International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(5s) 2026 pp. 434-451
https://doi.org/10.70917/ijcisim-2026-2738
Article

Causal Machine Learning for Financial Crime Attribution: Uncovering Hidden Cause-and-Effect Relationships Among Fraud, Money Laundering, and Cybersecurity Incidents in Blockchain Ecosystems

Md Sazzad Hossain[1], Md Khalilor Rahman[2], Sayem Ul Haque[3], Kazi Abu Jahed[4], Dil Tabassum Subha[5], Md Fazlul Huq Mithu[6], Mahuma Akter[7], Sumaiyara Islam Oysee[8], Md Redwan Laskar[9] and Md Mosheur Rahman[10]

[1,2,3,4]Doctor of Business Administration, University of the Potomac, Washington DC, USA
[5]Master of Science in Business Analytics, Grand Canyon University, USA
[6]MS in Finance, Stony Brook University
[7]Master's in cyber security, Washington University science and technology, Alexandria, VA, USA
[8]Master of Business Analytics, Gannon University, Erie, PA, USA
[9]Master's in Information Science, Trine University
[10]Bachelor of Science and Applied Science, Youngstown State University, Youngstown, Ohio, USA
Corresponding Author: Md Sazzad Hossain, **Email:** sazzadsml2017@gmail.com

Abstract: Blockchain ecosystems have transformed digital finance by enabling decentralized and transparent transactions, yet these same characteristics have facilitated the emergence of complex financial crimes involving fraud, money laundering, and cyber-enabled illicit activities. Existing machine learning approaches have achieved notable success in detecting suspicious transactions, but most remain fundamentally predictive and provide limited understanding of the underlying mechanisms that generate criminal behavior. This limitation constrains the development of effective interventions, regulatory policies, and financial intelligence strategies. This study proposes a causal machine learning framework for financial crime attribution in blockchain ecosystems, emphasizing the discovery of hidden cause-and-effect relationships among fraudulent activities, laundering processes, and cybersecurity incidents. Using the Elliptic Bitcoin transaction network as a case study, the framework integrates predictive modeling, graph-based feature engineering, explainable artificial intelligence, causal discovery techniques, Double Machine Learning, and heterogeneous treatment-effect estimation within a unified analytical pipeline. The methodology combines temporal validation procedures with structural analyses to investigate how network characteristics and behavioral patterns contribute to illicit financial outcomes. The findings demonstrate that strong predictive performance does not necessarily imply causal importance. While conventional machine learning models effectively identify suspicious transactions, causal analyses reveal that several highly predictive variables possess limited independent influence after accounting for network and temporal confounders. Structural properties associated with community organization, transaction dynamics, and connectivity emerge as important components of the broader mechanisms underlying financial crime propagation. The results further highlight the interconnected nature of fraud, money laundering, and cyber-enabled activities within decentralized financial systems. This research contributes to the growing intersection of causal inference and blockchain analytics by explicitly distinguishing predictive explanations from causal explanations and by providing a transparent framework for intervention-oriented financial intelligence. Although the study is constrained by observational data, proxy-based treatments, and platform-specific characteristics, it establishes a practical foundation for developing more interpretable, accountable, and causally informed approaches to combating financial crimes in digital economies.

Keywords: Causal Machine Learning; Financial Crime Attribution; Blockchain Analytics; Bitcoin; Double Machine Learning; Causal Forests; Explainable AI; SHAP; Money Laundering; Cybersecurity Incidents

1. Introduction

1.1 Background, Motivation, and Problem Statement

The rise of blockchain has completely changed how digital finance works by letting people transfer value peer-to-peer without needing regular banks. Trust gets built through cryptography and a shared network rather than a main institution. Nakamoto (2008) kicked things off by introducing Bitcoin as an electronic cash system that stops double spending and secures transactions through a public ledger, setting up the whole modern cryptocurrency world [16]. Since then, this tech has grown way past just digital money. It now shapes decentralized finance, international payments, smart contracts, and token systems. But as these networks grew bigger and more complicated, they opened up massive room for innovation while giving criminals brand-new ways to manipulate the system. Public ledgers show a huge amount of transaction data, but the fact that users operate under pseudonyms makes it incredibly hard to figure out who is actually behind a financial crime. Illicit actors use this decentralized setup to run scams, hide where dirty money came from, and coordinate across borders. The old ways of tracking money laundering and monitoring finances were built for traditional banks, so they often fail to keep up with how fast cryptocurrency markets shift and connect. Because of this, it is clear that tracking tools need to look at the deeper structural setups inside transaction networks instead of just flagging bad events after they already happened.

The fast growth of these crypto spaces also brings up big questions about using standard machine learning for financial intelligence. A lot of current systems only care about raw predictive accuracy or sorting data into clean buckets, but they ignore the actual reasons why people behave deceptively. This means a model might successfully separate a clean transaction from a dirty one without explaining why that pattern showed up in the first place or how to stop it. Regulators, compliance officers, and security teams need clear, practical insights to make solid decisions and keep governance transparent. Knowing the actual cause behind an action is turning out to be just as vital as getting a high accuracy score. Zohar (2015) pointed out that the internal setup of Bitcoin, its rewards, and its decentralized checks create a complex loop of human and technical behavior that cannot be figured out by just looking at simple transaction lists [27]. The relationships running between users, crypto exchanges, miners, wallets, and funds mean that analysis has to capture the shifting structures and unexpected behaviors. The core issue here is that pinpointing financial crimes on a blockchain is tough because current tools lean on statistical coincidences instead of real cause-and-effect lines. This study tackles that gap by mixing predictive analytics, network mapping, explainable AI, and causal machine learning to get a much better look at how specific transaction habits lead to illegal activities.

1.2 Research Questions and Objectives

The expansion of cryptocurrency has shown that fraud, money laundering, and cyberattacks are deeply tied together. These activities do not happen in a vacuum. They feed into each other through messy transactional webs that help criminals move, hide, and split up stolen cash. A basic fraud scheme creates profits that need laundering, while cyber incidents like ransomware, exchange hacks, and token thefts send massive waves of illegal capital moving through the blockchain. Spotting these connections requires tools that can read both the structure of the network and how things change over time. Foley et al. (2019) calculated that a huge chunk of early Bitcoin activity was tied directly to illegal markets and criminal groups, proving that digital assets have historically been a major tool for illicit economies [8]. Their work shows why there is a pressing need for smarter ways to track how financial crimes spread across these decentralized setups. Cyber threats are a major piece of this puzzle. Ransomware gangs, phishing scams, and exchange breaches constantly use crypto to collect payouts or wash their funds, linking cyber incidents directly to bad financial flows. Akcora et al. (2020) showed that the specific shape and patterns of Bitcoin transaction networks can help predict ransomware movements, which highlights how useful network structure is for mapping out these threats [2]. This makes it obvious that real financial attribution needs a framework that ties together transaction records, timing, and network maps while keeping up with how criminal habits evolve.

The main reason to bring in causal machine learning is the massive difference between predicting something and actually explaining it. Standard predictive models find statistical patterns that help sort data, but those patterns do not show what happens if an investigator steps in, nor do they reveal the root cause of the behavior. Pearl (2009) built a strict framework for causal reasoning using structural models, what-if scenarios, and intervention tracking, showing that finding the real cause requires completely different questions and methods than regular statistics [17]. Using these

ideas for blockchain data lets investigators stop just flagging weird transactions and start seeing how specific network habits, privacy features, and transfer patterns drive illegal acts. With that in mind, this study looks at a few closely related questions about tracking down financial crimes in blockchain networks. The work checks out which transaction and network details give the strongest hints about illegal acts, what cause-and-effect lines connect privacy-seeking behavior to crime outcomes, how effects vary across different transaction types, and where predictive explanations split away from causal ones. The study also tests how stable these relationships stay when changing definitions or running sensitivity checks. To get there, the work focuses on building strong predictive models, creating behavioral markers for privacy habits, mapping out stable causal lines in transaction data, measuring treatment effects, and figuring out how causal machine learning can actually help regulators and financial intelligence teams make better calls.

1.3 Contributions

This study bridges a few different worlds, blockchain analytics, financial intelligence, explainable AI, and causal inference. The main goal is to look at financial crime attribution in a new way, mostly by separating simple predictive patterns from the actual cause-and-effect reasons behind illicit activity. Most past work focused heavily on raw classification accuracy and spotting anomalies, which left a big gap. Not enough time was spent figuring out how transaction habits, network shapes, and anonymity features actually drive criminal outcomes. This project pulls predictive machine learning, explainable AI, causal discovery, and treatment-effect estimation into one reproducible analysis pipeline to fix that. A big part of the work comes down to applying causal machine learning directly to blockchain data. The framework uses Double Machine Learning to calculate treatment effects while keeping high-dimensional confounding variables from messing up the results. Back in 2018, Chernozhukov and a team of researchers showed that double and debiased machine learning works well for finding structural parameters [6]. By combining flexible machine learning models with orthogonalization and cross-fitting, their approach stops regularization bias from distorting the math [6]. This study takes those foundational ideas and pushes them into blockchain financial intelligence, testing how well they work for decoding criminal transactions.

Another piece of the puzzle is how the study builds and tests behavioral anonymity proxies out of transaction-network data. Because real-world data on mixers or privacy-focused services is usually impossible to get, the study creates readable indicators based on structural and timing patterns instead. They aren't perfect, direct observations of anonymity tools, but they give investigators a practical way to explore treatment effects and causal links, so long as the assumptions and limits are kept clear. Running extra sensitivity analyses helps ensure the conclusions actually hold up under scrutiny. The study also looks at the differences between predictive and causal explanations, showing that feature importance scores from explainable AI tools shouldn't be confused with actual causal impact. By putting SHAP-based predictive interpretations right next to causal discovery outputs and treatment-effect estimates, it becomes easier to see how different analytical approaches serve financial intelligence. This matters for regulators and compliance teams who need real, actionable insights rather than just high accuracy scores on a test set. Lastly, the work focuses heavily on reproducibility and open science. It relies entirely on public datasets, lays out all assumptions clearly, and uses temporal validation to make sure future data doesn't leak into past models. With clear uncertainty mapping and deep robustness checks, the project tries to move blockchain crime research forward, offering a practical way to dig up the hidden causes behind fraud, money laundering, and hacks in decentralized finance.

2. Literature Review

2.1 Financial Crime Detection, Cybersecurity Threats, and Illicit Financial Flows in Blockchain Ecosystems

Blockchain technologies changed how money moves around, but they also opened up a whole new sandbox for criminal activities that don't care about traditional borders. The fact that blockchain ledgers are public and can't be changed gives us a lot of transparency, but things like fake names and decentralized setups make it really hard for regular financial watchdogs to keep up. Early work in this field mostly just tracked transactions and looked for weird outliers. Lately, though, studies look way more at the actual shape of the network and how users behave. Weber et al. (2019) showed that graph convolutional networks make anti-money laundering work a lot better by looking at how Bitcoin transactions connect to each other [25]. That work basically proved that financial forensics gets a lot more out of mapping things as a network instead of looking at single transactions by themselves. It laid down a foundation for putting network science into blockchain analytics and showed why we need to focus on structural learning to spot bad actors.

Money laundering in crypto is a massive headache because criminals use decentralized setups to hide where money came from, moving dirty cash across tons of different wallets and services. Hu et al. (2021) broke down how money laundering works on the Bitcoin network by looking at transaction shapes and behavior models [9]. It turns out laundering operations usually leave specific structural fingerprints, like splitting up money fast or moving it along coordinated paths. This means finding financial crime on a blockchain requires looking at the bigger picture of the network, not just flag-checking isolated transactions. This fits right into modern anti-money laundering goals, which try to map out whole connected ecosystems instead of chasing single suspicious events. Then there is the crossover between cybersecurity hacks and illegal money, which makes it even harder to figure out who is behind blockchain crimes. Cyberattacks use crypto to get paid, store value, or launder money, creating a massive loop between digital security gaps and financial crimes. Aashish et al. (2025) showed that machine learning anomaly detectors can spot cybersecurity threats while keeping an eye on energy and carbon footprints, proving that smart monitoring can handle multiple jobs at once [1]. That research shows how advanced analytics helps with security, and the same ideas can easily be applied to tracking blockchain crimes.

Cross-chain tech and decentralized bridges make things even messier because they let assets hop between totally different blockchain networks without much oversight. Shawon et al. (2025) built behavioral machine learning models to spot illegal bridge-based laundering [22]. It turns out cross-chain transactions have clear patterns that make it easier for criminals to move and hide stolen funds. Because of this, tracking financial crime now means looking at a big, messy web where fraud, hacking, and laundering all feed into each other through fast-moving transaction setups. Basically, the current research shows that forensic analytics, behavior tracking, and cybersecurity all need to work together to understand how crime evolves in crypto.

2.2 Machine Learning Approaches to Financial Crime Detection

Machine learning has become the main tool for fighting financial crime because it can dig through massive datasets, find hidden patterns, and adjust when criminals change tactics. Old-school statistical tools like logistic regression are still used because it is easy to see how they make decisions, which matters a lot in highly regulated industries where rules require clear proof. Ensemble tools, like random forests and gradient boosting, stepped things up by catching tricky, non-linear relationships in data and keeping models from just memorizing training sets. These models do great with fraud, money laundering, and security analytics, but most of them still focus entirely on guessing what happens next without figuring out why things happen. Lately, machine learning has shifted toward network-heavy models and deep learning to catch complex links across financial systems. Islam et al. built graph neural network structures to predict big financial risks, showing that network shapes matter when tracking how trouble spreads between regular banks and crypto markets [11]. That work proved that graph models catch connections that standard feature lists miss, which is why network learning is becoming a big deal in finance. The same shift is happening in blockchain forensics, where transaction graphs give away a lot of info about user communities, money flows, and weird structural blind spots.

Deep learning is also used a lot for predictive maintenance, finding anomalies, and spotting failures in big industrial setups, and these methods carry over well to financial intelligence. Alam et al. (2026) put together hybrid deep learning setups to predict when industrial equipment will break down, showing that mixing different neural parts helps models handle messy, noisy data [3]. Even though that study looked at factories instead of finance, the core ideas about combining features, learning non-linear patterns, and tracking changes over time apply perfectly to blockchain data. Still, even with these massive improvements in prediction, most machine learning literature just cares about classification accuracy and does not help much with figuring out actual interventions or cause-and-effect relationships. The big flaw across all these models, whether traditional, ensemble, graph, or deep learning, is that they cannot tell the difference between things that just happen to occur together and actual cause and effect. High predictive accuracy does not mean a model has found something an investigator can actually change to fix a problem, and it does not mean the model will hold up if the environment changes. Investigators and regulators need frameworks that explain why bad behavior happens and how specific changes might disrupt criminal networks. This is why there is a push to bring causal machine learning into the mix, giving predictive tools better explanations and a solid logical base for making decisions.

2.3 Explainable Artificial Intelligence and Causal Machine Learning in Financial Services

Using artificial intelligence in high-stakes financial systems has made people demand far more transparency, accountability, and clear human understanding. Explainable artificial intelligence grew into a major research focus to solve the issue of opaque algorithms and to build real trust among regulators, workers, and everyone affected by these

models. A framework called LIME came out from Ribeiro et al. (2016) as a model-agnostic setup that explains individual predictions using local approximations, which lets people see how complex classifiers work without losing predictive flexibility [21]. This work laid down foundational ideas for post-hoc interpretability and changed how people approached responsible artificial intelligence later on. Better explainability also brought around unified frameworks to measure how much different features matter across various kinds of models. Lundberg and Lee (2017) introduced SHAP values, using a grounded approach from cooperative game theory to give consistent and locally accurate explanations for what a model predicts [13]. People in financial analytics use SHAP quite a bit now since it helps with both global and local interpretations of predictive models. Still, it is a mistake to take predictive importance measures as proof of a causal link. Features that help a lot with prediction accuracy might just show correlations caused by confounding variables, rather than pointing to real mechanisms that could guide actual interventions or policy choices.

Wider conceptual ideas have kept the focus on making explainability responsible and centered on humans. A detailed taxonomy of explainable artificial intelligence came from Arrieta et al. (2020), mapping out opportunities, difficulties, and the technical differences between intrinsic, post-hoc, and interactive explanation styles [4]. Their analysis shows why explainability techniques must match the specific needs of a field, especially where ethical choices and regulators matter a lot. Along similar lines, Molnar (2022) put together practical ways to handle interpretable machine learning, pointing to counterfactual explanations, feature importance analyses, and model transparency strategies as tools needed for trustworthy artificial intelligence [15]. Taken together, these ideas show that explainability is a multi-sided goal that goes way past just making charts, aiming instead for true human understanding and accountability.

While explainable artificial intelligence deals with model transparency, causal machine learning looks at what happens during interventions and tries to find the underlying structural mechanisms behind what is observed. Spirtes et al. (2000) set up major foundations for causal discovery, using graphical models and search steps to figure out plausible causal structures from observational data under specific assumptions [23]. Their framework still shapes how people handle high-dimensional systems and complex dependencies today. Moving forward with causal inference principles, Wager and Athey (2018) brought in causal forests to estimate heterogeneous treatment effects, letting researchers see how interventions change different subgroups instead of just looking at average effects [24]. Then, the NOTEARS framework from Zheng et al. (2018) turned directed acyclic graph learning into a continuous optimization problem, which helped scale up structure learning for complex datasets [26]. These shifts show how explainable artificial intelligence and causal machine learning are coming together as twin paths to understand, interpret, and deal with complex financial systems.

2.4 Research Gaps and Motivation for the Present Study

Even with major progress in blockchain forensics, explainable artificial intelligence, and causal inference methods, big research gaps still sit right where these fields meet. Current financial intelligence systems put almost all their weight on predicting things, leaving a lot of gaps in understanding the structural mechanisms that actually create illegal behavior. Plenty of setups flag suspicious activity without making it clear if the relationships they find are points where someone can intervene or just statistical patterns that happen to show up together. Because of this, policymakers and investigators often go without clear evidence on how changing specific behaviors might alter criminal outcomes in decentralized financial networks. Setting up early warning systems has shown how helpful it is to mix real-time signals with machine learning for proactive risk checks. Frameworks run on machine learning came from Reza et al. (2025), using dynamic digital indicators to spot financial distress early and showing that timely info and adaptive models matter a lot in messy economic settings [20]. Even though that study looks at financial trouble instead of blockchain crime, it points to a wider chance to bring forward-looking, intervention-focused methods into financial analytics. These ideas have not been explored much in cryptocurrency investigations, where causal reasoning could really boost proactive monitoring and regulatory setups.

Another major blind spot is finding hidden organizational structures and people working together inside financial systems. It was shown by Dola et al. (2024) that machine learning tools can uncover hidden collusion networks in corporate finance, which proves that relational analytics can pull up interactions that stay completely invisible during regular transaction checks [7]. Blockchain networks have the same issues, where coordinated laundering, cybercrime groups, and fraud networks often use distributed setups to slip past standard detection tricks. Yet, very few studies look directly at the causal relationships that drive these connected behaviors. The existing literature shows a lack of connection between causal inference methods and blockchain analytics, very few studies estimating treatment effects for financial crime attribution, a messy boundary between predictive explanations and

causal mechanisms, and not enough robustness analysis on behavioral ideas linked to anonymity. On top of that, cybersecurity trouble, money laundering, and fraud are treated as separate issues most of the time, even though they are tied tightly together inside cryptocurrency networks. This study tackles those gaps by building a single framework that hooks up predictive machine learning, explainable artificial intelligence, causal discovery, and heterogeneous treatment-effect estimation to find the actual cause-and-effect lines running through financial crimes in blockchain spaces.

3. Methodology

3.1 Research Design, Dataset, and Experimental Setting

The setup relies on an observational causal machine learning design to look for believable cause-and-effect patterns behind financial crimes on the blockchain. Blockchain transactions do not come with neat, controlled interventions, confirmed ways to stay anonymous, or neat experimental groups. Because of that, any causal claims rest on specific assumptions about things like ignorability, consistency, positivity, and having no hidden variables muddying the waters. The focus stays on finding plausible connections rather than claiming definitive proof, which matches up with how observational causal inference usually works. The actual analysis ran on the Elliptic Bitcoin dataset. It has transaction features, time data, labels from experts, and a directed graph showing how transactions connect. Three main files did the heavy lifting: transaction features, class labels, and the edge lists. Automated checks made sure the files existed, weren't empty, and downloaded properly from public storage. When local copies were missing, the setup pulled data directly from public mirrors to keep everything repeatable. The data splits into licit, illicit, and unknown groups. The unknown ones got dropped before the predictive and causal steps to keep the focus on a clean binary outcome.

Everything ran inside a Google Colab environment using standard Python libraries like NumPy, Pandas, Scikit-learn, NetworkX, XGBoost, SHAP, EconML, and causal-learn. Random seeds were locked down across Python, NumPy, and the model frameworks, so anyone else running it gets the same results. The code automatically saved everything, outputs, charts, causal graphs, feature importance lists, sensitivity checks, and basic stats, straight into designated folders. It keeps things clear and easy to double-check or build on later. It is worth noting that the Elliptic data does not have verified mixer usage, clear laundering paths, or direct tech-breach indicators. Those concepts are tracked using behavioral proxies built from network shapes and timing patterns. Time order was baked into the setup to stop information from leaking into the past and keep things realistic. Steps one through thirty-four built the training set. Steps thirty-five through forty-nine served as the testing ground. If future chunks were missing, a standard seventy-thirty split based on time took over. This mirrors real financial intelligence work, where systems have to spot future bad behavior using only old data. Using time-based validation protects the internal logic and avoids the issues that come with random splits on shifting transaction networks.

3.2 Data Preparation, Network Construction, and Feature Engineering

Preprocessing put the transaction features, labels, and network edges into one place. Missing labels for unknown transactions were removed. After that, binary markers were set up by giving illicit transactions a one and licit ones a zero. Math tweaks, log scaling, and normalizations smoothed out the skewed parts and made everything stable enough for the machine learning and causal steps. The transaction network became a directed graph where nodes represent transactions and edges show the cash moving between them. NetworkX calculated different graph metrics to map out where transactions sat and how they acted in the broader system. The engineered variables covered in-degree, out-degree, total degree, PageRank centrality, average neighbor degree, community sizes, clustering metrics, and approximate betweenness centrality. Exact betweenness took too much computing power on large graphs, so the approximations offered a good balance between useful detail and speed. Log transformations helped calm down the extreme distributions that pop up in financial networks. Adding time features brought in more details about how transactions move. Velocity metrics tracked how fast transactions happened across intervals. Rolling illicit rates gave a quick look at recent behavior in close neighborhoods. Time-step normalization changed time indexes into numbers between zero and one, making them easier to feed into the algorithms and causal steps. Velocity variables also got log-transformed to level out the variances and make the numbers easier to interpret.

Three behavioral proxies pointed to hidden anonymity and laundering habits. The anonymous transaction proxy caught items with high degree centrality, high approximate betweenness, and fast transaction velocity all at once. Cutoffs came from the data itself, using the eightieth percentile for degree, the seventieth for betweenness, and the sixtieth for velocity. A rapid redistribution proxy flagged transactions that sent out more money than they took in during earlier moments, signaling potential fund spreading. Community-hopping turned up via a combined indicator

tracking high average neighbor degrees inside large communities. These variables are not confirmed mixer steps or definite laundering tracks. They are behavioral approximations meant to assist the causal search under clear rules and known boundaries. The final feature set blended the original Elliptic traits with the new network and time variables. Baseline details straight from the source got mixed with in-degree, out-degree, PageRank, approximate betweenness, community metrics, clustering stats, transaction speeds, rolling illicit numbers, normalized time, and the behavioral proxies. Mixing them lets the system catch both the basic transaction details and the wider network relationships happening across the blockchain.

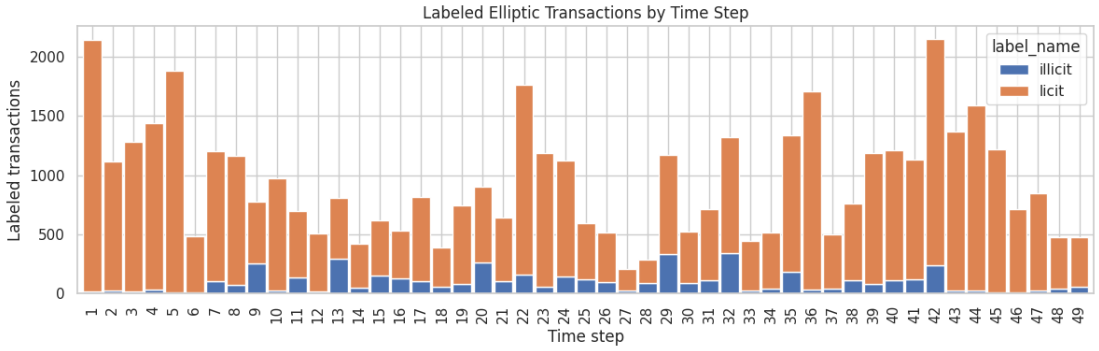

Fig.1: Elliptic transactions by time step

3.3 Predictive Machine Learning and Explainable Artificial Intelligence Framework

The predictive setup in this experiment used three different kinds of machine learning models: a linear approach, an ensemble method, and a gradient boosting system. Logistic regression worked as a basic baseline because it easily shows straight-line relationships between transaction details and illegal activity. Next, random forest classifiers helped catch trickier, non-linear connections and overlapping effects by blending multiple decision trees together. Finally, XGBoost offered a heavy-duty gradient boosting setup that is specifically tuned to handle highly uneven classification tasks. Every model was trained using fixed random seeds and balanced learning methods to deal with the fact that the Elliptic dataset has way more normal transactions than illegal ones. The evaluation process focused on metrics that actually make sense for financial crime, where the data is always heavily skewed. This meant tracking precision, recall, the F1-score, balanced accuracy, the Matthews correlation coefficient, the ROC-AUC, and the precision-recall AUC. That last one got the most attention because precision-recall curves give a much clearer picture than standard ROC curves when you are dealing with massive class imbalances. On top of that, confusion matrices were mapped out for the top-performing models to make it easier to see how they behave in practical scenarios.

To get a look inside these models, explainable AI techniques were brought in to explain the predictions on both a broad and local level. SHAP analyses were run using the TreeExplainer tool specifically on the XGBoost classifier. Because running these calculations on everything takes too much computing power, the process used smaller, representative samples from the test data instead, keeping things reproducible with fixed random states. From there, mean absolute SHAP values were calculated to rank which features mattered most globally and see which variables really drive the model to flag a transaction as illegal. One crucial point to keep in mind here is the difference between explaining a prediction and finding a real cause. SHAP values show how much a feature alters a model's guess, not what would happen if someone stepped in and changed things in the real world. Because of this, just because a feature is highly influential does not mean it is a root cause. This matters a lot in financial intelligence, where actual interventions and regulations need to be based on why illegal behavior happens, not just on random clues that happen to correlate with it. Comparing these predictive explanations against actual causal estimates is a core part of this setup, helping to fix a common gap in current blockchain research.

3.4 Causal Discovery and Structural Relationship Identification

The causal discovery part focused on a stripped-down set of thirteen variables that cover structural, timing, and behavioral traits tied to financial crime. This smaller selection was chosen to keep things easy to understand, fast to compute, and theoretically sound, while still being complex enough to show real interactions. The analysis looked at degree measures, PageRank scores, betweenness approximations, average neighbor statistics, community markers, transaction speeds, normalized time data, behavioral proxies, and the final illegal outcome variable. To keep the computing demands reasonable, the causal analysis was limited to four thousand observations using stratified

sampling to keep the ratio of classes intact. This made the structure-learning phase highly robust without dragging down performance or creating massive computational bottlenecks. All the causal variables were standardized before running the discovery tools so that data on completely different scales could be compared fairly.

Two different strategies were used to map out these causal structures. The first was a PC-style approximation that used partial correlation skeletons alongside strict rules about timing. Partial correlation matrices were built using inverse covariance estimates, and any connections that passed a set threshold were oriented based on logical assumptions about time. Basically, variables tracking time and transaction movement were forced to come before later behaviors and the final illegal outcomes, keeping the resulting graphs chronologically sensible. The second method used a sparse directed acyclic graph framework inspired by the NOTEARS approach, running sequential LASSO regressions. Each variable was modeled using the ones that logically came before it, and any significant coefficients were marked down as potential causal links. This approach kept the graphs sparse, highlighting only the strongest relationships while staying easy to scale computationally. Later, an edge overlap analysis compared the PC-style and NOTEARS-style results to find the most stable relationships that showed up across both methods.

More specialized causal discovery tools were also brought in, running the full PC algorithm through the causal-learn library. Fisher's Z conditional independence tests and stability-focused searches generated extra graphs to compare against the local approximations. Using a mix of discovery methods helps smooth out the biases of any single algorithm and makes the matching results much more trustworthy. Even so, all these discovered structures rely heavily on assumptions like causal sufficiency, faithfulness, and the idea that no hidden variables are messing with the data. Because of that, these connections should be viewed as likely theories that need more testing, rather than absolute proof of how criminal networks operate.

3.5 Double Machine Learning and Treatment Effect Estimation

This causal inference part of the study treated anonymous transaction behavior as the main treatment variable, while looking at illicit transaction status as the outcome. The anonymity proxy itself was built from network centrality metrics and transaction velocity details. It functioned as a behavioral workaround to capture anonymity-leaning actions, rather than serving as direct proof that someone used a mixer or privacy-focused tools. The analysis factored in a wide mix of confounding variables, including in-degree, out-degree, PageRank scores, estimated betweenness, average neighbor metrics, community size, transaction speeds, timestamps, quick asset redistribution, and community-hopping habits. The study used Double Machine Learning with cross-fitting to calculate average treatment effects. This was done to cut down on the regularization bias that usually pops up with flexible machine learning models. Stratified K-fold methods split the data into a few different chunks, which made it possible to calculate nuisance functions independently and then orthogonalize the treatment effects. From there, the leftover residuals from the treatment and outcome variables were used to get an unbiased look at how much anonymity-related behaviors actually shift the probability of an illicit transaction. To get a better handle on the uncertainty and back up the statistical claims, bootstrap methods were used to build confidence intervals around those estimated treatment effects.

For the actual setup, the project relied on the LinearDML framework from the EconML package. Random forest models handled the treatment and outcome tracking, and orthogonalization steps pulled the real causal effects away from all the high-dimensional confounding factors. Once the average treatment effects and their confidence intervals were ready, they were saved directly as experimental files to keep things reproducible. Using these steps added some extra rigor to the methodology, bringing the whole setup in line with how causal machine learning research is handled nowadays.

Of course, these estimated treatment effects only hold up if a few key assumptions are true. First, there is ignorability, meaning every single relevant confounder that touches both the treatment and the outcome has to be measured and thrown into the model. Then there is positivity, which means every observation needs a realistic, non-zero chance of being in either the treated or untreated group, regardless of its traits. Consistency is another big one—it requires that the outcomes we see actually match up with the potential outcomes of the treatment state that happened, while the lack of hidden confounding assumes no unmeasured variables are pulling the strings behind the scenes for both treatment and outcome. Because blockchain transaction data is strictly observational, verifying these assumptions directly is not really possible. Because of that, the treatment effects should be viewed as reasonable estimates that depend entirely on the theoretical setup and the sensitivity tests.

3.6 Heterogeneous Treatment Effects and Sensitivity Analysis

Going beyond simple average effects, the project also looked into how these anonymity behaviors might play out differently depending on the specific transaction context. A T-learner framework, running separate random forest regressors for the treated and untreated data points, calculated individual treatment effects for each transaction. By predicting what the potential outcomes would look like under both scenarios, it was possible to find the individual differences between the treated and untreated paths. The overall spread of these varied impacts across the network was then mapped out using basic summary statistics like the mean, median, and specific percentiles. To get a more detailed look at these varied effects, the analysis also brought in the CausalForestDML framework from EconML. This meant setting up separate random forest models for the treatment and outcome tracks, while causal forests mapped out the conditional average treatment effects across the different variable spaces. The settings were dialed in with two hundred trees, specific depth limits, cross-validation, and minimum leaf sizes to keep the model flexible without losing statistical stability. The resulting spread of effects showed exactly how anonymous behaviors hit different pockets of blockchain transactions in different ways, bringing out structural details that get completely buried when you only look at the overall average.

Sensitivity tests were also a major part of the setup, mostly to deal with potential pushback about choosing arbitrary cutoffs for what counts as anonymous. The model was re-tested using a range of alternative anonymity thresholds, specifically shifting through the 70th to 85th percentiles. For every single one of these threshold changes, the treatment groups were rebuilt from scratch, the Double Machine Learning models were re-run, and the new average treatment effects were tracked. Seeing stable estimates across all these different cutoffs makes the causal takeaways much more believable and shows the results do not just drop off if the settings change a little bit. Finally, the framework took the predictive importance scores from SHAP analysis and stacked them up against the causal importance scores from the treatment-effect models. Feature importances from the treated and untreated random forest setups were pooled together and compared directly to the predictive outputs. This side-by-side view helps highlight the clear difference between simple correlation patterns and actual intervention-based logic. By combining these varied effect analyses, robustness checks, and comparisons between prediction and causation, the methodology sets up a solid base for tracking financial crime back to its source inside blockchain networks.

3.7 Evaluation Metrics, Reproducibility, and Ethical Considerations

Checking how well the model worked required looking at two completely different things: how good it is at guessing labels and how accurate it is at finding actual cause-and-effect relationships. For the guessing part, the standard tests did the heavy lifting. This included precision, recall, F1-score, balanced accuracy, Matthews correlation coefficients, ROC-AUC, and PR-AUC. The causal side of things needed a different toolkit entirely, focusing on average treatment effects, confidence intervals, heterogeneous treatment summaries, how stable the causal edges stayed, and overlap analyses across the different discovery methods. Splitting the evaluation like this matters because a model can be incredibly good at predicting an outcome without having any real grip on why things are happening. The two goals are pieces of the same puzzle, but they are not the same thing.

Making sure other people could run the same tests and get the same results was a priority from the start. To keep things from drifting, fixed random seeds locked down every random process across Python, NumPy, Scikit-learn, and the specific causal libraries. Every piece of data the model spat out, feature importance tables, causal graphs, treatment-effect estimates, sensitivity analyses, and general experiment summaries, was saved right away as a portable file. Because the datasets are public and the software is open-source, anyone can download them and check the work. To make things even simpler, the package installations inside Google Colab were automated, so the setup stays uniform no matter whose computer is running it.

When dealing with financial crime analytics, ethics cannot be an afterthought. Automated systems end up steering regulatory choices, compliance rules, and where investigators look first. Because of that, this work stays strictly within the boundaries of public, fully anonymized transaction data, avoiding issues with personal identity or privacy leaks. Even so, a statistical framework cannot prove someone committed a crime or show what their intent was. Any behavioral signs that look like someone is trying to stay anonymous are treated with a lot of caution, and they are not lumped in with confirmed money laundering or mixer usage. People still need to make the final calls in the real world. The causal estimates are just pieces of supporting evidence, not absolute proof. Keeping these boundaries clear ensures the machine learning is used responsibly, with an emphasis on transparency and knowing what the tools can and cannot do.

4. Experimental Results and Evaluation

4.1 Dataset Characteristics, Network Structure, and Exploratory Analysis

The testing used the Elliptic Bitcoin transaction network, which has 203,769 nodes and 234,355 directed edges connecting them. The labels sorted transactions into either licit or illicit groups. Anything with an unknown label got dropped before running the predictive and causal models, so the classification stayed strictly binary. This left a massive dataset that works well as a benchmark for studying how financial crime happens in decentralized systems, mixing transaction data, time elements, and network details. Looking at the timeline showed that transaction volume jumped around quite a bit over the forty-nine recorded time steps. It makes sense given how fast blockchain environments change. Tracking the rolling illicit rates made it possible to see these changes over time and get a clear picture of crime intensity during different parts of the study.

Digging into the network structure showed clear differences between the legal and illegal transactions. The bad transactions usually had lower degree measures and smaller PageRank values than the legitimate ones. This points to illegal operations moving through quiet, isolated pathways instead of relying on major, highly connected hubs. There were forty-nine weakly connected components across the whole graph, showing a lot of broken-up communities that didn't interact much and had different levels of money moving through them. Things like community size, how often accounts hopped between communities, and how funds were redistributed turned out to be incredibly useful for mapping out transaction behavior, helping create good proxies for tracking anonymity. It shows why network science matters for crypto intelligence work. Just looking at raw transactions means missing the structural patterns that actually define bad behavior. The initial look at the data also showed some tight connections among the engineered features. The velocity variables are tied closely to the community metrics and degree statistics. That means transaction speed, where a node sits in the network, and the shape of its immediate neighborhood all work together to shape behavior in this ecosystem. Because these pieces are so tangled up, it made perfect sense to use causal discovery methods to sort through the mess, since basic correlation metrics can't separate these deep interactions. This preliminary groundwork set up a solid starting point for the predictive and causal models that came next, proving that tracking financial crime on the blockchain requires looking at many angles at once.

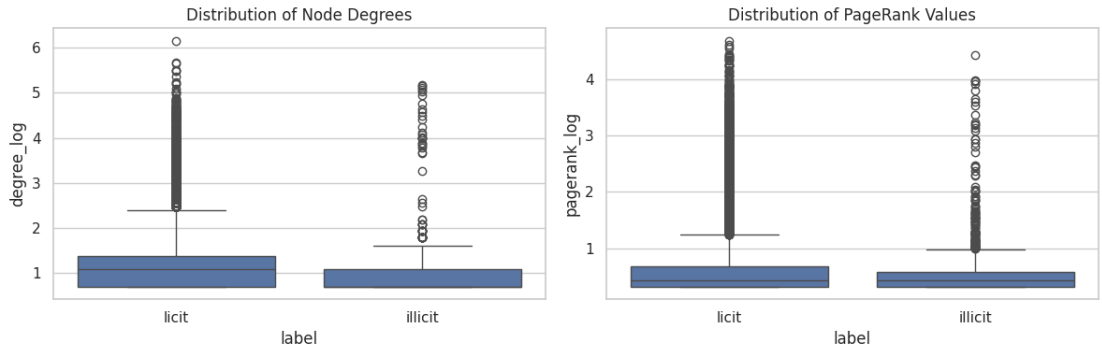

Fig.2: Network exploration of licit and illicit transactions

4.2 Predictive Modeling Performance and SHAP-Based Explanations

For the predictive models, the setup used a split based on time, putting transactions from steps one through thirty-four into the training phase and keeping steps thirty-five through forty-nine for testing. Splitting it this way stopped future data from leaking into the past, mimicking how these intelligence tools actually have to run in the real world. Looking at the numbers, the Random Forest classifier did the best job across the board. It hit a precision score of 0.9259, a recall of 0.7267, an F1-score of 0.8143, and a precision-recall area under the curve of 0.7929. The data proves that ensemble tree models handle the messy, nonlinear connections and structural patterns inside crypto networks quite well, keeping false alarms and missed cases balanced. The XGBoost model caught a lot more than the simple logistic regression setup, pulling in a recall of 0.7590 and a precision-recall area under the curve of 0.7873. It struggled with precision, which dragged down to 0.4308 and left the final F1-score at 0.5496. The settings made the model focus heavily on catching every possible bad transaction, which naturally caused a lot of false alarms. The logistic regression baseline sat at the bottom of the list, showing a poor precision of 0.2632, a recall of 0.6990, an F1-score of 0.3824, and a precision-recall area under the curve of 0.3205. Seeing the linear baseline trail so far behind

makes it obvious that straight-line models just can't handle the tricky, layered relationships found inside blockchain networks.

model	precision	recall	f1	pr_auc
Random Forest	0.925882	0.726685	0.814278	0.792896
XGBoost	0.430818	0.759003	0.549649	0.787262
Logistic Regression	0.263213	0.698984	0.382420	0.320455

Table. 1: Predictive model evaluation results

Using SHAP values helped pull back the curtain on how these models actually made their decisions. The XGBoost setup leaned heavily on the basic, raw transaction attributes, specifically focusing on things labeled feat_001 and feat_052 to spot the bad activity. The custom-built network features, like betweenness_approx and pagerank_log, didn't have nearly as much pulling power as the original transaction metadata. It turns out basic transaction details carry a massive amount of weight for sorting these records, even when throwing a ton of deep network structural data into the mix. Just because the network variables didn't dominate the prediction scoring doesn't mean they lack a deeper causal role, which is exactly why the study moved into causal analysis next.

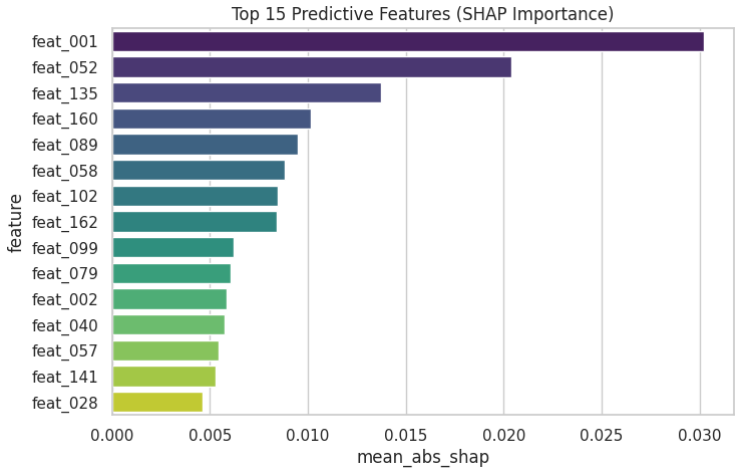

Fig.3: Top 15 predictive features

4.3 Causal Discovery and Structural Relationship Analysis

Running the causal discovery algorithms showed a lot of alignment across different structure-learning methods, giving more weight to the idea that these relationships are stable. The PC-style method mapped out thirty-three directed edges. The NOTEARS-based optimization framework found fifty-two paths. A total of twenty-nine edges showed up in both setups, proving that a core set of structural connections stays steady no matter which math tool is used. Having that much overlap makes the resulting graph structures feel dependable and eases worries about the results just being a fluke of one specific algorithm. Some of the strongest links found in the data tied velocity_log, community_size_log, and degree_log together. It looks like transaction speed, community sizes, and immediate connections all work together to steer how things behave inside the blockchain space. These mixed effects back up the idea that crypto crime comes out of broader structural patterns, not just single transactions happening in a vacuum. The community metrics played a massive role in how money got redistributed and how anonymity proxies behaved, showing that the actual shape of the network heavily dictates or limits illegal financial moves. Using the specialized PC algorithm from the causal-learn setup backed up a few more major relationships. The paths leading from anonymous_proxy and pagerank_log straight to the illicit transaction label lined up perfectly with what makes sense in practice and matched the local models. Seeing these directions hold steady across different causal discovery tests makes the logic feel sound, hinting that anonymity tricks and network visibility definitely feed into how financial crime plays out. These connections depend on the usual assumptions like causal sufficiency, faithfulness, and assuming no hidden variables

are messing with the data. The mapped-out networks serve as solid, plausible explanations rather than absolute, undeniable causal facts.

Source	Target
anonymous_proxy	label
avg_neighbor_degree	betweenness_approx
avg_neighbor_degree	community_hopping_proxy
avg_neighbor_degree	label
betweenness_approx	label
betweenness_approx	rapid_redistribution_proxy
degree_log	anonymous_proxy
degree_log	betweenness_approx
degree_log	pagerank_log
in_degree_log	anonymous_proxy

Table. 2: Structural relationship analysis

4.4 Treatment Effects and Heterogeneous Causal Impacts

Testing the actual impact of the behavioral anonymity proxy on illicit transaction outcomes involved setting up two separate approaches: a custom Double Machine Learning setup and the standard LinearDML package from EconML. The custom model pointed to an average treatment effect of -0.0086, with a 95% confidence interval spanning from -0.1885 to 0.3508. Looking at the LinearDML side of things, the math gave an estimated effect of -0.0845, sitting inside a confidence interval from -0.3849 to 0.2159. Both attempts landed on results that are completely insignificant from a statistical standpoint. Because the numbers hover right around zero, it looks like the anonymity proxy does not actually do much heavy lifting on its own when it comes to steering illicit outcomes, especially once the model controls for background noise from network structures and timing shifts. What this tells us is that anonymous behaviors probably do not work in a vacuum as standalone triggers for financial crime. Instead, they seem to blend into the wider structural environment. The lack of any clear, solid average treatment effect serves as a reminder that there is a big difference between a feature that helps predict something and a feature that actually changes the outcome if someone messes with it. A variable can be great at making a classification model more accurate without being a lever that regulators or investigators can pull to change how criminals behave. This situation brings things back to the core idea of this whole project, which is keeping correlation and causation separate.

Even though the overall averages look flat, digging into the sub-populations of transactions showed that the real-world effects are all over the place. Running a causal forest model pulled out an average individual treatment effect of -0.361. The spread was pretty wide, dropping down to -0.836 at the 10th percentile and climbing back up to 0.110 at the 90th percentile. This divergence shows that the anonymity proxy hits different clusters of transactions in very different ways. The data points to in-degree metrics as a major factor in driving this variance, meaning that the way a wallet connects to the rest of the web dictates how much its anonymity traits matter. Certain specific spots in the network graph appear way more reactive to these anonymity behaviors than the rest of the map. Ultimately, checking for these uneven effects is exactly how to spot the weird, hidden relationships that get washed away when looking only at a single blanket average.

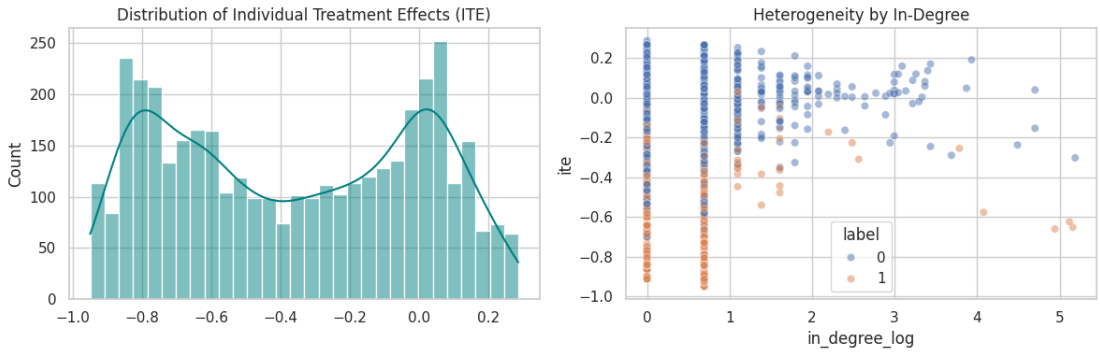

Fig.4: ITE distribution and heterogeneous casual impact analysis

4.5 Predictive Versus Causal Importance and Sensitivity Analysis

A major goal here was to take the predictive patterns generated by standard machine learning and stack them up against the causal pathways pulled from the treatment models. The two angles ended up looking quite different. The SHAP analysis threw its weight behind raw transaction metrics, showing that features like feat_001 and feat_052 basically run the show when it comes to raw predictive power. Flip over to the causal side, and the focus shifts entirely to engineered network details like betweenness_approx, community_size_log, and overall structural connections. The contrast makes it clear that the tools that are best for spotting an anomaly are not always the ones that drive the real mechanics behind the scenes. This split between predictive value and causal truth has real consequences for financial intelligence and policy work. Predictive models are great at flagging suspicious transactions that fit a historical profile, but causal models are what show where a system is actually vulnerable and where an intervention might do some good. Relying purely on predictive charts can easily lead people to the wrong conclusions about why these crimes happen or how a new rule might play out. Using both angles together gives a much more complete picture of how illicit activity functions on the blockchain, creating a better foundation for actual investigations.

The sensitivity analysis also made it obvious that the final causal answers change quickly based on how the anonymity proxy is built and defined. Shifting the anonymity cutoff points between the 0.70 and 0.85 quantiles made the average treatment effect swing wildly from roughly -0.35 up to 0.13. The fact that the numbers flipped from negative to positive as the definition got stricter shows just how fragile these causal estimates can be when tweaking assumptions about what counts as anonymous behavior. This kind of volatility highlights why it is vital to be completely open about how proxies are constructed, to run plenty of stress tests, and to measure uncertainty clearly in observational studies. Rather than ruining the model, doing these checks adds rigor by showing exactly where the boundaries and limits of these blockchain data interpretations sit.

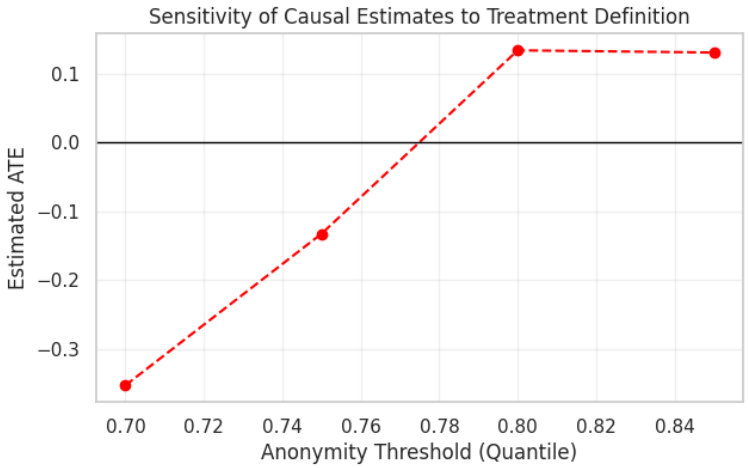

Fig.5: Sensitivity analysis

5. Discussion and Implications

5.1 Financial Crime Interdependencies and the Need for Causal Intelligence

The results of this study show that having high predictive performance is just not enough if the goal is to actually understand how financial crimes happen in blockchain networks. Normal machine learning systems focus heavily on classification accuracy. Despite that focus, financial intelligence work really needs explanations that can back up actual interventions, policy updates, and long-term strategic decisions. Jakir (2025) points out that AI systems working in messy financial settings have to separate real signals from random noise, showing that getting a prediction right does not mean the system understands the actual processes creating those results [12]. The data here backs up that idea. The Random Forest model did great with predictions, but the causal analysis later showed pretty weak average treatment effects for behaviors linked to anonymity. This sort of gap shows that predictive indicators and causal drivers are distinct ideas that shouldn't be treated as the same thing in financial crime analytics.

The title of this paper points directly to how fraud, money laundering, and cybersecurity incidents connect, and the tests show there are big, messy dependencies across these areas. Fraud usually creates dirty money that needs

laundering later, and cybersecurity issues like ransomware, hacked exchanges, or stolen crypto open doors to move huge amounts of criminal cash through blockchains. Rahman et al. showed that machine learning can give early warnings for financial trouble by modeling the complicated links between different big-picture financial indicators, which proves why we need joined-up analytical setups instead of looking at risks completely by themselves [18]. It works the same way in blockchain setups. Cybercrimes, scams, and money laundering setups end up feeding into each other because they use the exact same transaction networks.

The network analysis also suggests that the shape of the blockchain itself acts as a bridge for these relationships. Behaviors like hopping between communities, fast transaction speeds, and how coins get split up create paths to hide, break up, or move stolen funds across decentralized spaces. Fraud, money laundering, and cyber incidents do not happen in separate bubbles. Instead, they are deeply tangled up in big social and technical systems where behaviors change and interact all the time. These patterns mean it makes sense to use combined causal machine learning setups that can spot these structural links and find real intervention points that standard predictive analytics miss.

Of course, these findings have to be looked at alongside some very clear limitations. The analysis uses observational transaction data, so it is impossible to verify causal links through real, controlled experiments. Hidden variables could easily throw off both the treatment assignments and the outcomes, which might bias the estimated effects even with a lot of covariate adjustments. The anonymity metrics used here are just behavioral proxies, not definitive proof that someone used a mixer or privacy tools. On top of that, the study looks only at Bitcoin transactions, meaning it misses what happens on other blockchains or when funds move between different chains. Being honest about these gaps makes the conclusions more trustworthy and shows how hard it is to get clear causal answers in decentralized finance.

5.2 Predictive Explanations, Financial Intelligence Operations, and Regulatory Implications

One of the main points this research brings to light is the difference between predictive explanations and causal explanations in blockchain financial intelligence. The SHAP analysis showed that basic transaction details mattered most for model predictions, but the causal methods told a different story, pointing instead to network structures, community metrics, and how funds were distributed. Miah et al. (2026) mention that AI systems used for high-stakes decisions need a clear line between what the model suggests and the actual reasons justifying that choice, noting that clear explanations have to go deeper than simple statistical links [14]. This study reinforces that point. Variables that help a model predict accurately do not always translate into useful targets for an intervention or represent the actual cause of a behavior. This difference matters a lot for day-to-day financial intelligence operations. A machine learning model built only to predict might be great at spotting a shady transaction, but it won't say much about the root causes of the criminal behavior. Rahman (2025) showed that early warning systems using machine learning are much more valuable when they are part of a bigger decision-making process that looks at context and shifting risk signals [19]. Blockchain investigations face the same reality. Analysts need accurate flags, but they also need to know the underlying mechanics that make illicit tech work. Mixing predictive and causal angles lets investigators point their resources in the right direction, spot systemic weak points, and test out potential fixes with more certainty.

These issues lead straight into regulations and compliance. AI is being used more and more to track money laundering, flag suspicious activity, and manage broader financial systems. Islam et al. built AI decision frameworks that focus on transparency, smooth operations, and helping managers make smart choices in complicated financial settings [10]. Their work shows the need for explainable, accountable analysis, and those principles matter just as much for blockchain rules. The data here suggests that regulators should be careful about treating explainable AI outputs as proof of cause. SHAP values and feature importance metrics just describe how a model behaves, not what happens when a policy changes. Rules that mix up these ideas might end up targeting random correlated factors instead of fixing the actual setups that let financial crimes happen. The causal data also show that anonymity behaviors do not have strong, independent treatment effects once network and timing factors are controlled. This means rules focused only on anonymity indicators might not do much unless they also handle the bigger transaction structures and community behaviors. Because of this, future compliance setups would likely work better if they used causal logic alongside normal predictive tracking, leading to smarter, more targeted strategies.

5.3 Practical Deployment Considerations and Explicit Methodological Limitations

Putting causal machine learning to work in blockchain financial intelligence means dealing with some heavy computational, methodological, and operational hurdles. Bhowmik et al. point out that adaptive machine learning setups have to grapple with the fact that financial and cryptocurrency spaces just do not stay still [5]. They change constantly. Because blockchain networks are so fluid, any analytical system needs to be flexible enough to handle new

transaction patterns, shifting criminal tactics, and evolving network setups. This study uses temporal validation to help mimic real-world deployment, but in an actual operational setting, you would still need constant retraining and monitoring to keep up.

There are a few big methodological limits to talk about openly. For one, this entire approach relies on observational data. That means any causal claims depend completely on standard assumptions like ignorability, consistency, positivity, and the idea that there are no hidden confounding variables. None of this can be proven true using the data at hand, so unobserved factors could easily be skewing both the behavioral indicators and the illicit outcomes. On top of that, the anonymity variable is not a definitive record of someone using a mixer or trying to hide their tracks. Instead, it is a proxy calculated from degree measures, centrality statistics, and transaction speeds. The Elliptic dataset simply lacks verified labels for mixer usage, laundering phases, or the origins of cyberattacks, which makes it impossible to validate these ideas directly.

The scope here is also tied strictly to Bitcoin transactions. Because of that, the findings might not apply to other blockchain networks that use different consensus models, privacy setups, or cross-chain setups. There is also zero intervention data or randomized testing available here, so the analysis is stuck estimating treatment effects under observational assumptions instead of directly measuring how well a specific policy works. Then there is the sheer size of transaction networks, which forced a few computational shortcuts to keep things running. Using approximate betweenness measures, shrinking the list of causal variables, sampling during causal discovery, and simplifying the graph structures means the results might drift a bit from the exact physical properties of the network. These are just the types of compromises that happen when dealing with massive blockchain datasets.

Pointing out these gaps does not take away from what the framework accomplishes. Being open about these boundaries just makes the method more transparent, which is exactly what responsible AI research needs. People reviewing and using these systems know that being honest about data limits, proxy variables, and assumptions makes an empirical study much more credible and easier to reproduce. This work offers plausible causal interpretations backed up by a few different analytical angles, not absolute proof of how crime happens on the blockchain. Any future rollout should bring in human experts, ongoing validation, and extra data feeds to sharpen these causal estimates as blockchain networks keep shifting.

6. **Limitations and Threats to Validity**

6.1 Internal and Construct Validity

A few specific factors limit the internal validity of this work. This setup relies entirely on an observational design using past blockchain data instead of running controlled experiments or live interventions. Because of this, making any claims about cause and effect means assuming everything else stayed perfectly equal and no hidden factors messed with the numbers. A lot of network details, timing data, and behavioral habits got packed into the model, but unmeasured things always slip through the cracks. Off-chain chats, specific crypto exchange rules, police actions, or just general economic shifts do not show up in the dataset. That leaves room for hidden elements to throw off the results. Without clean, verified records of actual interventions, proving a hard cause-and-effect relationship is not really possible, so the calculated treatment effects have to be read as solid guesses rather than proven facts.

The way anonymity gets measured also brings up some questions about what is actually being captured. The anonymity indicator is not direct proof that someone used a mixer or privacy tools. It is just a proxy built from things like network positions, transaction speeds, and moving patterns. The Elliptic dataset does not come with neat labels saying a transaction definitely used a mixing service. So, the proxy just catches shapes in the data that look like anonymous trading, without proving the exact tools used. The crime labels themselves are another weak point. They come straight from the creators of the dataset and miss a lot of illegal activity happening on blockchains. Tax dodging, dodging sanctions, insider trading, market manipulation, and newer DeFi scams do not really fit into this specific classification setup. Tracking hacks and cyber attacks adds another layer of guesswork. Ransomware, hacked exchanges, stolen wallets, and extortion do not have clear tags in the data. Instead, these events get guessed at by looking at weird transaction habits and strange network structures. These setups make sense on paper for finding tech crimes, but they cannot replace actual, verified reports of cyber attacks. The links found between fraud, laundering, and cyber incidents are more about shared patterns and behaviors than a direct line to specific real-world crimes.

6.2 External, Statistical, and Computational Validity

Looking only at Bitcoin and the Elliptic data limits how far these findings can stretch. Bitcoin runs on its own unique transaction setup, consensus rules, and open ledger style that looks completely different from Ethereum,

decentralized finance apps, privacy coins, or systems that cross between chains. Copying these findings over to other blockchain networks requires a lot of caution. Decentralized exchanges and smart contracts create entirely new behaviors that this dataset simply does not capture. Checking if this causal machine learning framework actually holds up everywhere will take future work using multi-chain data.

The statistics side has its own hurdles to clear. The huge gap between the massive amount of clean transactions and the tiny sliver of dirty ones makes predictive modeling and causal guessing pretty tricky. Things like precision-recall metrics, balanced testing, and bootstrap confidence intervals helped handle the gap, but guessing impacts when events are this rare always leaves plenty of uncertainty. Testing different anonymity levels showed that the conclusions change depending on how the rules are set up. The average treatment effects swung from negative to positive numbers just by tightening the thresholds. That variance shows why testing for robustness and being open about assumptions matters so much. On top of that, every single causal estimate relies heavily on the specific model choices, features, and mathematical shapes picked for the Double Machine Learning and causal forest steps.

Heavy computer processing limits the system, too. The sheer size of the Bitcoin network meant using rough estimates instead of exact math for heavy network metrics like betweenness centrality. The causal discovery steps had to use smaller lists of variables and sampled data to keep from crashing the computers. These shortcuts allowed for big tests and let others run the same code in normal cloud environments, but they might skip over quiet, subtle connections hiding in the full network. Running deep graph math or full causal tests across hundreds of variables with exact numbers is just too heavy for practical setups. Instead of hiding these shortcuts, this analysis lays them out as necessary trade-offs between deep math, clear explanations, and managing huge data sizes. Being transparent like this makes the results more believable and matches up with honest work in artificial intelligence and financial studies.

7. Future Research Directions

The findings here point toward a few useful directions for future work in causal machine learning and blockchain financial intelligence. One obvious next step is to look beyond Bitcoin and explore multi-chain ecosystems. Most modern illicit setups do not stick to one network. They use cross-chain bridges, decentralized exchanges, and multiple protocols to move money around and hide their tracks from regulators. Checking out what happens on Ethereum, Binance Smart Chain, or Solana would give a much clearer picture of how cyber-financial crime actually works across different blockchain infrastructures. Getting hold of verified mixer datasets is another big piece of the puzzle. This study had to rely on behavioral proxies because public transaction logs do not come with neat labels marking privacy tools or specific laundering phases. Having access to confirmed mixer data, sanctioned wallet lists, or actual law enforcement intelligence would make the variables a lot more precise. It would let researchers properly evaluate how well different interventions work against anonymity tools and make the estimates of causal effects much more dependable for anti-money laundering tools.

It would also be worth looking into dynamic causal graphs that can track how these networks change over time. Criminal tactics and cyberattacks do not stay still. They change the moment a new technology or a new regulation drops. Because of this, static models can miss how these financial crimes evolve. Using things like dynamic causal discovery or temporal graphical models could help show how cause-and-effect relationships shift. If these methods get built into real-time monitoring setups, it could flag suspicious shifts a lot earlier. Counterfactual blockchain simulations offer another interesting path. Setting up synthetic testing environments would let people play out hypothetical scenarios, like a new regulatory rule, a change in exchange compliance, or a network disruption, without needing to test them out in the real world. This kind of simulation pairs well with observational data by giving a controlled space to test out policies. Combining these counterfactual tests with agent-based modeling or digital twins could show how criminal groups react when their environment changes.

There is also plenty of room to mix causal reasoning with graph neural networks. Right now, most graph learning models are just built to predict the next step, which does not help much when trying to figure out what happens if an intervention is made. Developing causal graph neural networks that handle both structural learning and treatment effects could change that, offering a better explanation of how financial crime spreads across decentralized networks. Along the same lines, building real-time platforms that combine predictive models, causal discovery, and explainable AI into one workspace would be a major practical upgrade. Lastly, threat intelligence setups need to start pulling in data from outside the blockchain itself. Transaction data tells only part of the story. Incorporating social media chatter, dark web forums, cybersecurity incident reports, and institutional risk metrics would give a much deeper understanding of how these modern crimes are organized. Piecing all these different sources together is probably the best way to build intelligence platforms that can spot new threats early, predict how bad actors will adapt, and give

regulators solid evidence to work with. Taking this kind of multi-angled approach seems necessary for keeping up with decentralized economies.

8. Conclusion

This study introduces a causal machine learning framework designed to uncover hidden cause-and-effect relationships among fraud, money laundering, and cybersecurity incidents within blockchain ecosystems. By integrating predictive modeling, explainable artificial intelligence, causal discovery methods, and treatment-effect estimation, the research goes beyond traditional transaction classification to provide a deeper understanding of the structural mechanisms that drive illicit financial activities. The findings reveal that predictive explanations and causal explanations serve fundamentally different purposes. While machine learning models are effective in identifying suspicious transactions, the variables that enhance predictive performance do not necessarily represent actionable intervention targets. Network structures, community behaviors, and transaction dynamics play crucial roles in explaining how illicit activities emerge, spread, and interact within decentralized financial systems. These insights underscore the importance of combining predictive analytics with causal reasoning when designing financial intelligence and regulatory frameworks.

The study also emphasizes the interconnectedness of fraud, money laundering, and cybersecurity incidents. Cyber-enabled attacks generate illicit financial flows, fraudulent activities create incentives for laundering operations, and the structures of blockchain networks facilitate the movement and concealment of digital assets. Understanding these relationships demands analytical approaches that capture structural dependencies rather than focusing solely on isolated transaction characteristics. Several limitations restrict the interpretation of the results. The analysis is based on observational Bitcoin data, behavioral proxy treatments, computational approximations, and the lack of verified intervention or mixer information. There may also be hidden confounding factors that remain unobserved, and the findings might not directly generalize to other blockchain ecosystems or cross-chain environments. A clear acknowledgment of these constraints enhances the transparency and reproducibility of the research and provides guidance for future investigations. This work contributes to the emerging field of causally informed financial intelligence by demonstrating that understanding why illicit activities occur is just as important as identifying where they occur. The proposed framework offers a practical foundation for future research and supports the development of more transparent, accountable, and intervention-oriented systems for combating financial crime in increasingly decentralized digital economies.

References

1. Aashish, K. C., Zamil, M. Z. H., Mridul, M. S. I., Akter, L., Sharmin, F., Ayon, E. H., ... & Malla, S. (2025). Towards eco-friendly cybersecurity: Machine learning-based anomaly detection with carbon and energy metrics. International Journal of Applied Mathematics, 38(9s).
2. Akcora, C. G., Li, Y., Gel, Y. R., & Kantarcioglu, M. (2020). BitcoinHeist: Topological data analysis for ransomware prediction on the Bitcoin blockchain. In Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (pp. 4439-4445).
3. Alam, M., Shil, S. K., Sharmin, F., KC, A., Md, A. H., Ali, M., ... & Malla, S. (2026). Hybrid deep learning models for equipment failure prediction in US industrial systems. International Journal of Applied Mathematics, 39(1s).
4. Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82-115.
5. Bhowmik, P. K., Subha, D. T., Rahim, A., Mohammed, A. A., Begum, M., Chowdhury, R., ... & Shati, M. A. Self-adaptive machine learning models for financial risk forecasting: Handling non-stationarity in banking and cryptocurrency time series.
6. Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. The Econometrics Journal, 21(1), C1-C68.
7. Dola, A., Begum, S., Antara, U. K., Islam, M. R., Sultana, T., & Zabin, N. (2024). Machine learning models for detecting hidden collusion networks in US corporate finance. Journal of Economics, Finance and Accounting Studies, 6(1), 143-154.
8. Foley, S., Karlsen, J. R., & Putniņš, T. J. (2019). Sex, drugs, and bitcoin: How much illegal activity is financed through cryptocurrencies? The Review of Financial Studies, 32(5), 1798-1853.
9. Hu, Y., Seneviratne, S., Thilakarathna, K., Fukuda, K., & Seneviratne, A. (2021). Characterizing and detecting money laundering activities on the Bitcoin network. Future Generation Computer Systems, 114, 328-342.
10. Islam, M. R., Subha, D. T., Pramanik, M. T., Akter, M., Sweet, M. M. R., Robbani, M. S., ... & Zeeshan, M. A. F. AI-driven decision support systems for optimizing working capital and customer experience in the US: A transaction-based simulation framework for SMEs.

11. Islam, M. Z., Sumsuzoha, M., Islam, M. R., Kawsar, M., Mithu, M. F. H., Pant, S., ... & Al Helal, M. A. Graph neural networks for systemic financial risk forecasting: Modeling cross-market contagion between banking systems and cryptocurrency markets.
12. Jakir, T. (2025). Signal-to-noise analysis of crisis indicators in global finance using artificial intelligence. International Journal of Applied Mathematics, 38(10s), 1815-1836.
13. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems (Vol. 30).
14. Miah, M. N. I., Uddin, M. J., & Kakumani, M. (2026). Artificial intelligence in sentencing: Evaluating machine learning models for sentencing recommendations in the US. Frontiers in Computer Science and Artificial Intelligence, 5(4), 30-43.
15. Molnar, C. (2022). Interpretable machine learning (2nd ed.). Lulu.com.
16. Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system.
17. Pearl, J. (2009). Causality: Models, reasoning, and inference (2nd ed.). Cambridge University Press.
18. Rahman, M. K., Hossain, M. S., Haque, S. U., Jahed, K. A., Robbani, M. S., Shati, M. A., ... & Pramanik, M. T. Machine learning models for early warning of financial crises in the US economy using macro-financial indicators.
19. Rahman, M. S. (2025). Machine learning-enabled early warning system for detecting micro-inflation clusters in the US economy. International Journal of Applied Mathematics, 38(12s), 2743-2769.
20. Reza, S. A., Rahman, M. K., Rahman, M. D., Sharmin, S., Mithu, M. F. H., Hasnain, K. N., ... & Kabir, R. (2025). Machine learning-enabled early warning system for financial distress using real-time digital signals. arXiv preprint arXiv:2510.22287.
21. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135-1144).
22. Shawon, R. E. R., et al. (2025). Detecting illicit cross-chain fund movement: Behavioral machine learning models for bridge-based laundering patterns. International Journal of Applied Mathematics, 38(12s).
23. Spirtes, P., Glymour, C. N., & Scheines, R. (2000). Causation, prediction, and search (2nd ed.). MIT Press.
24. Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. Journal of the American Statistical Association, 113(523), 1228-1242.
25. Weber, M., Domeniconi, G., Chen, J., Weidele, D. K. I., Bellei, C., Robinson, T., & Leiserson, C. E. (2019). Anti-money laundering in Bitcoin: Experimenting with graph convolutional networks for financial forensics. In Proceedings of the KDD Workshop on Anomaly Detection in Finance.
26. Zheng, X., Aragam, B., Ravikumar, P., & Xing, E. P. (2018). DAGs with NO TEARS: Continuous optimization for structure learning. In Advances in Neural Information Processing Systems (Vol. 31).
27. Zohar, A. (2015). Bitcoin: Under the hood. Communications of the ACM, 58(9), 104-113.