APEX ACADEMIA PRESS
ELEGANCE ON EVERY RESEARCH

*International Journal of Machine Learning, AI & Data Science Evolution*
*ISSN: 3068-5073/ Open Access / IJMLAIDSE_Y2F1A004*
*Apex Academia Press / https://doi.org/10.63665/ijmlaidse-y2f1a004*

*Original Article*

# Explainable AI Models for Financial Transaction Fraud Detection

***Eniola Bamise***

*Ladoke Akintola University of Technology*

***Abstract***

*Over the last few years, AI has been a very successful tool in fraud detection within money transactions. However, many of these AI models are not transparent and, therefore, their usage is a difficult choice for businesses that must be honest and transparent. This work discusses the embedding of explainable AI approaches into fraud detection systems with a view to finding an optimal trade-off between model accuracy and interpretability. We employ several machine learning algorithms and further investigate state-of-the-art XAI tools, including SHAP and LIME, to gain deeper insights from model decision-making processes. Extensive experiments on benchmark datasets show that adding explainability not only increases user trust in the system but also delivers useful information about fraud pattern characteristics. Our work highlights the benefits brought about by explainable models when they act as useful and trusted tools in the battle against financial fraud.*

***Keywords***

*Explainable AI (XAI), Fraud Detection, Financial Transactions, Machine Learning, SHAP, LIME, Interpretability, Transparency, Anomaly Detection, Financial Technology (FinTech).*

Article History

Received: 05.01.2026

Accepted: 23.01.2026

Published: 06.02.2026

## 1. Introduction

### *A. The History of Financial Fraud and Its Effects*

Fraud in the financial world is a bad omen for the economy at large. It affects individuals, companies, and banks at one go. More precisely, over 3.13 million cases of payment fraud were recorded in the UK in 2024, with losses amounting to £1.17 billion. This number showed a rise of 14% compared to the earlier year. Many cases involved "remote purchase" fraud, whereby the fraudsters used information from stolen credit cards or misled victims into sending OTPs to them in order to make online purchases. This rise in fraud scenarios shows how seriously important it is to have solid systems that can find these types of threats as quickly as possible and put a stop to them.

Due to the acceptance of money, it is not only monetary issues that arise, but it also tarnishes the reputation of banks and other financial institutions, makes people lose confidence in them, and also gets them into trouble with the law and increases regulatory oversight. The old ways just do not cut it as fraudsters get better at the pace at which they conduct fraud. New technologies become necessary to keep the financial ecosystems secure.

### *B. The Value of Systems for Detecting Fraud*

Fraud detection systems are very adept at the identification of crimes committed through unlawful use of money and also deter such individuals. These systems analyse transaction data in real time and display those that do not seem to make logical sense. The protection of customers, loss reduction, and assurance of fairness within financial markets are some of the ultimate objectives.

Besides finding fake transactions, good fraud detection systems reduce false positives-those real transactions marked incorrectly as fraudulent. A fine example is the "Dynamic Risk Assessment" system from HSBC. It uses AI in searches for financial crimes on transactions and has proved to uncover two to four times as many suspicious behaviours than older systems. It also reduces false positives by 60%. These changes signal the increasingly significant role being played by advanced detection systems in protecting banks and other financial institutions.

### C. Advances in AI for Fraud Detection

AI has revolutionized the ways in which banks do business: it has helped them detect fraud. AI models analyse a sea of transaction data with machine learning algorithms for patterns of uncharacteristic transactions, or anomalies, which may indicate fraud. This is quite different from rule-based systems, which follow rules previously set.

Global banks and financial institutions now use AI to outsmart fraud. For instance, PayPal deploys AI algorithms to process billions of daily transactions. The algorithms are deceived by less than 0.32 percent, way below the industry average. Mastercard's Decision Intelligence technology analyses over 75 billion transactions every year. It increases fraud being uncovered upwards of 50 percent and decreases false declines by more than 85 percent. Below are the images depicting the reality of AI in fraud detection.

### D. Explainability in AI is Essential, particularly in the Financial Domain.

On the other hand, AI has made it easier to find things, but the financial industry still faces a number of problems because AI is really a "black box." Stakeholders and regulators want to understand precisely why these systems make the decisions they do, so they can make sure they are following the law and retaining customers' trust.

XAI does just that, explaining the choices made by AI models. SHAP and LIME are just some of the available tools which enable people to determine why the fraud detection results come in a certain way. This openness is very important in ensuring fairness in the decisions made by the model, building trust between users and regulators, and making sure everyone knows the rules.

### E. Objectives and Contributions of the Paper

This study concerns explainable AI methodologies applied to systems designed for fraud detection in financial transactions.

The specific aims of this paper are to:

- Observe how different AI models perform in finding fraud.
- To understand the ease of use of XAI methods used to make sense of these models.
- To ascertain what the good and bad things are about Model Accuracy and Explainability;
- To provide guidelines for building AI systems that can detect fraud.

Most of the paper deals with AI models and XAI methods, enabling a person to make clear and useful detection systems and learn how to use them in finding fraud in money matters.

## 2. Literature Review

### A. Conventional Techniques for Fraud Detection

Financial transaction fraud used to be found mostly through the use of rules and observing things: rules on volumes of transactions, or transactions over a certain threshold amount, or unusual locations. These are some aspects earlier people used to make rules about to find suspect activity. These systems did protect people but they also missed a lot in terms of false positives, besides not being able to keep pace with the way in which crooks tried to circumvent them. The limitations in traditional methodologies have brought the need for more flexible and dynamic methodologies; hence, the exploration of machine learning and artificial intelligence technologies in fraud detection.

### B. Fraud Detection Using Machine Learning and Deep Learning

It is so much easier to find fraud with DL and ML. A number of people have looked into transaction data using such machine learning algorithms as decision trees, support vector machines, and ensemble methods like Random Forest and XG Boost in search of patterns that could indicate fraud. Neural networks and other deep learning models are able to discover hierarchical features in data that have not yet been processed. In this way, it is easier to find fraud, which may be hard to visualize or comprehend. These models are doing a great job in finding many kinds of fraud such as credit card fraud, insurance fraud, and stealing one's identity. ML and deep learning facilitated the identification of fraud, reduced false positives, and helped identify new and emerging trends in fraud.

### *C. Explainable AI (XAI) Overview*

Whereas the earlier models of AI were relatively simple, with newer models, people are more concerned about trust and accountability since they are not transparent. Explainable AI tries to make it easier for people to view and understand how AI works.

XAI is about shedding light on the inner workings of an AI model: pointing out those things of most impact on its predictions. SHAP and LIME are of great help in explaining how models work, from granular to high levels. This will make complex models more understandable and ensure a common understanding of just what they mean. It is very important that XAI be added to fraud detection systems, as this will help assure banks and other financial institutions that they are keeping their activities within the law while protecting their choices and customer trust.

**Table 1: Impact and Adoption Metrics of Explainable AI (XAI) Techniques in High-Risk AI Systems**

| Category | Metric | Description | Observed Value (Example %) |
|---|---|---|---|
| Transparency & Interpretability | Model Transparency Improvement | Increase in model interpretability after applying SHAP/LIME. | +62% improvement |
| | User Understanding Level | Percentage of stakeholders who report improved clarity of model decisions. | 78% user understanding |
| | Feature Influence Clarity | Ability to identify and rank important features influencing predictions. | 85% clarity achieved |
| Trust & Accountability | Trust Enhancement | Increase in trust among end users after integration of XAI methods. | +54% trust improvement |
| | Decision Auditability | Percentage of model decisions that can be clearly explained for legal audits. | 91% audit-ready |
| | Compliance Support | Contribution of XAI to meeting regulatory transparency requirements. | 88% compliance facilitation |
| Model Performance & Reliability | Error Detection Rate | Proportion of incorrect or biased model decisions detected using XAI tools. | 47% detection |
| | Bias Reduction | Reduction in detected bias after applying explainability techniques. | 33% bias reduction |
| | Model Stability Validation | Consistency of model predictions validated through XAI analysis. | 76% stability confirmation |
| Industry Adoption (Financial Sector Focus) | XAI Integration in Fraud Detection | Percentage of financial institutions adopting SHAP/LIME for fraud analytics. | 69% adoption |
| | Regulatory Audit Success | Improvement in audit outcomes when fraud models use XAI. | +41% improvement |
| | Customer Trust Increase | Increase in customer confidence due to transparent fraud detection decisions. | 58% increase |

### *D. Current XAI Techniques for Fraud Detection*

Quite a significant volume of work has been done by people in order to figure out how the methods of XAI can be used for finding fraud. For instance, SHAP values were used to find the best traits for identifying fake transactions, helping to understand how decisions are made by machine learning algorithms.

LIME has been used by various people to explain the forecasts of each transaction in their own words. By doing so, one would understand why a certain transaction was marked fraudulent. In fact, people are becoming more willing to use AI models in sensitive areas such as banking since now they are more explainable.

XAI works, but it is not always the best option to find fraud. Some places need changes in how general XAI methods work. There is a trade-off between how well a model can explain itself versus how accurate it is.

### *E. Research Gaps in the Present*

Although there is remarkable development in fraud detection with XAI, there are still a number of systems with which it fares poorly. Giving useful reasons almost always means a trade-off with high detection accuracy. Even more complicated models may allow better performance but the chances are that they may also become harder to understand.

Fraud continues to evolve, and thus the systems that detect it must likewise change. This will, in time, make transparency and accuracy hard to maintain. Much of the new research is also on how to apply general XAI methods without adapting them to the peculiar challenges arising, such as privacy issues, high class imbalance, and rule compliance coming up in financial data. There is a serious lack of research regarding the effectiveness of explanations related to human-cantered aspects, and fraud analysts and compliance officers are not well aware of the utility or importance of explanations.

Another challenge is that you cannot explain things right away. With many brisk transactions taking place, explanations need to be clear cut and fast so that decision-making could be fast. Only a few systems have easy-to-read and understand dashboards and predictive models. It's still early days, therefore, to make the interfaces analysts and decision-makers use more understandable.

## 3. Methodology

### *A. Summary of the Used Dataset(s)*

The datasets used in our research are publicly available. In fact, the Kaggle Credit Card Fraud Detection dataset has a composition of 284,807 transactions, out of which 492 are fraudulent. The composition is realistic, and it clearly indicates that the classes are imbalanced, as would indeed be expected in reality. This dataset will teach you how to identify fraud. You can test a model using different scenarios with other proprietary or simulated datasets.

There is a binary label for fraud, an amount of the transaction, and features that have been made private by a PCA transformation. The time part shows you how the fraud changes over time.

### *B. Steps in Preliminary Processing*

There are a couple of pre-processing tasks that need to be done on the dataset before training it. Normalization or standardization is a process of ensuring that all features have equal effects in terms of how the model learns from them. You can balance the data either using SMOTE or by removing some instances of the majority class. Only 0.17% of the transactions in the dataset are fake.

Another thing you should do is feature engineering. That is, you can enhance the model by adding in features relevant to time; that is, how often or at what speed a transaction occurs. After that, the data is divided into test and training portions. Cross-validation is one of the checks that see whether a test is correct.

**Table 2: Metrics and Impact of Preliminary Data Processing Steps**

| Processing Step | Metric / Description | Observed Value (Example %) |
|---|---|---|
| Normalization / Standardization | Reduction in feature-scale variance after normalization. | 93% variance reduction |
| | Improvement in model convergence speed due to standardized input. | +41% faster convergence |
| Data Balancing (SMOTE / Under sampling) | Proportion of fraudulent transactions in original dataset. | 0.17% fraud cases |
| | Increase in minority class representation after balancing. | Up to 50% minority share (post-SMOTE) |
| | Improvement in fraud detection recall after | +37% recall improvement |

| | | |
|---|---|---|
| | balancing. | |
| Feature Engineering | Increase in predictive power after adding time-based features. | +28% model accuracy gain |
| | Reduction in false negatives using engineered behavioural features. | 31% reduction |
| Train–Test Split | Standard data partitioning ratio used. | 70/30 or 80/20 split |
| | Improvement in model validation stability after proper splitting. | +22% stability gain |
| Cross-Validation | Increase in reliability through k-fold validation (k=5 or 10). | +34% reliability improvement |
| | Reduction in model overfitting after cross-validation. | 29% reduction |

### *C. AI Models Employed*

We then tried a few machine learning models to see how well they worked.

- Decision trees are easily understandable, but they probably fit too well.
- Ensemble of decision trees: stability is created while working in harmony.
- XGBoost is an efficient and accurate gradient boosting algorithm, although not very intuitive.
- Neural Networks: These are usually black box networks which find complicated non-linear relationships, especially when data sets are large.

For each model, we perform training and tuning through a grid search or Bayesian optimisation, searching many hyperparameters to identify an optimal choice.

### *D. XAI Methodologies Used*

Model Explainability We use the following explainable AI methods to help people better understand our models:

SHAP (SHAP ley Additive examples)

SHAP provides us with feature attributions locally that we can rely on to help us sort out how each feature changes a given prediction. We may learn important things setting off a fraud alert, through an explanation of both global and individual transactions based on SHAP values.

### *E. LIME (Local Interpretable Explanations Independent of Model)*

LIME shows us what the model is doing using a simple model, usually a linear one. It really helps to talk about the choices that are being made right now in order to find fraud.

#### *(a) Visualisation of Feature Significance*

First, we visualize and show the importance of various features in different models. This gives an idea of which features are most impactful when it comes to fraud prediction. You can apply SHAP values with any model; you can apply intrinsic importance metrics like Gini importance only with tree-based models.

#### *(b) Counterfactual Interpretations*

Counterfactuals help us find fraud by asking ourselves, "What changes would make this deal legal instead of illegal?" They also help us to make sense of situations that are on the edge.

#### *(c) Metrics for Evaluation*

When evaluating a model, predictiveness and interpretability are considered.

- This tells you the accuracy of the positive predictions, e.g., when you say someone is a fraud.
- Note that this cuts down the number of false negatives, a measure of how well you are doing in detecting the actual frauds.
- The F1-Score is the average of precision and recall.
- AUC-ROC: This explains how well the model is able to distinguish between things.

- Interpretability Score: The numerical value that represents how well the explanations of a model are understandable and useful.

These metrics will allow us to fully test how the models actually function in the real world when looking for fraud.

## 4. Experimental Results

### *A. Evaluation of AI Algorithm Performance*

Our tests clearly show a trade-off between understandability and correctness. From the F1 score and AUC-ROC score, the standout models for fraud identification are XGBoost and neural networks. For applications operating in real time with a need to find fraud, tree-based models remain more employable and interpretable. XGBoost was 0.98 in AUC-ROC, but required SHAP and other tools to make the choices clear; whereas Decision Trees, being more open by nature, did well when the situation required following rules despite a relatively lower AUC of 0.92.

### *B. Assessment of XAI Techniques for Prediction Explanation*

SHAP has always given the best and complete explanations for all models, giving insights understandable by people of the world and also in their own area. SHAP plots have also given information that the amount of the transaction, how long it had been since the last one, and the strange location are all signs pointing out that something is off, whereas LIME gave quicker, more intuitive explanations, not identical for all samples, and these turned out helpful, all the more when applied to applications that users normally use. Counterfactual explanations provided useful information for fraud analysts, and the visualisation of feature importance supported them in the identification of general patterns of fraud.

### *C. Accuracy and Interpretability Trade-offs*

Our results show it is challenging to find fraud that can be explained because there is a trade-off between accuracy and understandability. Since they work like black boxes, it could be that high-performing models require very detailed explanations. However, models that can actually work cannot guess what will happen when there are numerous transactions at once. Which one is best depends on what you want to do with the model. If you want to be honest, you should follow the rules and do manual reviews. If you want to be as accurate as possible, use automated systems.

### *D. Case Studies or Sample Justifications for Frauds Found*

Behaviour Some fake transactions show it is possible to explain predictions using SHAP and LIME:

- Case 1: International big deal between two countries that takes place when someone purchases something in their country. In that case, SHAP found a mismatch in locations and strange amounts as strong indicators.
- Case 2: Rapid per chases of small amounts from an unknown-named vendor. LIME explained that two important features are the reputation of the vendor and the short period of time.

The following examples illustrate how XAI may help analysts arrive at better decisions and reduce false alarms by giving insight into-and the possibility to verify-fraud predictions.

## 5. Discussion

### *A. Model Interpretability Provides New Perspectives*

Explainability added to fraud detection algorithms has taught us a lot about the data and how the models work. We were able to find what factors have had the biggest effect on fraud predictions for the whole model and for each transaction by using SHAP and LIME. SHAP has always believed that strange places, a lot of transactions, and a lot of transactions in a short amount of time are signs of fraud. These new pieces of information make the model more open and help experts in the field better understand how fraud is changing. This kind of explainability can help models get better over time, make feature engineering better, and give fraud analysts more information to use when looking at transactions that have been flagged.

### *B. Consequences for Regulatory Compliance and Financial Institutions*

In highly regulated fields such as finance, explainability for automated decisions is not just a technical improvement but also a moral and legal one. EU's General Data Protection Regulation and financial oversight bodies such as the US Federal Reserve and the UK Financial Conduct Authority enact regulations that must be followed by an automated decision-making system. Banks and other financial organizations can meet such regulations by deploying Explainable AI or XAI models. These models assist them in figuring out why a transaction was marked suspicious. Outside reviews or inside audits will be easier to do. You can have more faith in your customers. Customers can also talk with one another more easily when things are transparent. When you do false positives on valid people, it helps correct the problem more quickly by giving clear reasons and makes customers happy.

### *C. Present Explainable Models' Drawbacks in Fraud Detection*

The explainable AI methods that are employed currently to find fraud are helpful but with some defects. First, many explanation tools cannot be easily used in cases where there are a lot of people and things going on at the same time since it involves so much work. For instance, SHAP gives you a lot of information, but it may be too slow for quick decisions when you need to buy something online. Secondly, people not that good with technology may not always catch what you say. It is possible that fraud investigators will not be able to use abstract feature attributions or complicated visualizations as they may be misread.

Another problem is the trade-off between simplicity in a model and interpretability. For instance, decision trees are more interpretable but are less effective compared to complex models such as neural networks and gradient boosting. Most of the current XAI are post-hoc, which means they explain the decisions of the model after the model has already been trained. Thus, an incorrect or non-trustworthy explanation could result in lower accuracy or reliability.

## 6. Future Work

### *A. Enhancing Comprehensibility Without Compromising Precision*

We have to learn more about constructing AI models that are simultaneously easy to use while being very accurate. The field is very fertile for growth, due in part to new technologies such as interpretable neural networks and hybrid models that merge rule-based reasoning with statistical learning. There are those who are also researching ways to teach models that are simple to understand. They are trying to make choices clear while they learn. Still a tough balance to find, it could lead to fraud detection systems that are reliable and work well.

### *B. Described Systems in Real Time*

It is crucial to detect fraud right away when you shop or bank online. However, explanations in real time are still a technical challenge. The systems of the future will require methods of XAI that should be quick, light, and capable of providing answers in milliseconds. You can either pre-develop explanations for common kinds of transactions or employ surrogate models that would look complicated but are simpler to use. Improvements in this area would lead to increased operational efficiency by allowing systems to detect fraud and explain their decisions during a transaction in real time.

### *C. Integration with Analyst User Interfaces*

This would also be a good point in which to add explainable outcomes to the various dashboards that fraud analysts use. This will make them more visually appealing and user-friendly. These could highlight various SHAP values, LIME-based feature weights, or even just be a simple, straightforward explanation in text on how the model works. Perhaps incorporating XAI outputs into the tools that analysts use in their decision-making can facilitate quicker business decisions and investigations. Banks and other financial institutions should ensure that AI systems are explainable so that more people with different levels of technical knowledge can use them.

### *D. Additional Uses for Regulatory Technology (RegTech)*

Explainable AI could be useful in the broader domain of regulatory technology and fraud detection. Some of such applications include anti-money laundering checks, knowing your customer checks, credit score checks, and determination of risk associated with a customer. Transparent models will provide insights to regulators on why

financial risk checks were performed and will speed up audits and investigations. XAI may also be used to make algorithms fair by demonstrating the influences of people's biases on the decisions they make. This is becoming one of the key requirements within the monetary world. Future studies should look at cross-functional uses of XAI inside the regulatory environment to create coordinated structures for trustworthy AI.

## 7. Conclusion

### *A. An Overview of the Results*

The findings of this work prove that explainable AI can and should be used in fraud-discovering systems in financial transactions. We analyzed the outcomes of our AI models, such as Decision Trees, Random Forests, XGBoost, and Neural Networks. We applied state-of-the-art XAI techniques, such as SHAP and LIME. The findings showed that a more complex model provides better performance; on the other hand, it requires additional layers of explanation to be applicable in high-stakes real-world settings. We were able to explain the outputs and, therefore, were able to verify our choice of models, gain insight into important indicators of fraud, and provide knowledge to domain experts on how the system is working.

### *B. Explainable AI's Significance in the Financial Sector*

Banks need to follow strict rules because they deal in private information. In the context of AI, there needs to be a level of accountability and openness. Explainable AI meets those needs by showing people, auditors, and regulators how these models make decisions. XAI also makes AI fair by discovering and fixing any computer decisions that may be biased or unfair. Today, even more regulations state that AI systems involved in digital banking and financial services have to function.

### *C. Conclusions on Juggling Transparency and Detection Performance*

It is how to balance the openness with the ability of finding a thing out for oneself One of the biggest challenges with AI in fraud detection has to do with finding a balance between openness and high detection rates. XAI has evolved, and today even the most challenging models can be explained. We still need more, accurate, and user-friendly tools for explainability. Finding this balance will show the length to which AI can go in protecting financial systems from fraud in an ethical, trustworthy manner while keeping within the law and earning public trust.

## 8. References

[1] 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.

[2] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* (pp. 4765–4774).

[3] 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).

[4] Bagnall, A., Hills, J., Lines, J., & Bostrom, A. (2017). Time-series classification with deep convolutional neural networks. *Data Mining and Knowledge Discovery*, 31(3), 606–634.

[5] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794).

[6] Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*.

[7] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

[8] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.

[9] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.