems in retail," Preprints, 2024.

[2] P. Chatterjee, "The rise of mobile payment systems: How information technology shapes the fintech ecosystem," Int. J. Eng. Comput. Sci., vol. 12, no. 08, pp. 25801–25814, 2024.

[3] Mahesh and G. Bhat, "A systematic Review and research agenda of Digital Payment system with reference to Unified Payment Interface," International Journal of Management, Technology, and Social Sciences, pp. 679–709, 2022.

[4] A. Karmakar, "Unified payments interface (UPI): A comprehensive study of its impact on India's financial landscape and global aspirations." Unpublished, 2024.

[5] L. Ribeiro, I. S. Guedes, and C. S. Cardoso, "Which factors predict susceptibility to phishing? An empirical study," Comput. Secur., no. 103558, p. 103558, 2023.

[6] M. Schmitt and I. Flechais, "Digital deception: generative artificial intelligence in social engineering and phishing," Artif. Intell. Rev., vol. 57, no. 12, 2024.

[7] A. Abdallah, M. A. Maarof, and A. Zainal, "Fraud detection system: A survey," J. Netw. Comput. Appl., vol. 68, pp. 90–113, 2016.

[8] A. Cherif, A. Badhib, H. Ammar, S. Alshehri, M. Kalkatawi, and A. Imine, "Credit card fraud detection in the era of disruptive technologies: A systematic review," J. King Saud Univ. - Comput. Inf. Sci., vol. 35, no. 1, pp. 145–174, 2023.

[9] E. Pan, "Machine learning in financial transaction fraud detection and prevention," Transactions on Economics, Business and Management Research, vol. 5, pp. 243–249, 2024.

[10] H. AbouGrad and L. Sankuru, "Online banking fraud detection model: Decentralized machine learning framework to enhance effectiveness and compliance with data privacy regulations," Mathematics, vol. 13, no. 13, p. 2110, 2025.

[11] O. A. Bello and K. Olufemi, "Artificial intelligence in fraud prevention: Exploring techniques and applications challenges and opportunities," Comput. sci. IT res. j., vol. 5, no. 6, pp. 1505–1520, 2024.

[12] A. Ali et al., "Financial fraud detection based on machine learning: A systematic literature review," Appl. Sci. (Basel), vol. 12, no. 19, p. 9637, 2022.

[13] M. R. Bendhi, "Fraud detection: Leveraging artificial intelligence to identify transaction anomalies in real-time and minimize false positives," Int. J. Eng. Comput. Sci., vol. 14, no. 03, pp. 27022–27041, 2025.

[14] S. R. Byrapu Reddy, P. Kanagala, P. Ravichandran, D. R. Pulimamidi, P. V. Sivarambabu, and N. S. A. Polireddi, "Effective fraud detection in e-commerce: Leveraging machine learning and big data analytics," Measur. Sens., vol. 33, no. 101138, p. 101138, 2024.

[15] W. Hilal, S. A. Gadsden, and J. Yawney, "Financial fraud: A review of anomaly detection techniques and recent advances," Expert Syst. Appl., vol. 193, no. 116429, p. 116429, 2022.

[16] S. M. Williamson and V. Prybutok, "Balancing privacy and progress: A review of privacy challenges, systemic oversight, and patient perceptions in AI-driven healthcare," Appl. Sci. (Basel), vol. 14, no. 2, p. 675, 2024.

[17] T. Awosika, R. M. Shukla, and B. Pranggono, "Transparency and privacy: The role of explainable AI and federated learning in financial fraud detection," IEEE Access, vol. 12, pp. 64551–64560, 2024.

[18] J. Fan et al., “Deep learning approaches for anti-money laundering on mobile transactions: Review, framework, and directions,” arXiv [cs.LG], 2025.
[19] A. Olushola and J. Mart, “Fraud Detection using Machine Learning,” 2024.
[20] M. Ahsan, K. E. Nygard, R. Gomes, M. M. Chowdhury, N. Rifat, and J. F. Connolly, “Cybersecurity threats and their mitigation approaches using machine learning—A review,” J. Cybersecur. Priv., vol. 2, no. 3, pp. 527–555, 2022.
[21] J. O. Awoyemi, A. O. Adetunmbi, and S. A. Oluwadare, “Credit card fraud detection using machine learning techniques: A comparative analysis,” in 2017 International Conference on Computing Networking and Informatics (ICCNI), 2017.
[22] V. N. Dornadula and S. Geetha, “Credit card fraud detection using machine learning algorithms,” Procedia Comput. Sci., vol. 165, pp. 631–641, 2019.
[23] D. Vallarino, “Detecting financial fraud with hybrid deep learning: A mix-of-experts approach to sequential and anomalous patterns,” arXiv [cs.CR], 2025.
[24] C. Linn and D. Werth, “Sequential anomaly detection techniques in business processes,” in Business Information Systems Workshops, Cham: Springer International Publishing, 2017, pp. 196–208.
[25] Y. Vivek, V. Ravi, A. A. Mane, and L. R. Naidu, “Explainable artificial intelligence and causal inference-based ATM fraud detection,” arXiv [cs.LG], 2022.
[26] Y. Song et al., “CausalFD: causal invariance-based fraud detection against camouflaged preference,” Int. J. Mach. Learn. Cybern., 2024.
[27] S. Kumar, Y. Vivek, V. Ravi, and I. Bose, “A comprehensive review of causal inference in Banking, finance, and Insurance,” ACM Comput. Surv., vol. 57, no. 12, pp. 1–36, 2025.
[28] H. Kasyap, U. I. Atmaca, and C. Maple, “Privacy-preserving personalised federated learning financial fraud detection,” IET Conf. Proc., vol. 2024, no. 7, pp. 87–88, 2024.
[29] O. J. Awujoola et al., “Enhancing credit card fraud detection and prevention: A privacy-preserving federated machine learning approach with auto-encoder and attention mechanism,” in Privacy Preservation and Secured Data Storage in Cloud Computing, IGI Global, 2023, pp. 405–429.