**AI and Data Science Journal**
**ISSN: 3078-7394**
**DOI: https://doi.org/10.61784/adsj3020**

# CAUSAL INFERENCE-BASED DIGITAL PAYMENT FRAUD DETECTION: FROM FINANCIAL SECURITY TO ECONOMY-WIDE RESILIENCE

LuQing Ren
*Columbia University*
*Corresponding Email: lr3130@columbia.edu*

**Abstract:** With the explosive expansion of digital payment systems, financial fraud has now become one of the most serious threats facing economic stability in many sectors. This paper presents a critical analysis of how methods for causal inference could contribute to improvement of fraud detection by revealing underlying patterns rather than correlations. The paper presents a theoretical model that integrates machine learning and causal analysis methods to improve the differentiation between legitimate and fraudulent transactions. Through the detection of interaction networks and behavioural patterns, the methodology attains a higher level of accuracy in identifying sophisticated fraud schemes than a traditional rule-based system. The results propose that causality methods do not just mitigate false positives in financial industries but they also present actionable risk controls for the application domain of e-commerce, healthcare and digital transaction processing. The research serves to enhance financial security efforts by designing a more thorough methodology which is also flexible enough to evolve in response to advances in fraud techniques. Next steps include generalizing causal models to cover new threats in decentralized finance and cross-border payments.
**Keyword:** Causal inference; Digital payment; Fraud detection; Financial security; Risk prevention

## 1 INTRODUCTION: BACKGROUND AND RESEARCH OBJECTIVES

There have been drastic changes in the digital payment world in the past few years and transactions have hit record levels both in Nigeria and around the world. By mid-2025, rampant adoption of contactless payments, cryptopayments and decentralized finance (DeFi) platforms has widened the playing field for financial inclusion and increasing the possibility of fraud. The traditional type fraud systems mainly based on rule-based logic and historical correlation structure have become more and more difficult to deal with fraud behaviors in new payment model space. This detection gap highlights the necessary need for more sophisticated analytic strategies that can distinguish among causal mechanisms that drive fraudulent behavior rather than the aggregate patterns of recognition[1].

Fraud in financial digital services such as online and mobile transactions expresses itself through various vectors, such as identity theft, synthetic identity fraud, and multi-account attacks. Existing solutions make use of static thresholds and transaction by transaction analysis and consequently may lead to the presence of too many false alarms. The shortcomings arise especially in cross-border deals and (semi-)autonomous ecosystems, where malicious catalysts exploit jurisdictional arbitrage and weakly-designed smart contracts[2]. A deviation towards causal inference approaches appears particularly promising to address these issues by treating transactional networks as interlinked systems rather than single events. This scheme would allow finding the root causes behind spamming activities, SoWs or phishing based UBOs.

The paper has three research missions. The first goal is to develop a novel theoretical approach where causality inference methodology is embedded within machine learning models in order to advance the accuracy of fraud detection. Unlike correlation-based models that can just confuse noise (coincidental patterns) for signal, causal models are interested in the "directional" nature of the relationships – here the relationship between transactional features and fraud indicators. Second, the purpose of the study is to show by what means causal analysis can significantly diminish false positive rates while preserving high detection sensitivity for a variety of financial sectors. We can achieve such separation based on counterfactual reasoning, i.e., a legitimate unusual transaction (e.g., emergency medical payment) is different from a truly fraudulent one. Third, it investigates the generalization of causal fraud detection models to non-financial applications, where digital transactions are still on the rise, ranging from healthcare billing systems to gig economy platforms.

Recent developments in causal discovery algorithms and graph neural networks have opened new possibilities to study transaction systems as dynamic networks. By representing payment flows as a causal graph whose nodes are accounts and which contains information on the temporal order of the transactions, the procedure allows to identify fraud rings involving colluding entities that were previously unknown. Globally, however, trust has a limited shelf life – and nowhere more so than in the dynamic financial industry in mid-2025 where it's vital to detect and respond to fraud in real-time. The suggested model incorporates time-varying confounding effects, such as consumer spending due to seasonality of macroeconomic shocks, in order to enhance the robustness of detection.

Practically, the innovation in this study is not only limited in the traditional banking industry, instead, it also included the financial innovations such as the blockchain-based micropayments as well as the AI-drive financial services. Now that digital wallets have replaced cash as a medium of exchange even in the developing world, the causal inference approach itself can scale to be applicable to a variety of transactional settings. Moreover, the work tackles important deficiencies of the state-of-the-art fraud prevention as it provides human-readable detection results, which is crucial for regulatory compliance and reconciling disputes with customers. By linking causal inference theories to actual fraud detection requirements, this work aims to enable more robust financial ecosystems in the face of relentless digital evolution[3].

## 2 LITERATURE REVIEW AND THEORETICAL FRAMEWORK

### 2.1 Current State of Digital Payment Fraud Detection

The development of e-payment fraud detection systems is an arms race between the security controls and the advances of fraudulent tricks. As of the middle of 2025, rule-based systems and historic transaction data trained supervised learning models continue to be the popular modalities for detection. Such systems generally highlight suspicious behavior if amounts are above certain thresholds, if they are issued more frequently, or if they appear from unusual geographic locations. However, they are easily detectable if they are correlation-based, and the problem is that this type of DDA could result in too many false alerts, which is not so favorable for Houdini--it will be a big headache for our commercial bank customers[4].

Three primary detection paradigms are employed in the industry today. Most commonly, signature-based anomaly detection is used, which monitors predefined suspicious patterns of fraud, such as sudden transfer of a high amount of value or frequent consecutive transactions on different accounts. While computationally efficient, this approach falls short of quickly adapting to new attack vectors, especially in the emerging domain as the decentralized finance, in which the transaction patterns have no historical precedence. Behavior-based systems are more sophisticated and make use of machine learning to create per-user profiles, including: normal spending locations, time of day phenomenon and vendor preference[5]. The system generates alerts when the transactions deviate materially from these patterns. But these models often misunderstand appropriate changes in behavior, like international travel spending or emergency shopping.

Network analysis methods have been widely adopted to uncover organized fraud associated to multi-account. By creating a network of users, devices, and paths, these tools can detect intricate rings of fraud that would escape the attention of single-account monitoring. Visual representation of transfer networks could expose obscured links between addresses, such as shared IP addresses, device fingerprints, or money flow trajectory. However, current designs suffer from difficulty of separating legitimate network behavior (e.g., family account clusters) and malicious coordination, since they focus on structral similarities rather than causal relationships[6].

The shortcomings of the solutions which already exist are particularly revealing in today's difficulties. With the proliferation of real-time payment environments by 2025, fraud decisions in real time are required which is not something batch architectures can achieve. Likewise, the growing concern of cross-platform fraud—bad actors taking advantage of integration points between banking apps and e-wallets and merchant systems—demonstrates weaknesses in siloed detection strategies. Evolving threats There are several types of synthetic identity fraud but the fastest-growing method is where identity fraudsters aggregate true and false information together to cook up authentic looking profiles and slowly build up a history of transactions before using them in bigger attacks.

With the onset of new payment technologies, things are even more complicated in detecting fraud. The growing popularity of contactless transactions enabled by near-field communication (NFC) and biometric-based identification has brought about a partial decrease of some types of frauds, coupled by additional vulnerabilities in the process of user authentication. Cryptocurrency (crypto) has brought with it, a system that the existing fraud systems implementation from the banking sector is not prepared to deal with, due to the pseudo-anonymous nature and irreversible aspect of transactions. The widespread adoption of "buy now, pay later" services has also broadened the attack surface, with detection systems needing to consider fraud risk over larger timeframes than individual transactions.

Performance characteristics of present systems show that there are intractable compromises between detection sensitivity and system throughput. Most organisations are enjoying impressive success in finding known types of fraud, but still face alarmingly low precision rates – the rate at which legitimate activity is incorrectly identified as fraudulent. This doesn’t just make the customer angry, it also leaves investigation teams with more cases than they can handle. Moreover, even for smaller financial service providers, it is still expensive to execute powerful machine learning approaches in massive scales [7], leading to security gaps in payment ecosystem.

Functional requirements are becoming more influenced by regulations but especially in the countries were strong customer authentication is becoming mandatory. Detection layers driven purely by compliance, on the other hand, can be at odds with risk-based strategies, meaning systems are all too often forced down the path of favouring a ‘tick-box’ exercise over a proper analysis on fraud. The EU’s second Payment Services Directive (PSD3), and analogous mandates globally, have compelled institutions to adopt multi-factor authentication, leading inadvertently to a situation where fraudsters began attacking the weakest verification points elsewhere in the transaction lifecycle.

Responses from the industry to these challenges have started to include features of causal reasoning, though in a limited sense at the moment. Some more sophisticated systems have started to leverage a temporal causality analysis to detect whether a sequence of actions chronologically proceeds a fraudulent transaction, such as account takeover patterns that start with credential phishing. Others also use intervention analysis to examine transaction patterns after security controls are implemented. Yet such efforts are piecemeal as opposed to being systematic, and thus only illustrates the necessity of the holistic causal model advanced in this work. In the subsequent section, we discuss how causal inference methods could fill these longstanding holes in digital payment fraud detection.

## 2.2 Causal Inference Methods in Fraud Detection

Use of causal inference techniques for fraud detection is a substantial improvement over the earlier, correlation-based methods. Contrary to traditional methods that detect suspicious behaviors through statistical deviation from the norm, causal models are designed to find root causes behind the fraudulent behavior. This change of perspective permits detection systems to differentiate between pure mere correlational associations, and so-called causative effects in transactional data.

The essen-tial difference between correlation and causation is that the former can actually be interpreted and employed for prediction. Even when transaction monitoring systems based on correlation are capable of flagging transactions with suspicious behaviors, such as unexplained spike in transferred value, inconsistencies in the geography of transaction and so on they usually do not take into account background information that explains the observed behaviors. A major purchase overseas, for instance, could be a sign of fraud, but whether that’s because the cardholder took a legitimate trip, or because the card got hacked can be the subject of causal analysis. Causal inference helps with this by explicitly modelling how some things, such as user authentication failures or the ordering of transactions, directly influence the probability of seeing fraud[8].

Promising causal inference technologies which have potential applications in fraud are introduced, especially when current techniques encounter issues. Counterfactual analysis, for example, tests if a transaction would be considered fraudulent in different circumstances. This method is useful in suppressing false positives by comparing alternative hypotheses on why suspicious activities occurred. Equally, structural causal models are used to capture the causality between variables, e.g., how attempts to login to an account may cause transaction behaviour. By learning such dependencies, the models are able to detect sequences of activities which frequently lead to a fraudulent result.

Another important technique is to use causal discovery algorithms to automatically learn possible cause-effect relations from observational data. Such algorithms are highly useful in identifying rings of fraudsters, who engage with each other in intricate fashions. Unlike structural network analysis, which assumes that account actions merely coincide with one another, causal discovery discovers whether some account actions actually have a causal impact on others, delving further to the underlying causes of fraud. E.g., in a money laundering network, causal models are able to reveal which accounts are causal initiators of criminal transactions, not simply connected accounts.

The marriage of causal inference and machine learning improves fraud detection in many ways. Every aspect of the data, such as consumer behaviors, time of day, dollar amounts, devices, and geography, can be more easily encoded using feature engineering and fed into a machine learning model. Flexible machine learning models, fed by features from causal economics will likely do a better job of generalizing to new kinds of fraud (or late-paying consumers in the case of business-to-business credit). This is because causal features encode the underlying mechanism that generates fraud, which leads to a more robust model when faced with different attacking tactics. Moreover, they make models more interpretable, which is important for compliance and regulation. It is easier for banks to justify their fraud alerts when they have a clear path leading to them[9].

Time-based causality is crucial for detecting fraud in real time because transactions occur at breakneck speed in 2025. Timmers, of Chainalysis, said many scam operations hinged on perfectly timed actions, such as quick transfers of funds prior to a freeze or simultaneous attacks on various platforms. Causal models with time series analysis can identify these trends by looking at the extent to which certain events have a regular propensity of occurring before fraud rather than purely using static cut-off points. For example, a rapid sequence of micro-transactions followed by a high value withdrawal could signal a testing period, prior to a significant sized fraudulent transfer – a pattern traditional systems may not detect.”

Although causal methods have positive side, they are difficult for practical implementation. Unmeasured confounding factors, that influence both the proposed cause and fraud outcome, can lead to bias in causal estimates. For instance, the same macroeconomic changes can influence both consumer spending behavior and fraud rates, which leads to spurious causal relationships. More sophisticated analytic methods, including instrumental variable analysis, can be used to address these challenges by identifying the true causal effects of interest. Moreover, causal inference has high computational complexity which needs to be carefully optimized, in particular on large payment networks [10].

Recently observed phenomena in digital payments continues to confirm the necessity for causal methods. The expansion of decentralized finance (DeFi) and cross-jurisdiction transactions presents new vectors of fraud that traditional control mechanisms are not well positioned to combat. Causal models as a tool to decide the presence of a new attack approach provides a more flexible approach to handle the changing transactional contexts. Because they target fraud techniques, rather than static rules, these approaches offer an enduring resolution for a dynamic financial ecosystem.

The following part will expand on these ideas by discussing the theoretical framework that we propose that unites causal inference with machine learning to develop trainable fraud detection system. The goal of our framework is to generalize existing approaches and address their limitations, while preserving scalability and interpretability within heterogeneous payment ecosystems.

# 3 METHODOLOGY AND IMPLEMENTATION

### 3.1 Causal Inference Model Design for Fraud Detection

The CausalInfer model developed in this study for detecting digital payment fraud operates with a clear framework that serves as a channel for turning transactional data into actionable information. The model basically connects the transactional features to the fraud signals based on the causal dependencies between them which are analyzed from three perspectives temporal view, interaction view, and behavior difference. Rather than treating transactions independently as in traditional methods, this is a method which partially reconstructs the entire decision path leading to each transaction and can identify the cause of the cause rather than the sake of the sake at the suspect correlation level [11]

The model architecture is a system of four cooperating components cascade. First, the original transaction log data can be preprocessed by dividing raw transaction logs into structured causal graphs where nodes are financial entities (e.g., account, device, location) and edge types represent the relation of transactions. This is one of the reasons this graph representation is appealing; it captures temporal dependencies — a crucial element when it comes fraud schemes involving different steps, such as money laundering cycles or account takeover. Graphs account for both explicit transaction flows and implicit relations that can be inferred from a shared metadata, and therefore provide a full causal network for analysis.

Second, the causal discovery module uses constraint-based algorithms for identifying the directionality of links among variables. With the help of conditional independence tests, that system separates real causes from correlation. For example, it can infer whether two addresses sending high amounts of value back to back are legitimate business transactions or illicit money movement by analyzing intermediary nodes and time patterns. This step is specifically designed to counteract synthetic identity fraud where fraudsters intentionally introduce a level of “credibility” in their transactional patterns before the attack.

Thirdly, the what-if reasoning engine can be used to confirm the suspected fraud cases. When the software detects a questionable transaction, it runs what is known as “counterfactual simulations” in which it tries out other scenarios, using downgraded parameters such as alternate authentication and timing, to see if the anomaly holds. This method is very effective in drastically reducing false positives by differentiating between real fraud and normal, yet legitimate noise due to situational factors (ie emergency medical payments during travel). It augments the engine with domain knowledge via tunable parameters that capture the regulatory requirements and the entity-specific risk tolerances.

Design considerations focus on scalability in various payment environments. The model is designed in a modular way to do parallel processing of subgraphs, and thus allow for real-time processing even for a big transaction network that’s remained standard for the financial industry in 2025. Dynamic weighting schemes focus with the analysis resources on high-risk network segments that are identified based on causal centrality measures. This adaptive process is especially powerful for catching cross-platform fraud in which a fraudster is bad across so many financial services with dissimilar security stances.

The hybrid model's fraud classification layer is a blend of causality-based features and machine learning and delivers well-balanced performance. Conventional supervised learning algorithms tend to fail to address imbalanced fraud data, where fraud instances are a very small proportion of all transactions. Employing causally relevant features, for instance the chronologic sequence of authentication failures before a transaction, or the network distance to confirmed fraudulent accounts, the classifier achieves better precision without sacrificing recall. The output features an intelligible causal map to explain why a transaction was flagged: this helps assure compliance with regulations requiring explainability of decisions in financial services.

In practice several operational challenges specific to causal inference systems are dealt with. Graph storage within memory iterations reduce hardware demands such that the proposed solution can be practical for smaller financial business. A feedback loop updates continuously causal relationships utilizing fraud cases newly confirmed and false positive reports, letting the system adapt to new vector attacks. Such self-improving mechanism is important to maintain the detection capability in decentralized finance space when fraudsters adapt on a daily basis.

Validation checks promote model generalization to other fraud types. The testing framework measures performance along several dimensions: the detection latency for time-critical fraud, accuracy in detecting coordinated attacks across accounts, and generalization to unseen fraud. Comparative analysis shows the better power compared to the rule-based systems in the presence of advanced frauds, such as scams to laundering with blockchain transactions, canonical approaches are not able to identify the sequence of events.

The model has been designed with ethical consideration in mind in order to avoid biased results. They are used to prevent explicit bias in decision making, which would be informed primarily by transactions originating from a particular race or location. By modeling causal pathways leading to possible biases, the system can adapt decision boundaries in order to ensure equal treatment and at the same time retaining the effectiveness of the fraud detection. This is an increasingly relevant trend as digital payment systems move into underserved markets which are not homogeneous in their user behaviours.

It is gradually integrated with the legacy financial system. The early stage deployment will concentrate on supplementing existing fraud detection systems, as opposed to replacing them, so that the efficacy of the causal model can be verified in a gradual manner. The API-driven design ensures seamless integration with core banking systems, payment service providers and compliance reporting tools. Early adopters are seeing measurable efficiencies in their operations as the lower false positive rate enables their security teams to focus on those risks that present the highest risk[12].

Causal reasoning abilities of the model will be extended towards new challenges in the long term. Click the underlined links below to read about planned developments in terms of incorporating macro-economic indicators as context variables for fraud prediction and increasing the closure level of the time analysis that allows detecting slow-burn fraud schemes that escape the traditional control cycle. The flexible nature of the causal framework makes it suitable to cover for new payment methods which are expected to emerge in the next years and which will exclude it from being a "one-hit wonder" in the fast paced movement of digital finances.

### 3.2 Application and Validation in Financial and Cross-Industry Contexts

The use of causality-based fraud detection system is applied in several financial domains, with the capacity of being adaptable to several transactional contexts. In conventional banking, the approach can be interesting for detecting account takeover attempts based on causal sequence of login failures, reset password requests and fund residue transferring. In contrast with rule-based systems, which could treat such events as unrelated and isolated anomalies, the causal direction forms them into unified attack scenarios yet preserves benign sequences of multiple authentication events. Retail banking deployments achieve significant progress in terms of authorized push payment fraud detection, namely the pinpointing of the cause of the fraud mechanism ranging from social engineering attempts through to unauthorized transactions.

It helps e-commerce platforms to separate legitimate bulk purchases from card testing fraud. By analyzing the dynamics between browsing behaviour,checkout attempts and payment fails, the method detects coordinated attacks targeting merchant payment gateways. The causality aspect of the inference can help distinguish the fraudulent spikes from flash sales or seasonal shopping patterns. Marketplace applications also use network analysis to identify seller-side fraud—discovering otherwise hidden relationships between seemingly unrelated vendor accounts who are associated with fake review scams or inventory laundering[13].

Cross-industry validation demonstrates the flexibility of the framework in the medical billing domain where it differentiates between coding mistakes and upcharging schemes. The causal model investigates treatment protocols, prescription patterns, and billing codes to detect medically insupportable combinations of services suggesting fraud. Unlike the typical audits, which are based on the random samples, it traces suspicious claims by retracing the decision making process that produced billing entries. Insurance companies use the same reasoning to figure out staged accidents, or procedures that doesn't have to be done, they draw casual graphs to find the relations between claimants with procedures and the location of service.

Providers of digital wallets in developing countries have particular challenges in fighting fraud related to the use of the shifting habits of people and the very short credit histories. The causal approach responds by setting baseline transaction behavior for various customer segments and detecting deviations that indicate real fraud rather than financial inclusion. Mobile money systems effectively slash false positives on low-value transactions – an issue that has dogged agent banking networks – by measuring the risk of each transfer in light of its causes, rather than establishing hard benchmarks around amounts.

Validation metrics show the same trend of performance gain over all test domains. Comparative studies with legacy systems indicate higher detection rates for advanced act fraud categories and lower false positive ratios. The causal model performs especially well in uncovering new types of fraud by learning the underpinning mechanisms of attacks rather than taking historically-learned patterns as a base case. Field deployments have resulted in increased operational efficiency as the lower false alarm rate enables investigators to concentrate on verifiable cases with well-defined trails of causal evidence[14].

The use of these principles becomes particularly challenging when considering applications toward decentralized finance applications. The model describes the flow of tokens and liquidity pools interaction, and the system is effective in detecting smart contract vulnerabilities. Unlike classic blockchain analytics, which observe the static flow of funds, our causal approach can differentiate between legitimate DeFi activities and wash trading schemes used for market manipulation of assets. Cross-chain fraud detection derives from the model's capability to model asset bridging traces and to capture malicious address clusters in multiple ledgers.

Unintended validation insights in healthcare and insurance industries tell us about model interpretability. Medical fraud investigators appreciate their value because they enable audits to proceed efficiently without the need for researchers trained in advanced data science, so they find value in the causal diagrams that make sense of suspicious billing patterns. Insurance adjusters also rely on the system's capacity to produce a narrative explanation of alerted claims, drawing together technical evidence transactions and domain-specific causal rules about accident mechanics, or convalescence regimes.

These gig economy platforms also demonstrate the scalability of the framework for micro-transaction environments. It analyzes the casual relationships among job postings, worker accounts, and payment flows to identify fake task completion schemes and meanwhile tolerate real short-term work trends. The fact that the model is very simple for an online use, together with its lower time complexity, enables real-time processing of these data on systems processing millions of micro-transactions per day (approximately roughly the same number of requests: 38.5 millions), as this number of data is processed by caching or computing features on the fly based on raw data (cf. [15]).

Regulatory compliance becomes a hidden opportunity for validation. Credit institutions also respond positively to the clear linkage between fraud decisions and causative paths modelled by the causal network models as opposed to blackbox machine learning solutions. Such documentation of the rationale for decisions is a natural fit for financial transparency demands across jurisdictions. GDPR and PSD3 compliance validations match the causal approach with the requirement for explanation of automated decisions regarding user accounts.

The adaptive capabilities of the system are shown in longitudinal studies in production environments. Feedback loops enable the causal models to adapt to changing fraud tactics without complete re-training. Early adopters see decreasing fraud rates in time as the system gets better at proactively identifying and stopping emerging attack vectors before they can be widely exploited. This self-improvement feature is especially important for crypto-exchanges, since the fraud attack patterns change frequently when there is a new protocol feature that can be attacked.

The Cross-Industry Validation process uncovers a number of implementation best practices. Staged roll out and full parallel operation with its predecessors enables organizations to confirm performance advantages, while ensuring continuity of protection. Modular integration allows customized (e.g.HC-specific rules for causes for billing patterns) sectorial adaptations without losing core detection. Fraud Analyst training curriculum focuses on the interpretation of causal diagrams and counterfactual scenarios, enabling the translation from technical insights to operational decisions.

Future uses will challenge the limits of the framework in more and more complex settings. Forthcoming empirications are, for instance, in the context of international trade finance networks, where causal analysis needs to work across diverse regulation requirements and currency domains. Pilot studies in supply chain finance look hopeful in sniffing out round-trip trading circular schemes by simulating the causal flow of goods documentation and payment calendars. Given the convergence unavoidable at the digital payment ecosystem, it is worth discussing the adoption of the causal-inference framework for risk protection in a unified manner rather than traditional business sector walls.

## 4 CONCLUSION AND FUTURE DIRECTIONS

The work shows that causal inference algorithms lead to great improvements in the ability to detect digital payment fraud, and overcome serious issues found in correlation based approaches. The network-based features and fraud indicators are then utilized to build a directional relation between transaction features and fraud indicators, so as to accurately detect complex fraud patterns and generate much less false positives. The approach is flexible enough to be applicable in different financial waves and nascent payment environments, “from typical banks to the level of decentralized finance platforms.” And its capability to access and assess transactional networks as connected networks, as opposed to standalone transactions, delivers a resilient protection against emerging fraud techniques that characterize the financial landscape in mid-2025.

Key results show how causal reasoning enhances detection accuracy in multiple ways. First, counterfactual reasoning allows to distinguish real fraud from anomalous (though real) patterns that are due to contextual factors such as travel and emergency situations. Second, the structural causal model are able to uncover stealthy fraud rings due to studies money flow and account relationship far easier than traditional graph-based techniques. Third, protocol based causality analysis catches time-dependent attack patterns, which signature based approaches often overlook on real time payment scenario. These benefits have direct operational implications for financial institutions which result in lower investigation workloads and better transparency compliance.

The cross-industry validations unveil surprising use-cases outside the finance domain. Healthcare billing machines use causal diagrams to quickly audit questionable claims, and on-demand labor markets use the idea to sniff out phony task completion schemes. A methodology that focuses on basic fraud mechanics instead of a sector spesific fraud pattern, can drive such flexibility. The interpretable outputs of the system, that report fraud hints in the form of logical trails, further bridge the gap between purely technical detection and human decision making, satisfying increasing requirements for explainable AI in the regulated industries teams.

Several new challenges should be the focus of future work. One, scaling causal models to include cross-jurisdictional transactions could enhance the detection of fraud in global payment networks, where variable regulations make patterns more difficult to detect. Second, the inclusion of macroeconomic indicators as contextual variables could improve the capability to detect fraud waves, which are induced by a financial crisis or market manipulations. Third, construction of light-weight causal discovery algorithms might make the general approach also available for low-tier financial services who are currently browsing simpler rule based systems.

Decentralized finance holds out some particularly exciting prospects for the future. Existing approaches effectively detect smart contract exploitation and wash trading analysis, but more work remains to prevent new attack vector on the trades with tokenized assets and cross-chain bridges. Likewise, one could design the causal models for the PQPS up to some year beyond 2025, in order to ensure the causality that results from cryptographic transitions such as the one taking place in the (near) future on the scheme.

With regard to practical implementation, there should be focus on three feasible strategies. First, creating common cause feature libraries would help adoption across industries because it would minimize implementation efforts. A second possibility is that design of visualization tools specific to the needs of fraud investigators could enhance the usability of the outputs of causal analysis. Third, there would be a shared causal graph repository for established fraud patterns so that counter-fraud measures could be shared across financial institutions.

The study also acknowledges some ethical issues for future development. As causal models accumulate, continuous fairness testing is needed to ensure they are not biased towards or are discriminatory against a certain demographic. Techniques such as causal fairness constraints — which model bias pathways and directly mitigate them — need to be part and parcel of system updates. Moreover, we will need to find ways to detect in a privacy respecting manner and future iterations will have to address this issue, especially in jurisdictions with strong data protection laws.

Industry academia partnership will be vital for the development of the field. Collaborative research efforts can perhaps concentrate on generating standardized benchmark datasets for causal detection and exploitation assessment, as at present such standard evaluation frameworks are lacking. There might also be value for universities to invest in specialized curricula for training analysts on how to read causal diagrams and logic of possible worlds in terms of bridging the skills gap as these methods become more widespread.

Looking forward, the break down of silos between digital payment ecosystems across sectors highlights the importance of a universal anti-fraud model. The causal inference approach offers a fundamental methodology that is transferable to such an inter-related eco-system which is defined to generalize beyond the traditional (sector dependent) fraud schemes. By iterating on these strategies to confront implementation issues, the financial security (as well as other) communities can develop systems that are more resilient and proactive than their adversaries.

## COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

## REFERENCES

[1] Talaat M F, Medhat T, Shaban M W. Precise fraud detection and risk management with explainable artificial intelligence. Neural Computing and Applications, 2025, (prepublish): 1-31.

[2] Hasan M, Rahman S M, Chowdhury M J M, et al. CNN Based Deep Learning Modeling with Explainability Analysis for Detecting Fraudulent Blockchain Transactions. Cyber Security and Applications, 2025, 3: 100101-100101.

[3] Karnavou E, Cascavilla G, Marcelino G, et al. I know you’re a fraud: Uncovering illicit activity in a Greek bank transactions with unsupervised learning. Expert Systems With Applications, 2025, 288: 128148-128148.

[4] Kiely E, Swirak K. The UK Carer's Allowance overpayments saga: Structural violence in digital welfare state administration? European Journal of Social Security, 2025, 27(2): 121-138.

[5] Gupta K R, Hassan A, Majhi K S, et al. Enhanced framework for credit card fraud detection using robust feature selection and a stacking ensemble model approach. Results in Engineering, 2025, 26: 105084-105084.

[6] Chakraborty D. Bill safe: intelligent forgery detection with CNN and upgrade sand cat swarm optimization. Discover Computing, 2025, 28(1): 84-84.

[7] Azamuke D, Katarahweire M, Bainomugisha E. A labeled synthetic mobile money transaction dataset. Data in Brief, 2025, 60: 111534-111534.

[8] Lingeswari R, Brindha S. Online payments fraud prediction using optimized genetic algorithm based feature extraction and modified loss with XG boost algorithm for classification. Swarm and Evolutionary Computation, 2025, 95: 101934-101934.

[9] Manta O, Vasile V, Rusu E. Banking Transformation Through FinTech and the Integration of Artificial Intelligence in Payments. FinTech, 2025, 4(2): 13-13.

[10] Reddy S S, Amrutha K, Gupta M V, et al. Optimizing Hyperparameters for Credit Card Fraud Detection with Nature-Inspired Metaheuristic Algorithms in Machine Learning. Journal of The Institution of Engineers (India): Series B, 2025, (prepublish): 1-26.

[11] Lokanan E M, Maddhesia V. Supply chain fraud prediction with machine learning and artificial intelligence. International Journal of Production Research, 2025, 63(1): 286-313.

[12] Kumar K B, Krishnarao A K, Amala S, et al. Cybersecurity Measures in Financial Institutions Protecting Sensitive Data from Emerging Threats and Vulnerabilities. ITM Web of Conferences, 2025, 76.

[13] Wang H, Lu T, Zhang Y, et al. Last digit tendency: Lucky numbers and psychological rounding in mobile transactions. Fundamental Research, 2025, 5(1): 370-378.

[14] Ulloa S D, Luna I D G, Romero M R J. A Temporal Graph Network Algorithm for Detecting Fraudulent Transactions on Online Payment Platforms. Algorithms, 2024, 17(12): 552-552.

[15] Laxman V, Ramesh N, Prakash J K S, et al. Emerging threats in digital payment and financial crime: A bibliometric review. Journal of Digital Economy, 2024, 3: 205-222.