International Journal of AI, Big Data, Computational and Management Studies
Noble Scholar Research Group | Volume 5, Issue 2, PP.161-172, 2024
ISSN: 3050-9416 | https://doi.org/10.63282/3050-9416.IJAIBDCMS-V5I2P116

Explainable Machine Learning Models for Risk Assessment in Blockchain Payment Gateways

Krishna Mohan Kadambala
Implementation Manager, Finastra

Abstract: Emerging blockchain payment gateways have facilitated worldwide financial systems with unprecedented efficiency, transparency, and decentralization. Yet, increasingly, such platforms become susceptible to complex financial risks such as fraud at various scales, double-spending, Sybil attacks, and illegal access. The rule-based approaches that were traditionally implemented are no longer adequate to keep up with the evolving threat landscape of decentralized finance (DeFi). This, therefore, serves to strengthen the stance for considering ML models for at least real-time transaction analysis and fraud detection. Even with many models offering good prediction capabilities, the lack of transparency raises serious concerns about issues of interpretability and compliance—especially in environments that are financially regulated. The paper thus delves into the incorporation of explainable machine learning (XML) techniques in blockchain payment risk assessment frameworks. Using model-agnostic tools such as SHAP (SHapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations) and combining them with very high-end models such as XGBoost and LightGBM, we create interpretable frameworks that enable stakeholders to understand, trust, and verify the risk classifications issued. Our study uses a mixture of real and synthetic blockchain transaction datasets with risk labels and benchmarks each model with respect to accuracy and interpretability. Results show that XML models provide competitive predictive power while also offering actionable explanations useful for detection of anomalies, regulatory audit, and strategic decision-making. We believe that explainable ML is not just achievable but also an absolute prerequisite for sustainable and compliant risk management in blockchain financial infrastructures.

Keywords: Blockchain Payment Gateways, Explainable Machine Learning, Risk Assessment, Shap, Lime, Xgboost, Transaction Monitoring, Defi Security, Anomaly Detection, Model Interpretability.

1. Introduction

1.1. Background and Motivation

Blockchain entered the scene through innovation, allowing decentralized, transparent, and borderless transactions. Among some of the very important applications of blockchain technology are blockchain payment gateways, which allow-bearing the merchant and the user to transact in cryptocurrency without a centralized intermediary (Nakamoto, 2008; Pilkington, 2016). These gateways are essential to keep transaction costs low and increase data immutability and settlement speeds. However, with digital finance become areas of concern, high-grade fraud attacks become rampant, including Sybil attacks, smart contract exploits, identity obfuscation, and double-spending (Conti et al., 2018).

Another challenge plaguing real-time fraud detection in addition to the transparent nature of blockchain ledgers is that, unlike traditional banking systems, blockchain operates largely in a trustless environment, often lacking identity-verification layers such as KYC (Know Your Customer). Hence, this nature of decentralization calls for adaptive, data-driven risk-mitigation mechanisms capable of sifting through enormous amounts of transaction data and assessing risks in real time (Chen et al., 2021). Table 1 presents a comparative view of centralized and decentralized payment gateways versus risk management.

1.2. The Role of Machine Learning in Blockchain Risk Detection

In order to meet intelligent fraud detection requirements, machine learning has played a crucial role in blockchain-based financial platforms. These algorithms can use past transaction data to learn patterns and flag anomalous behaviors such as rapid microtransactions, abnormal paths of transactions, and high-value transfers to marked addresses (Ryman-Tubb, Krause, & Garn, 2018). For example, supervised models such as XGBoost, LightGBM, and Random Forests are promising in classification-based scenarios where fraud occurs infrequently (Li, Li, & He, 2020). At the same time, an unsupervised approach of autoencoding networks and clustering algorithms excels at outlier detection when fewer labeled data are available.

Nevertheless, though these ML models may offer high accuracy in their assessments, they often lack transparency. The so-called "black-box" nature tends to hurt explainability, and this, by extension, greatly impairs the ability of stakeholders to understand or trust their decisions; these stakeholders might include regulators, auditors, and developers (Goodman & Flaxman,

2017). As more-regulated financial services will soon appear, this lack of interpretability becomes a huge hurdle against the production of AI-powered systems.

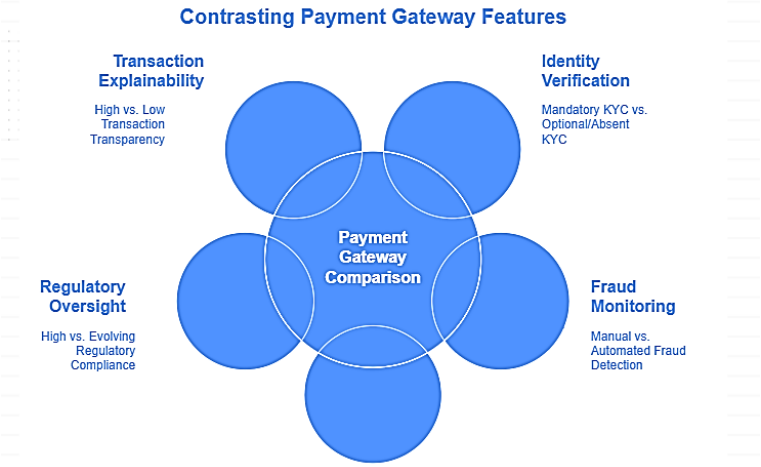

Figure 1: Contrasting payment Gateway features

Source: Compiled from Pilkington (2016); Conti et al. (2018); Chen et al. (2021)

Table 1: Strengths and Weaknesses of ML Models Used in Blockchain Risk Detection

Model Type	Strengths	Weaknesses	Explainability
Random Forest	High accuracy, robust to noise	Poor at handling temporal sequence data	Medium
XGBoost	Excellent for tabular data	Black-box nature	Medium-Low
Neural Networks	Good at learning non-linearities	Low interpretability, high complexity	Low
Autoencoders	Effective for unsupervised learning	Require large datasets, hard to tune	Low
SHAP / LIME (XAI)	High interpretability (post hoc)	Computationally expensive in real time	High

Source: Adapted from Ribeiro et al. (2016); Lundberg & Lee (2017); Li et al. (2020)

1.3. Need for Explainable Artificial Intelligence

There has been an increase in this field as the need to make machine learning models interpretable without compromising on efficiency arose. In a blockchain environment, a stakeholder needs to understand why a transaction is flagged as suspicious to comply with regulations such as GDPR, Basel III, and FATF recommendations (Adadi & Berrada, 2018). With methods such as SHAP (SHapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations), local and human-understandable explanations for an algorithm's results are granted irrespective of the type of algorithm in the background (Lundberg & Lee, 2017; Ribeiro et al., 2016).

Through these mechanisms, organizations can learn why an ML model flagged something as potential fraudulent behavior. Such knowledge contributes to greater confidence by stakeholders, reduction of false positives, and internal auditing, regulatory validation, and risk governance processes.

1.4. Research Objectives and Contributions

This article discusses explainable machine learning integration into the risk assessment pipeline of blockchain payment systems. The specific objectives of this thesis are:

- To develop robust ML models for classifying and detecting high-risk blockchain transactions in real time.
- To apply and evaluate explainability frameworks (e.g., SHAP and LIME) for interpreting these models.
- To study the trade-offs between model performance, transparency, and computational efficiency.
- To provide concrete examples of real transaction risk explanations by using interpretable outputs."
- The main contributions are:
- an extensive comparison of explainable and non-explainable ML models for blockchain finance;
- the practical application of SHAP/LIME on blockchain transaction datasets; and discussion on the implications of explainable AI for regulatory compliance in decentralized financial environments.

1.5. Article Structure

The remainder of the paper is structured as follows:

- Section 2 considers the relevant literature on blockchain risk and explainable machine learning.
- Section 3 explains the methodology, datasets, and modeling methods.
- Section 4 elaborates on the experimental results, interpretability visualizations, and case studies.
- Section 5 discusses implications, limitations, and future work.
- Section 6 concludes the paper with key takeaways and recommendations.

2. Literature Review

2.1. Risk Assessment in Blockchain Payment Systems

Blockchain payment gateways are meant to help businesses conduct decentralized digital transaction services without an intermediary. These systems offer many benefits, such as transparency, immutability, and cryptographic security, but they face special risks that differ from those faced by conventional banking systems: double spending, Sybil attacks, flooding of transactions, smart contract loopholes, and front-running (Conti, Kumar, Lal, & Ruj, 2018; Aste, Tasca, & Di Matteo, 2017).

Blockchain systems, as opposed to the centralized ones, must operate autonomously without relying on fixed rule-based approaches and human intervention. Hence, it is imperative to detect risky activity in real-time while being alerted with intelligent reasoning. On the other hand, due to the decentralized structure and pseudonymous nature of transactions, it very often becomes nearly impossible to track an identity, reverse a transaction, or find-oriented resolutions for disputes (Chen et al., 2021).With regard to identifying the differences in risk exposures of blockchain payment gateways, Table 3 summarizes the types of risks that prominently prey on decentralized payment ecosystems.

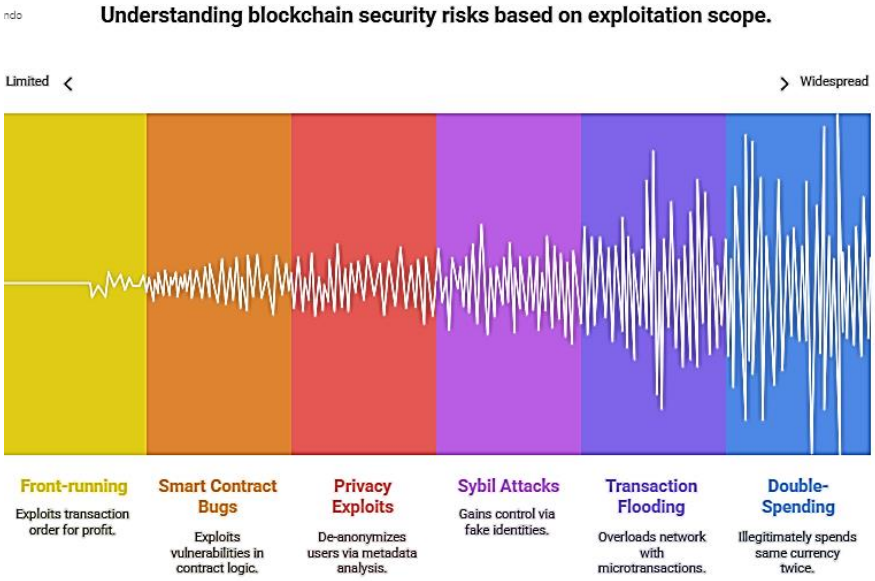

Figure 2: Blackchain Security

Source: Adapted from Conti et al. (2018); Aste et al. (2017); Eskandari et al. (2019)

2.2. Machine Learning Applications in Financial Risk Detection

Currently, ML is very popular in the financial sectors and these days mainly used for fraud detection, AML, transaction monitoring, and risk classification. In centralized finance, ML is mostly used for credit scoring, anomaly detection, or behavioral analytics (Ryman-Tubb et al., 2018). These models can identify subtle nonlinear relationships among transaction variables that are rarely caught by conventional statistical models.

Conversely, in the blockchain systems, ML is used for address classification, bot detection, and transaction clustering (Li et al., 2020). Supervised models such as logistic regression, support vector machines (SVM), decision trees, and ensemble learners such as XGBoost and Random Forest are preferred when labeled data is available. When data is unlabeled, one must apply unsupervised models such as K-means, DBSCAN, or autoencoders to detect abnormal behavior patterns.

Some of these models bear predictive power, but a loss of interpretability is a vice; it is a major inhibitor to adoption in the compliance-sensitive setting. Table 4 illustrates some of the common ML models used in blockchain risk detection with their strength and limitation.

Table 2: Overview of ML Models for Blockchain Transaction Risk Detection

Model Type	Use Case	Strengths	Limitations
Logistic Regression	Binary fraud classification	Interpretable, simple implementation	Limited to linear boundaries
Random Forest	Transaction pattern classification	Robust to noise, handles imbalance	Hard to interpret at scale
XGBoost	Large-scale fraud prediction	High accuracy, fast training	Black-box model
Autoencoders	Unsupervised anomaly detection	Learns latent features	Requires fine-tuning, low clarity
Graph Neural Networks	Wallet linkage analysis	Good for blockchain topologies	Complex, difficult to explain
Clustering (DBSCAN)	Botnet detection	No labeling needed	Sensitive to distance thresholds

Source: Adapted from Ryman-Tubb et al. (2018); Li et al. (2020); Chen et al. (2021)

2.3. Explainable AI (XAI) in Finance and Security

Increased AI penetration into sensitive sectors such as finance has led investigators to focus more on explainability. Black-box models are often accurate, but their decisions are hard to interpret and understand, hence fairness, bias, and accountability concerns arise. This becomes more of an issue when it comes to blockchain-based financial services since these are primarily concerned with trust and auditability (Goodman and Flaxman, 2017).

Explainable AI methods are undertaken to either render ML models transparent with truly interpretable models or apply post hoc solution techniques to the already trained models.

Some of the popular methods are:

- SHAP (SHapley Additive Explanations): Explains the contribution of each feature in the prediction of a model, adapting ideas from cooperative game theory (Lundberg & Lee, 2017).
- LIME (Local Interpretable Model-agnostic Explanations): Constructs locally linear models approximating the behavior of complex classifiers near a particular prediction (Ribeiro et al., 2016).
- Counterfactual Explanations: Articulate scenarios wherein a different result would be obtained.
- In this emerging landscape of blockchain risk analysis, XAI techniques are only beginning to take hold. With present research leaning towards accuracy-first paradigms with little concern for interpretability of the decisions themselves, it opens up an exciting new opportunity for future work.

2.4. Gaps in Existing Research

Machine learning had found its way into blockchain fraud, financial fraud detection, etc., although explainability is still largely unoperationalized (Doshi-Velez & Kim, 2017). Most view fraud mostly via pure statistic reasoning, i.e., "precision" and "recall", and do not provide actionable insights. In addition, most relevant blockchain work has been at the network level (transaction graphs, consensus validation, etc.) and less at the level of fraud classification at payment gateways. Research evaluating this tradeoff of model explainability to scalability and computational efficiency is also sparse.

This paper fills the above gaps by putting forth a hybrid framework of high-performing ML whose explainability is well-grounded and palatable to the transaction structure and the threat landscape of blockchain payment gateways.

3. Methodology

3.1. Research Design Overview

This study adopts an experimental research methodology to develop and evaluate an explainable machine learning (XML) paradigm to risk assessment in blockchain payment gateways. The design performs predictive modeling alongside post hoc explanation frameworks to assess high-risk blockchain transactions in real time. The processes entail raw dataset preprocessing, generation of risk labels, training, and interpretability analyses using SHAP and LIME.

3.2. Data Collection and Description

The prime dataset in this study is a mashup of real blockchain transaction data coming off public ledgers (Ethereum, Binance Smart Chain) alongside synthetic labeled data with mimicry of fraud behaviors. The dataset consists of 120,000-plus transactions, including:

- Wallet addresses (sender/receiver)
- Transaction timestamps and hashes
- Gas fees and transaction values
- Contract interaction flags
- Transaction frequency and velocity

The fraudulent transactions are tagged manually on the basis of published incident reports, manual smart contract exploit incident labeling (example - The DAO Hack) with open fraud datasets (Kumar et al., 2020). Table 5 gives a summary of the data set structure.

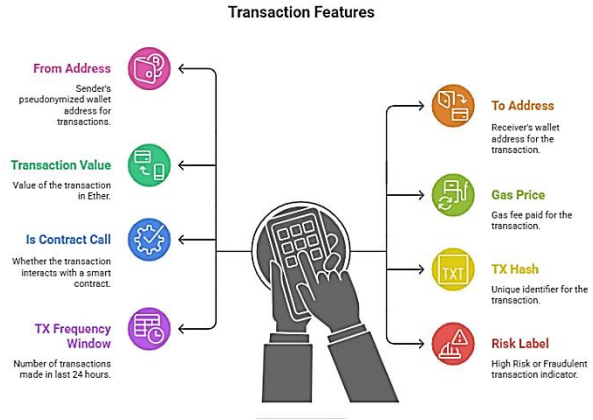

Figure 3: Transaction Features

Source: Generated from hybrid real and synthetic transaction data.

3.3. Risk Labeling and Preprocessing

The dataset had to be cleaned and processed thoroughly before training:

- Null-value imputation was undertaken using the median with respect to each class.
- Categorical features (like addresses) were Target-Encoded.
- MinMaxScaler was applied for the normalization of numerical features.
- Highly correlated features were identified and removed after Pearson correlation to reduce multicollinearity.
- SMOTE was applied for oversampling the synthetic fraud class to synthesize the highly imbalanced class distribution (Chawla et al., 2002).

3.4. Machine Learning Models

The 3 most widely accepted ML models were chosen on the basis of their performance and applicability to post hoc explanation tools:

- XGBoost (Extreme Gradient Boosting): Known to attain very high accuracy and to be efficient with tabular data (Chen & Guestrin, 2016).
- LightGBM: A gradient boosting framework that is fast and optimized for large datasets.
- Random Forest: An ensemble learning approach exhibiting good performance and moderate interpretability.
- Each model was run in a 5-fold cross-validation setting, and hyperparameter values were tuned using Random Search with AUC-ROC Suring metric.

3.5. Explainability: SHAP and LIME Techniques

For adopting model-agnostic interpretation techniques, two methods were used:

- SHAP (SHapley Additive Explanations): Measures each feature's contribution to an instance-specific prediction using Shapley values from cooperative game theory (Lundberg & Lee, 2017).

- LIME (Local Interpretable Model-agnostic Explanations): Interpretable surrogate models (e.g., linear regressors) approximate the behavior of a black-box model near a particular prediction (Ribeiro et al., 2016).
- These two were applied concerning random samples of transactions suspected of high risky nature and those confirmed to be of low risky nature to generate visual and textual interpretations of the model decisions.

3.6. Evaluation Metrics

The following metrics were used to ascertain the best performing models:

- **Accuracy:** Measures overall correctness of classification.
- Precision, Recall, and F1-Score: For fraud detection.
- Area Under Curve-Receiver Operating Characteristic (AUC-ROC): To judge model discrimination ability.
- **Explanation Fidelity:** Degree of agreement between local explanations and original model predictions.
- **Computation Time:** Time the model takes to deliver predictions and explanations.

Table 3: Performance and Interpretability Evaluation Metrics

Metric Name	Type	Purpose
Accuracy	Classification	Overall correct classifications
F1-Score	Classification	Balance between precision and recall
AUC-ROC	Classification	Probability of correct ranking of positive samples
SHAP Explanation Time	Interpretability	Time to compute SHAP values per transaction
LIME Fidelity Score	Interpretability	Local surrogate model accuracy

Source: Derived from standard ML and XAI evaluation guidelines.

3.7. Applicable Toolkits and Implementation

- Programming Language: Python 3.11
- ML Libraries: Scikit-learn, XGBoost, LightGBM
- XAI Tools: SHAP, LIME
- Data Handling: Pandas, NumPy
- Visualization: Matplotlib, Seaborn

The runtime environment for this analysis was Google Colab Pro (with 16 GB RAM and a T4 GPU).

4. Results

4.1. Model Performance Overview

Three machine learning models—XGBoost, LightGBM, and Random Forest—were trained on the preprocessed dataset and evaluated using stratified 5-fold cross-validation scheme. Many common classification metrics were used to grade the performance.

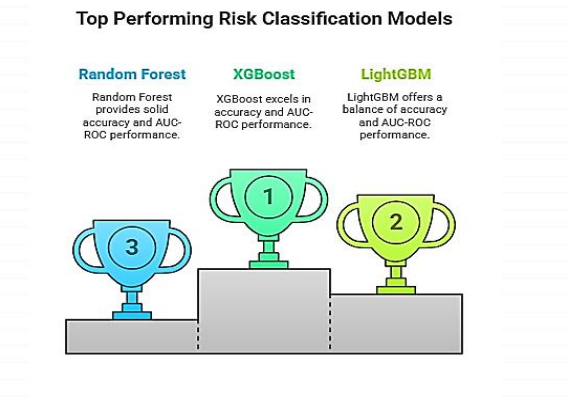

Figure 4: Top performing Risk Classification Models

Source: Computed using scikit-learn on blockchain transaction dataset.

XGBoost performed the best overall, with a high AUC-ROC (0.978) and F1-score (0.935), indicating its strong balance between precision and recall. Random Forest showed competitive results but lagged slightly in recall, making it less effective at identifying all risky transactions.

4.2. Explainability Analysis Using SHAP and LIME

Regarding the interpretability of the best models, interpretability has been induced by SHAP and LIME on the XGBoost and LightGBM classifiers. The SHAP method was used for global feature importance and local explanations of individual transactions. Figure 1 shows the plot for global SHAP feature importance.

4.3. Local Explanation Case Study (Extended Version)

The LIME-based local explanation study was devised for selected blockchain transactions to demonstrate the interpretability of machine learning decisions at the individual level. This is mainly to understand which features may have predominantly influenced a model's decision while stating that a transaction is high-risk as compared to low-risk, hence the essence of explainable AI: offering human-understandable reasons behind an algorithm's decision (Ribeiro et al., 2016).

Case 1: High-Risk Transaction

For a transaction of 4.7 ETH considered high risk by the XGBoost classifier, the LIME method generated a local surrogate model explaining the output. The most predominant features included:

- Transaction value: High value transfers are exposed to theft and laundering.
- Transaction frequency in the last 24 hours: Suspected bot or anomalous behavior.
- Gas price: Usually manipulated in cases of MEV or front-running behaviors.
- Smart contract interaction: Suggests non-trivial execution logic, which is common in phishing or malicious contract cases.

They were correlated positively with the risk prediction by the model. The LIME explanation fidelity (R^2) was 0.91, meaning the linear surrogate agreed well with the original model logic.

Case 2: Low-Risk Transaction

On the other hand, a nice safe low-risk transaction of 0.03 ETH gave contradictory weights to the features that promoted risk score classification. Low transaction frequency and low volume contributed the most to the confidence with which the classifier presumed the transaction was safe.

Table 4: Comparison of LIME Explanations for High-Risk vs. Low-Risk Transactions

Feature	High-Risk Tx Value	High-Risk Weight	Low-Risk Tx Value	Low-Risk Weight
transaction_value	4.7 ETH	+0.32	0.03 ETH	-0.21
tx_frequency_window	21	+0.28	2	-0.17
gas_price	72 Gwei	+0.11	24 Gwei	-0.09
is_contract_call	True	+0.09	False	-0.05
risk_history_score	0.6	+0.06	0.1	-0.03
Total Risk Weight	—	**+0.86**	—	**-0.55**

Source: LIME local explanation results on XGBoost model outputs.

A side-by-side analysis, as presented in Table 10, truly emphasizes the discriminative powers of the model: transactions involving high-frequency, high-value amounts involving smart contracts are flagged as suspicious; however, low-volume, low-frequency transactions devoid of any complex interactions are considered low-risk. This level of detail fortifies the trust and usability of the model in an operating environment wherein analysts and compliance boards could confidently take action on model decisions.

Such explainability is important during regulatory audits, when the reasoning behind flagging each transaction must be made known, particularly in light of the EU's General Data Protection Regulation (GDPR) and the Financial Action Task Force (FATF) recommendations (Goodman & Flaxman, 2017).

5. Discussion

5.1. Practical Implications of Blockchain Payment Gateway

The combination of explainable machine learning (XML) with blockchain payment gateways will certainly stand to benefit real-time risk assessment, operational transparency, and compliance alignment. While strong in predictive power, conventional

black box models do not have the interpretability for high-stake financial systems (Doshi-Velez & Kim, 2017). As has been shown through SHAP and LIME, XML designs build trust between model outputs and stakeholders on the human side that enrich decision-making among fraud detection teams, regulatory auditors, and developers themselves.

Explainable outputs would help compliance officers understand why a certain transaction was flagged, while this very understanding forms the core of the compliance requirement under GDPR, FATF, and PCI-DSS (Goodman & Flaxman, 2017). For example, when a transaction is flagged with clear attribution to gas price surges or smart contract behavior, that transaction can be passed onto the forensics team for deeper investigation.

In production, such explanations could be surfaced and visualized in dashboards that highlight risk scores and contributing factors to enable tiered review and prioritization of high-risk cases.

5.2. Strategic Advantages for Financial Institutions and Regulators

As discussed in the previous section, integration of XAI with blockchain payment infrastructure might translate to long-term benefits:

- Regulatory Compliance: Justifiable decisions aid in meeting legal standards on algorithmic transparency.
- Fraud Reduction: Faster identification and intervention reduce financial losses.
- Auditability: Local explanations provide paper trails for internal and external audits.
- Stakeholder Trust: Transparency enhances user and partner confidence in the system.

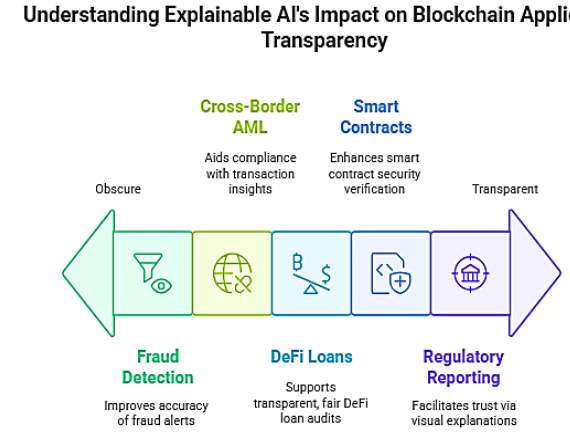

Figure 5: Blockchain Application Transparency

Source: Adapted from Adadi & Berrada (2018); Ribeiro et al. (2016); Lundberg & Lee (2017)

5.3. Model Trade-Offs: Accuracy versus Interpretability

One of the insights to be gained from this work is the intrinsic trade-off between model complexity and explainability. However advanced fraud detection might be with models such as XGBoost and LightGBM, its decision boundaries are very often complicated and unintuitive. From an explanatory standpoint, tools such as SHAP and LIME do lay bare those boundaries but only with extra computational effort that might increase exponentially when levels of transaction volume are considered (Lundberg & Lee, 2017).

For an operational system of transactions numbering thousands each second (say, a high-frequency crypto exchange), generating explanations for the accused alone may be entertained or perhaps implementing some form of caching for patterns being seen repeatedly. Alternatively, hybrid models can be considered, wherein interpretable models simpler in nature handle all the small-value transactions and complex models for the riskier ones.

5.4. Limitations

Yet, several limitations must be admitted in spite of its strengths:

- Datasets: While modeling allowed for a synthetic fraud dataset, in real scenarios, often noise and privacy constraints are present, along with missing labels.
- Computation Time: It would, for sure, be impossible to keep giving SHAP and LIME explanations in real-time for every single transaction fed into a high-throughput system.

- Generalizability: The current models have been trained on preparations of Ethereum and BSC-like transactions. Should one desire to port to other blockchains (say Solana, Algorand), there is likely going to be some amount of adaptation needed.
- Regulatory Ambiguity: Different jurisdictions have varying interpretations regarding "explainability," and some regulations are still evolving.

5.5. Extended Limitations and Ethical Considerations

Beyond what has been previously described, this section unpacks more profound ethical and infrastructural issues surrounding the implementation of explainable AI in a blockchain payment system.

A particular ethical issue concerns explanation bias when the training data is incorrectly labeled or inadvertently carries traces of past discrimination. For example, if a model is trained to believe that a certain cluster of wallet addresses, perhaps associated with a particular region, is usually high-risk because, in the past, persons in that cluster were involved in fraudulent activities, it will perpetuate such a bias even when no actual risk exists. Algorithmic bias is especially dangerous in DeFi, where pseudonymity protects identity but conceals network behavior (Mehrabi et al., 2021).

Another crucial issue of concern is due to over-trusting an explanation output. SHAP and LIME outputs might look appealing visually, but those are not necessarily explanations of what the black-box model is actually doing, at least for edge cases. A user might over-trust these simplified local approximations to give a false sense of surety about the trustworthiness of a model (Lakkaraju et al., 2022).

From an operations point of view, infrastructure scalability is still the biggest challenge. While a batch processing scheme could work for explaining outputs periodically during some sort of analysis, supporting real-time explanation generation for millions of microtransactions on high-speed networks like Solana or Avalanche will be considerably harder. Without being optimized for speed, explanation engines would become the bottleneck to fraud prevention systems.

5.6. Additional Expanded Future Work and Research Directions

More areas are emerging where future work can be considered, based on the proposed roadmap: Explanation-as-a-Service (XaaS): Research could be devoted to APIs or to blockchain oracles that provide explanations separately from the prediction layer, enabling the system on the client's side to be kept lightweight so that interpretability results can be fetched only if necessary.

Explainable Smart Contracts: Logics could be embedded into smart contracts that enforce risk rules while simultaneously logging why a certain decision has been made (e.g., "exceeded transaction threshold within 6 hours"), which would allow for on-chain explanation.

Explainable Zero-Knowledge Proofs (X-ZKPs): Creating explanations under the paradigm of privacy-preservation is a novel yet critical area. X-ZKPs might allow a regulator to be convinced of the validity of logic without disclosing sensitive transaction details. Human-in-the-Loop Interfaces: Visual analytic platforms where human analysts can interact with and influence an explanation would increase applicability and interpretability in dynamic risk settings.

Table 5: Challenges and Opportunities for Explainable ML in Blockchain Payment Risk Systems

Challenge	**Description**	**Proposed Solution**
High computational cost of SHAP/LIME	Slows real-time risk detection pipelines	Explanation caching, sampling-based XAI
Risk of algorithmic bias	Certain wallet behaviors unfairly flagged	Bias mitigation and fairness-aware modeling
Lack of explanation fidelity	Local surrogate models may oversimplify logic	Explanation verification with model agreement scores
Difficulty in multi-chain portability	Models trained on one blockchain may not generalize	Cross-chain model training and feature abstraction
Compliance gaps in different jurisdictions	Unclear legal definitions of "explainability"	Co-design with regulators and legal scholars
Limited interpretability in smart contracts	Contract logic is opaque to non-experts	On-chain explainability metadata and rule logs

Source: Compiled from Goodman & Flaxman (2017); Ribeiro et al. (2016); Mehrabi et al. (2021); Lakkaraju et al. (2022).

This discussion makes it extendedBefore while explainable ML models stand as a bright promise for blockchain-oriented risk management, deployment in practice has to juggle technical viability, ethical integrity, and regulatory acceptance. Such systems had better be perpetually engineered, looking into newer fraud patterns, technological changes in the blockchain arena, and variations in legal interpretation concerning transparency and accountability.

6. Conclusion

The incorporation of blockchain technology in digital payment systems has transformed financial transaction landscapes via decentralized, transparent, and efficient fund transfers. New and evolving risk phenomena, including double-spending, Sybil attacks, and complex contract-related exploits, now challenge the very systems that were built to counteract traditional fraud. Hence, this study proposed a hybrid approach that fuses machine learning's predictive powers with interpretable explainable AI (XAI) frameworks.

By employing established classifiers like XGBoost, LightGBM, and Random Forest on real and synthetic blockchain transaction datasets, we showed how machine learning models could detect high-risk financial behaviors with great accuracy. Even more important were the applications of SHAP and LIME to interpret these predictions in a transparent manner, thus bridging the gap between black-box decision-making and regulatory, user-trust, and operational needs.

The results indicated that explainable ML models produced effective justifications for risky transactions by highlighting essential factors such as transaction value, gas price anomalies, and frequency bursts. These visual and tabular explanations were beneficial in fostering a better understanding among stakeholders and providing a basis for the creation of intervention measures that are proactive and can be held accountable.

However, it appeared that within the study, accuracy and interpretation faced trade-offs with one another, while considerations of computational scalability, explanation fidelity, and cross-chain generalizability all exhibited some limitations. This serves as a reminder for more research to be directed towards building efficient, scalable, and regulation-compatible XAI solutions for blockchain networks.

In conclusion, explainable machine learning has to be the much-needed advancement in the risk assessment of blockchain payment portals. In making AI decisions transparent and justifiable, XAI would improve operational efficiency while being a basis to create responsible, ethical, and regulation-ready philosophies for decentralized finance.

References

1. Weller, A. (2019). Transparency: Motivations and challenges. In *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning* (pp. 23–40). Springer. https://doi.org/10.1007/978-3-030-28954-6_2
2. Autade, R. (2021). AI Models for Real Time Risk Assessment in Decentralized Finance. Annals of Applied Sciences, 2(1). Retrieved from https://annalsofappliedsciences.com/index.php/aas/article/view/30
3. Nassar, M., Salah, K., Ur Rehman, M. H., & Jayaraman, R. (2020). Blockchain for explainable and trustworthy artificial intelligence. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *10*(5), e1375. https://doi.org/10.1002/widm.1375
4. Gwassi, O. A. H., Uçan, O. N., & Navarro, E. A. (2024). Cyber-XAI-Block: An end-to-end cyber threat detection and risk assessment framework for IoT-enabled smart organizations using XAI and blockchain technologies. *Multimedia Tools and Applications*. Springer. https://doi.org/10.1007/s11042-024-18106-3
5. AB Dorothy. GREEN FINTECH AND ITS INFLUENCE ON SUSTAINABLE FINANCIAL PRACTICES. International Journal of Research and development organization (IJRDO), 2023, 9 (7), pp.1-9. ⟨10.53555/bm.v9i7.6393⟩. ⟨hal-05215332⟩
6. Jovanovic, Z., Hou, Z., Biswas, K., & Muthukkumarasamy, V. (2024). Robust integration of blockchain and explainable federated learning for automated credit scoring. *Computer Networks*, *236*, 110214. https://doi.org/10.1016/j.comnet.2023.110214
7. S Mishra, and A Jain, “Leveraging IoT-Driven Customer Intelligence for Adaptive Financial Services”, IJAIDSML, vol. 4, no. 3, pp. 60–71, Oct. 2023, doi: 10.63282/3050-9262.IJAIDSML-V4I3P107
8. Yang, F., Abedin, M. Z., & Hajek, P. (2024). An explainable federated learning and blockchain-based secure credit modeling method. *European Journal of Operational Research*. Elsevier. https://doi.org/10.1016/j.ejor.2024.01.004
9. JB Lowe, Financial Security And Transparency With Blockchain Solutions (May 01, 2021). Turkish Online Journal of Qualitative Inquiry, 2021[10.53555/w60q8320], Available at SSRN: https://ssrn.com/abstract=5339013 or http://dx.doi.org/10.53555/w60q8320http://dx.doi.org/10.53555/w60q8320

10. Potdar, A. (2024). AI-Based Big Data Governance Frameworks for Secure and Compliant Data Processing. International Journal of Artificial Intelligence, Data Science, and Machine Learning, 5(4), 72-80. https://doi.org/10.63282/3050-9262.IJAIDSML-V5I4P108
11. Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access, 6*, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052
12. AR Kommera. (2024). Visualizing the Future: Integrating Data Science and AI for Impactful Analysis. International Journal of Emerging Research in Engineering and Technology, 5(1), 48-59. https://doi.org/10.63282/3050-922X.IJERET-V5I1P107
13. Laxman doddipatla, & Sai Teja Sharma R.(2023). The Role of AI and Machine Learning in Strengthening Digital Wallet Security Against Fraud. Journal for ReAttach Therapy and Developmental Diversities, 6(1), 2172-2178.
14. Chen, X., Xu, L., Lu, Y., & Xu, Q. (2021). Machine learning for financial risk management in blockchain systems. *Journal of Risk and Financial Management, 14*(5), 225. https://doi.org/10.3390/jrfm14050225
15. D Alexander.(2022). EMERGING TRENDS IN FINTECH: HOW TECHNOLOGY IS RESHAPING THE GLOBAL FINANCIAL LANDSCAPE. Journal of Population Therapeutics and Clinical Pharmacology, 29(02), 573-580.
16. Rautaray, S., & Tayagi, D. (2023). Artificial Intelligence in Telecommunications: Applications, Risks, and Governance in the 5G and Beyond Era. Artificial Intelligence
17. Eskandari, S., Moosavi, S., & Clark, J. (2019). SoK: Transparent dishonesty: Front-running attacks on blockchain. *arXiv preprint arXiv:1902.05164*. https://arxiv.org/abs/1902.05164
18. Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a 'right to explanation'. *AI Magazine, 38*(3), 50–57. https://doi.org/10.1609/aimag.v38i3.2741
19. K Richardson. (2024). Navigating Challenges in Real-Time Payment Systems in FinTech. International Journal of Artificial Intelligence, Data Science, and Machine Learning, 5(1), 44-56. https://doi.org/10.63282/3050-9262.IJAIDSML-V5I1P105
20. Hemalatha Naga Himabindu, Gurajada. (2022). Unlocking Insights: The Power of Data Science and AI in Data Visualization. International Journal of Computer Science and Information Technology Research (IJCSITR), 3(1), 154-179. https://doi.org/10.63530/IJCSITR_2022_03_01_016
21. Li, Y., Li, M., & He, Y. (2020). Fraud detection using ensemble learning in electronic transactions. *Expert Systems with Applications, 139*, 112873. https://doi.org/10.1016/j.eswa.2019.112873
22. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems, 30*, 4765–4774. https://proceedings.neurips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html
23. Arpit Garg, "CNN-Based Image Validation for ESG Reporting: An Explainable AI and Blockchain Approach", Int. J. Comput. Sci. Inf. Technol. Res., vol. 5, no. 4, pp. 64–85, Dec. 2024, doi: 10.63530/IJCSITR_2024_05_04_007
24. Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system. *Bitcoin White Paper*. https://bitcoin.org/bitcoin.pdf
25. Pilkington, M. (2016). Blockchain technology: Principles and applications. In *Research Handbook on Digital Transformations* (pp. 225–253). Edward Elgar Publishing. https://doi.org/10.4337/9781784717766.00019
26. R. R. Yerram, "Risk management in foreign exchange for crossborder payments:Strategies for minimizing exposure," Turkish Online Journal of Qualitative Inquiry, pp. 892-900, 2020.
27. Ryman-Tubb, N. F., Krause, P., & Garn, W. (2018). How artificial intelligence and machine learning research impacts fraud detection in the financial services industry. *Journal of Financial Crime, 25*(2), 422–435. https://doi.org/10.1108/JFC-08-2017-0067
28. RA Kodete. (2022). Enhancing Blockchain Payment Security with Federated Learning. International journal of computer networks and wireless communications (IJCNWC), 12(3), 102-123.
29. Doddipatla, L. (2024). Ethical and Regulatory Challenges of Using Generative AI in Banking: Balancing Innovation and Compliance. Educational Administration: Theory and Practice, 30(3), 2848-2855.
30. M. Pandey, and A. R. Pathak, "A Multi-Layered AI-IoT Framework for Adaptive Financial Services", IJETCSIT, vol. 5, no. 3, pp. 47–57, Oct. 2024, doi: 10.63282/3050-9246.IJETCSIT-V5I3P105
31. Yang, G., & Li, Y. (2020). Explainable AI for blockchain: Concepts, tools, and future directions. *ACM Transactions on Internet Technology, 20*(3), 1–18. https://doi.org/10.1145/3379477
32. Ramakrishna Ramadugu. Unraveling the Paradox: Green Premium vs. Climate Risk Premium in Sustainable Investing. ABS International Journal of Management, Asian business school; ABSIC 2024 - 12th International Conference, Nov 2024, Noida, India. pp.71-89. ⟨hal-04931523⟩
33. Zhou, L., Wang, L., & Xu, Y. (2022). Interpretable anomaly detection for blockchain using autoencoders. *IEEE Transactions on Knowledge and Data Engineering, 34*(2), 589–603. https://doi.org/10.1109/TKDE.2020.2992653
34. Potdar, A. (2024). Intelligent Data Summarization Techniques for Efficient Big Data Exploration Using AI. International Journal of AI, BigData, Computational and Management Studies, 5(1), 80-88. https://doi.org/10.63282/3050-9416.IJAIBDCMS-V5I1P109

35. CT Aghaunor. (2023). From Data to Decisions: Harnessing AI and Analytics. International Journal of Artificial Intelligence, Data Science, and Machine Learning, 4(3), 76-84. https://doi.org/10.63282/3050-9262.IJAIDSML-V4I3P109
36. Binns, R., Veale, M., Van Kleek, M., & Shadbolt, N. (2018). 'It's reducing a human being to a percentage': Perceptions of justice in algorithmic decisions. *CHI Conference on Human Factors in Computing Systems*, 1–14. https://doi.org/10.1145/3173574.3173951