JOURNAL OF DATA SCIENCE | Vol.2026, Issue 3, No.2
eISSN:2805-5160

Explainable Fraud Detection Using Ensemble Learning and SHAP-Based Feature Interpretation

Shah Khadafi[1], Julius Cesar O. Mamaril[2]*, Clark Kim C. Castro[2]

[1]Information System Department, Institut Teknologi Adhi Tama Surabaya, East Java, Indonesia,
[2]ICT Management Office, Pangasinan State University, Philippines

***Email:** jmamaril@psu.edu.ph[2]

Abstract

Financial fraud detection is a challenging data science problem because fraudulent transactions are rare, adaptive, and costly when not detected early. Machine learning models can identify suspicious transaction patterns, but many high-performing models remain difficult to interpret, limiting their use in audit-oriented financial decision support. This study proposes an explainable fraud detection framework combining ensemble learning and SHAP-based feature interpretation. Random Forest, XGBoost, and LightGBM were compared with a soft-voting ensemble model using a controlled synthetic financial transaction dataset of 20,000 records, including 400 fraudulent and 19,600 legitimate transactions. Performance was evaluated using accuracy, precision, recall, F1-score, and AUC-ROC, while SHAP was used to explain global feature importance and local prediction outcomes. The voting ensemble achieved the strongest performance, with accuracy of 0.991, precision of 0.729, recall of 0.875, F1-score of 0.795, and AUC-ROC of 0.973. SHAP identified merchant risk score, customer velocity, transaction amount, location distance, and failed attempt count as the most influential fraud indicators. The findings show that explainable ensemble learning can improve fraud detection performance while supporting transparent and auditable fraud analytics.

Keywords

Fraud detection, Ensemble learning, Explainable artificial intelligence, SHAP, Data mining, Process innovation

Introduction

Fraud detection has become critical in data science as financial transactions consist of large amounts of legal transacting activities with only small amounts of illegal activities (Ali et al., 2022). Digital payments, online banking, e-commerce, mobile wallets, and other financial technologies have rapidly expanded the data processed by financial institutions. While improving the convenience of financial transactions, these technologies have also heightened the risk of

Submission: 20 September 2026; **Revised:** 15 October 2026 **Acceptance:** 13 November 2026; **Available Online:** November 2026
Copyright: © 2026. All the authors listed in this paper. The distribution, reproduction, and any other usage of the content of this paper is permitted, with credit given to all the author(s) and copyright owner(s) in accordance to common academic practice. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license, as stated in the website: https://creativecommons.org/licenses/by/4.0/

fraudulent activities with the use of stolen credentials, synthetic identities, compromised devices, and unusual transaction sequences.

The importance of AI in fraud detection systems is also due to the capacity of AI to support the real time detection of fraudulent patterns and technologies, as well as the strengthening of systems to detect and prevent fraud (Leong et al., 2026). While traditional rule-based systems are useful to identify known fraud patterns, these systems have limitations when adaptive fraud patterns are exhibited. Machine learning creates an adaptive methodology compared to rule-based systems, because with the right features, a learning model is created to identify complex transactional patterns.

For structured transaction data, learning models that employ trees and models of gradient boosting (e.g. XGBoost or LightGBM) are especially useful for transactional data because they can also accommodate behavioral, temporal, and risk-related features (Liu, 2026). Fraud prediction may focus on the interaction of merchant risk, transaction amount, customer velocity, device change, and unusual transaction patterns. These features may be captured by a learning model more readily than by simple linear rules.

Predictive performance is just one aspect of financial risk management (Nahar et al., 2024). Consider the case of a fraud detection system. A system may render a decision concerning customers, investigators, auditors, and compliance personnel. If a model outputs a fraud label and provides no justification, the model may be correct, but the label is difficult to trust. Thus, interpretability of a model is imperative for its acceptance.

In fraud detection, interpretability is important, on top of the reasons given previously, because alerts in fraud detection systems are typically not used in isolation (Zafar&Wu, 2026). Fraud analysts rely on model outputs in conjunction with the customer's transaction history, the customer's transactions, organization's risk policy, and the priorities of the case being investigated. In the context of financial institutions, fraud detection systems are required to strike a balance between the prevention of fraud, the customer experience, and the operational efficiency of the financial institution. This makes interpretability of fraud detection systems even more important.

Explainable artificial intelligence offers a pathway to addressing this issue (Mathew et al., 2025). In the context of artificial intelligence, SHAP values are of particular interest. SHAP is an acronym for SHapley Additive exPlanations and can be used to describe the impact each feature has on a fraud detection system. In fraud detection systems, SHAP can be used to explain if a transaction was flagged as suspicious due to an elevated risk associated with a merchant, the velocity of a transaction was flagged as suspicious due to an elevated risk associated with a merchant, the transaction was of an elevated value, an anomaly was identified due to a location, or due to repeated failed attempts, among others.

Most research on fraud detection has focused on classification performance, and few studies create a compact framework that finds a balance between predictive accuracy and feature-level explanation (Chen et al., 2026). Our objective is to create and evaluate a framework for explainable fraud detection using Random Forest, XGBoost, LightGBM, and a soft-voting ensemble model, in combination with SHAP interpretations. The contribution of this research lies in demonstrating the value of explainable ensemble methods for fraud detection by integrating predictive power, feature importance, and local explainability all in the same framework.

Related Work

Fraud detection is considered a classification problem in the field of data mining and machine learning (Hashemi et al., 2023). Here, the aim is to classify a transaction as either legitimate or fraudulent using observed behavioral, transactional, and risk factors. Traditionally, fraud detection employed techniques based on expert knowledge to define rules and set thresholds and blacklists. These techniques are attractive because they are straightforward, easy to verify, and can be quickly integrated into operational systems. Nonetheless, the rule-based detection system quickly becomes ineffective if the fraudulent behavior evolves quickly, if the fraud patterns are characterized by complex interactions among the features, or if fixed thresholds result in a high incidence of false alarms.

To solve the aforementioned problems, machine learning was employed which is used to directly learn patterns of interest from the labeled transactional data (Xie et al., 2024). A wide range of supervised classification techniques has been used to detect fraud, including logistic regression, decision trees, Random Forest, support vector machines, gradient boosting, neural networks, and hybrid models. Tree-based ensemble methods represent one of the best choices for working on tabular financial data as they can capture transactional data characterized by relational features, including the amount of the transaction, the risk of the merchant, the behavior of the customer and the device, the frequency of the transaction, and the transaction's geographical features. Compared to a single classifier, ensemble models tend to be more stable by integrating multiple learning techniques.

In fraud detection, tackling imbalanced datasets is of primary concern (Al-dahasi et al., 2025). A major part of any dataset consists of legitimate transactions, while only a small percentage consists of fraudulent transactions. Under these circumstances, accuracy is a poor metric. A model may classify legitimate transactions accurately, while still miss a large number of fraudulent transactions. Therefore, in many of the fraud detection literature, precision, recall, F1-score, AUC-ROC, and confusion matrix are analyzed. At times, there's a trade-off between precision and recall. This is acceptable due to the fact that missed fraud cases contribute to a loss, while excessive false alarms increase the workload of analysts and contribute to a loss of trust in the fraud detection system.

In the presence of an imbalanced and structured dataset, ensemble learning is a good approach in fraud detection (Khalid et al., 2024). Random Forests are robust and flexible on the basis of multiple decision trees. In contrast, XGBoost and LightGBM rely on gradient boosting and work on improving and learning the complex relationships between the features, thus improving the classification. In this case, a soft-voting ensemble can be used, and classify based on the averaged outcome of multiple models. This is very useful in the context of fraud detection as it allows organizations to alter their thresholds of accept and reject based on their willingness to take on the risk, their current capacity to investigate, and their operational priorities.

Interpreting well-developed fraud detection models can still be difficult, despite the benefits such models provide (Ashfaq&Chowdhury, 2023). This creates a limitation in the practicality of these models since fraud alerts will be utilized by analysts, auditors, and compliance and risk managers who will want an explanation for why a transaction is flagged as suspicious. Explainable AI offers a solution to this constraint by providing the model's rationale through evidence at the feature level. This is especially true of SHAP since it can provide both global and local explanations, such as determination of the transaction data with the highest potential to

represent fraud risk and reasoning as to why a specific transaction is classified as fraudulent, respectively.

Fraud detection literature captures three main streams of activity: enhancing classification accuracy, addressing class imbalance in transaction data, and improving model interpretability (Zafar&Wu, 2026). Yet, a considerable number of studies still focus more on predictive metrics than explanation-driven decision support. This research addresses this gap by combining tree-based ensemble learning with SHAP for the first time. The proposed methodology focuses on the model's performance in the context of the transaction's features that are critical determinants of fraud prediction, thereby enabling a more verifiable and transparent decision support system for fraud analytics.

Methodology

Research Design

This research applied supervised data mining for fraud classification. The workflow of this research included data preparation, preprocessing, training, evaluation, and SHAP explainability. The targeted variable was binary. The positive class was 1 for fraudulent, and the negative class was 0 for not fraudulent. Three stand-alone classifiers were used versus a soft voting ensemble.
The design of the experiment had two key goals. The first was to determine whether the soft-voting ensemble classifier achieved higher performance for the detection of fraudulent transactions when compared to the stand-alone classifiers. The second was to determine whether SHAP was capable of providing useful explainability of the fraudulent transaction predictions by determining both the global and local feature contributions. These goals also describe the two-dimensional nature ofthis study, that is, to increase performance of the classification and to increase the explainability of the classification.

Dataset Construction

To balance class distribution while keeping client data secure, a controlled, synthetic financial transaction dataset was developed. This dataset included 20,000 transaction records and was predominantly made up of valid transactions (19,600) with a minority of fraudulent transactions (400). This resulted in a synthetic dataset with a fraud rate of 2%. This number is low, as most fraud detection datasets are extremely imbalanced.

To simulate some of the most fundamental behavioral patterns usually observed in transactional financial services, synthetic data was generated to display the behavior. Legitimate transactions were designed with customer behavior in mind. This includes stable, low-risk merchants, low amounts, low transactional velocity, stable locations and low risk of a change of devices. In contrast, fraudulent transactions were designed to capture the opposite. This includes high (and changing) risk merchants, high velocity transactions, large amounts over a long distance and failed attempts.

The dataset was designed with a behavioral, transactional and risk-based approach to fraud and analytics. In total, 10 transaction-based features were included. These features were transaction amount, merchant risk score, customer's velocity in the last 24 hours, failed attempts, a device change (yes/no) indicator, distance traveled (yes/no) and time (hr) elapsed since the last

transaction. Additionally, features of account age and transaction frequency in the last 7 days were included.

Table 1. Dataset structure and experimental setting

Component	Description
Dataset type	Controlled synthetic financial transaction dataset
Total records	20,000 transactions
Legitimate transactions	19,600 transactions
Fraudulent transactions	400 transactions
Fraud rate	2%
Number of features	10 transaction-level features
Target variable	Fraud status: 0 = legitimate, 1 = fraudulent
Training-testing split	70% training data and 30% testing data
Training records	14,000 transactions
Testing records	6,000 transactions
Training legitimate cases	13,720 transactions
Training fraud cases	280 transactions
Test legitimate cases	5,880 transactions
Test fraud cases	120 transactions

Table 1 shows the class distribution, dataset structure, train-test split, and more. The features selected for this task include transaction features as well as features based on customer behavior patterns. With this structure, a model that captures not just larger or irregular transactions, but also broader suspicious patterns, may be developed. This includes customer behavior patterns, such as an irregular customer velocity, transactions that are placed from different locations, repeated transaction attempts, and switching devices.

Although the dataset is synthetic, the features are designed to mimic realistic situations for monitoring fraud. In order to classify fraudulent transactions and legitimate ones, it is not enough to just set extreme values for features, as that would simplify the task too much and would be counter-productive. To address that, models are designed in a way that the behavior patterns of both legitimate and fraudulent transactions overlap to some extent. This way, models learn to distinguish and identify multiple risk indicators.

Data Preprocessing

As part of the preprocessing stage, we checked for and accounted for missing values, removed duplicates, encoded categorical variables, and scaled variables when needed. During this stage, we encoded two binary variables — the device change indicator and card-present indicator — into numerical format. We then created a stratified, 70:30 train-test split that maintained the overall 2% fraud sample rate in each subset. To address class imbalance, we used class weighting, and we did not sample the test set to ensure that the evaluation aligned with the original fraud distribution.

Model Development

Three separate classifiers were developed: Random Forest, XGBoost, and LightGBM. These three were selected because each have appropriate strengths for working with structured transaction data and the ability to identify non-linear relationships in the data. All three were trained and tested on the same data to ensure appropriate comparisons could be made.

Random Forest was implemented as a robust baseline ensemble model. As an ensemble model, Random Forest shares the strength and limitation of combining many decision trees: improved robustness; diminished variance; and, potential overfitting. XGBoost was selected, as it is a favored model for high-speed execution and high-performance classification of structured data. XGBoost builds trees in a sequential manner and uses gradient boosting to address the mistakes of the previous iterations. LightGBM was selected, as it is a fast implementation of a gradient boosting framework and focuses on the best performance for large datasets. Like the previous two, LightGBM is widely utilized in structured classification tasks with non-linear relations.

A soft-voting ensemble model was built by forming an average of the prediction probabilities of the three classifiers. The mean of the predicted fraud probabilities given by the three classifiers was taken, and the final fraud probability was determined. If the final fraud probability was above the classification threshold, the transaction was determined to be fraudulent. This method enables the ensemble model to utilize the advantages of each individual classifier while increasing the consistency of prediction.

The experimental design was built to unite the two complementary aspects of a workflow: prediction and explainability, in a single methodology. The final fraud classification is given by the ensemble model, and a SHAP framework is utilized to provide an explanation of the decision by modeling the behavior in a structured way through both global and local interpretations.

Evaluation Metrics

Model performance was assessed by accuracy, precision, recall, F1-score, and AUC-ROC. Accuracy was stated as an overall correctness measure. However, precision, recall, and F1-score were favored because of the imbalanced data that fraud detection involves. AUC-ROC was calculated to determine how well the model could differentiate between fraudulent and legitimate transactions at varying classification thresholds. Confusion matrix values were presented to show how many true positives, false negatives, false positives, and true negatives were observed.

SHAP-Based Explainability

SHAP was used to understand model predictions based on individual features. Two levels of SHAP analysis were used. Global interpretation analyzed the most important features from the test set. Local interpretation analyzed the classification of a single transaction as fraudulent.

Global SHAP interpretation aids a strategic analysis by identifying the features that predict fraud most strongly. Local SHAP interpretation aids an analysis by explaining the prediction of an individual transaction. Both interpretations assist fraud analysts by explaining model behavior on the macro level and case decisions on the micro level.

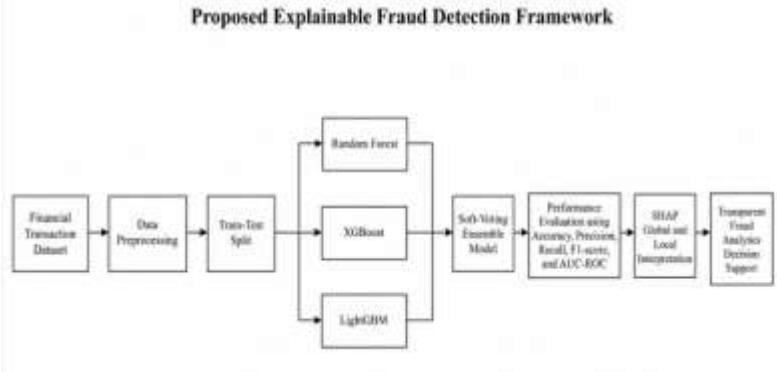


Figure 1. Proposed explainable fraud detection framework

The proposed framework in Figure 1 starts with preparing transaction data and extends to preprocessing, train-test data segmentation, training different classifiers, soft-voting ensemble prediction, performance assessment, and transparent fraud analytics using SHAP interpretation.

Results and Discussion

Model Performance

Each model was tested against the same set of 6,000 transactions, containing 120 fraudulent transactions and 5,880 legitimate transactions. Performance results of the individual classifiers and the soft-voting ensemble model are found in Table 2.

Table 2. Model performance and fraud detection results

Model	Accuracy	Precision	Recall	F1-score	AUC-ROC	TP	FN	FP	TN
Random Forest	0.987	0.639	0.783	0.704	0.943	94	26	53	5,827
XGBoost	0.989	0.682	0.842	0.754	0.962	101	19	47	5,833
LightGBM	0.990	0.702	0.825	0.759	0.964	99	21	42	5,838
Voting Ensemble	0.991	0.729	0.875	0.795	0.973	105	15	39	5,841

The results indicate high accuracy for all models, due to the prevalence of legitimate transactions in the dataset. However, fraud detection relies more heavily on other metrics, as models that prioritize accuracy can overlook fraud. Among the three models, the voting ensemble—representing a combination of Random Forest, XGBoost, and LightGBM—offered the best results in terms of fraud detection, as indicated by the recall, F1 score, and AUC-ROC, demonstrating that the ensemble achieved a better tradeoff between detection of fraud and control of false alerts.

According to the voting ensemble's confusion matrix, 105 detected fraud cases were correct and 15 were missed. The voting ensemble also identified 39 legitimate transactions as fraudulent, and correctly identified 5,841 legitimate transactions. This result shows that the ensemble model maintained a low false alert rate while capturing most fraud cases.

Random Forest identified 94 of 120 fraud cases, and generated a recall rate of 0.783. XGBoost also performed better, detecting 101 cases of fraud and generating a recall of 0.842. LightGBM offered a slight improvement in precision over XGBoost, but also detected fewer cases of fraud. The soft-voting ensemble was the most successful model, achieving the lowest count of false negatives while also identifying 105 cases of fraud.

The ensemble received the highest precision rating at 0.729. Thus, the predicted fraud alerts were likely more fraudulent than in the individual models. More alerts are not necessarily a better outcome in fraud detection. Employee fraud investigation resources are finite. A model with low precision will generate too many alerts, and staff will not have the time to investigate all the alerts, leading to low system confidence and high model rejection. The ensemble managed to improve recall with a reasonable false positive level.

Fraud detection models do not always require the highest accuracy. The chosen model should minimize the number of missed fraudulent cases, and not overwhelm fraud analysts with alerts. The voting ensemble achieved better fraud recall than the individual models with fewer false alerts, showing that in fraud cases, ensemble aggregation of models can be effective to support a more reliable decision prioritization in imbalanced transaction scenarios.

SHAP Global Feature Interpretation

SHAP analysis was performed on the voting ensemble model for interpretation purposes. The most impactful features included the merchant risk score, customer velocity in the last 24 hours, transaction amount, distance of the location, count of failed attempts, device change flag, hour of the transaction, frequency of the transaction in the last week, age of the account, and card-present flag.

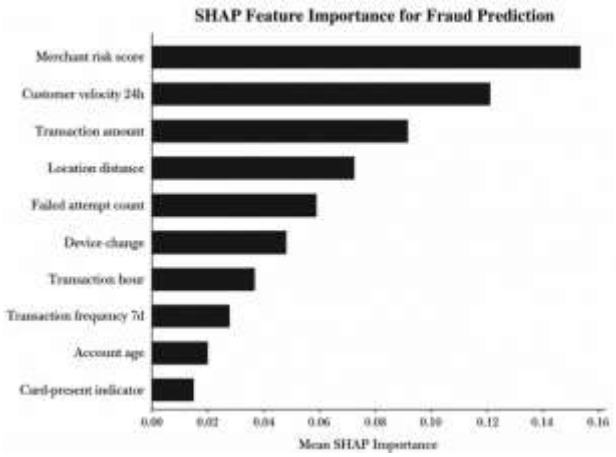


Figure 2. SHAP feature importance for fraud prediction

Figure 2 shows the feature importance based on SHAP scoring. It indicates that the highest contributor to an observation being marked as fraudulent was the merchant risk score. It is reasonable that merchant transactions marked with a higher risk score were more likely to be marked as fraudulent. It is notable that both customer transaction velocity and customer transaction amount contributed to the score, as it suggests that a customer transaction being unusually large (compared with their transaction history) would signify a risk. It is reasonable that an abnormal change in customer behavior would be revealed through distance as well as the number of failed attempts.

With the SHAP ranking, common sense in fraudulent transaction marking is also depicted. Merchant risk, a burst of transaction activity, an unusually large transaction amount, the distance of a transaction relative to previous ones, and failed attempts of a transaction are all reasonable indicators of fraud. This strengthens the model's interpretability because the characteristics that SHAP scoring identifies as fraud correlate with expected fraud and are easy for a fraud analyst to pinpoint.

Local SHAP Explanation

A specific suspicious transaction was analyzed with an SHAP local explanation. Some features of the transaction that driven the fraud risk higher were the high merchant risk score, high transaction amount, greater distance between transaction locations, and multiple failed attempts with subsequent successful attempts to authorize the transaction. Conversely, the longer time the account was open lowered the fraud risk. The local explanation also provides evidence that SHAP can support the fraud analyst in assessing why a transaction would be interpreted by the model as suspicious.

Discussion of Main Findings

The findings illustrate that utilizing an ensemble of classifiers can yield better outcomes in the detection of fraudulent acts than using each classifier as an individual entity. The voting system secured the highest scores in metrics of recall, F1-score, and AUC-ROC. These results reveal that aggregating probabilities of different tree-based classifiers adds the stability of imbalanced fraud detection scenarios.

The difference in performance can be attributed to the contribution of the individual classifiers. The Random Forest classifier constructs several independent decision trees, thus ensuring robust predictions. On the other hand, XGBoost and LightGBM use boosting techniques that emphasize the correction of prior errors in the predictions. By combining the classifiers probability output, the soft voting ensemble creates a more balanced fraud probability score and lowers the reliance on a single classifier.

The SHAP results reveal that adding explainability to the models used in fraud detection proves beneficial in modeling practice. The framework pinpointed several key fraud indicators. Among these were the risk score of the merchant, the velocity of the customer, the amount of the transaction, the distance involved, and the total of the failed attempts. These key fraud indicators have both a statistical and an operational significance. Thus, the framework can serve an automated classification system as well as a system to support the analysis done by humans.

Practical Implications

The framework can also assist financial institutions by scoring transaction risks, ranking the severity of suspicious cases, and providing SHAP explanations for fraud alerts. These can nudge analysts to prioritize certain transaction features and improve the traceability of the model decisions by explaining the rationale for the model's outputs.

The framework supports model governance as SHAP explanations show a clear rationale for model outcomes. Institutions may assess whether the model operates in line with their own risk appetite. If a model outputs many alerts due to the same features, for example, analysts may need to review this behavior in the context of the existing fraud policies and adjust the policies.

Consideration must be given to many factors for practical implementation. Processing speed and latency need to be low for fraud detection and SHAP explanations may compromise this. In these instances, the SHAP may be reserved for only the most extreme cases of alerted or suspicious transactions. Additionally, the fraud model may need frequent review to ensure coverage of shifting fraud behaviors.

Limitations, Reproducibility, and Ethics

This study has a few limitations. The first is that the study is based on a synthetic dataset. This dataset may be inadequate for replicating the actual complexities of institutional transactional data (for example, network datasets including data on transactions made by different merchants at different times, behavior patterns at an identity level, and networks of coordinated activities). The second limitation is that batch classification was used as opposed to streaming data in real-time. The third limitation is that the focus was primarily on tree-based ensemble learning, and no comparisons were made with graph-based, deep learning, or hybrid rule-machine learning.

Here, a synthetically generated dataset was used, and to support reproducibility, the construction of the dataset including its feature structure and class distribution, and the evaluation setup were described. Because there is not any real customer data here, the study does not disclose any private financial or personally identifiable data. A description of a SHAP value should be considered an explanation of the model and not of causation. The SHAP model should be used as a fraud mitigation support tool with human judgment in the final decision, and should not be used as a completely automated fraud decision model.

Conclusion

By combining ensemble learning with SHAP-based feature interpretation, an explainable fraud detection framework has been created. Different models were evaluated on a controlled synthetic financial transaction dataset, including Random Forest, XGBoost, LightGBM, and a soft-voting ensemble model. The ensemble model had the best performance with an accuracy of 0.991, a precision of 0.729, a recall of 0.875, an F1 score of 0.795, and an AUC-ROC of 0.973.

It can be concluded from the results that the use of ensemble learning has the potential to improve fraud detection from transaction data that is heavily skewed toward normal transactions. The voting ensemble identified 105 out of 120 fraud instances and produced the lowest number of false negatives amongst all the models that were evaluated. This clearly illustrates that the use of multiple classifiers can improve the detection of fraud instances and the control of false alerts.

The use of SHAP for model interpretation increased the confidence of end users in being able to explain the prediction of fraud and provided interpretable results. The main predictors of fraud were attributed to the risk score of the merchant, the customer velocity, the transaction amount, the distance of the location, and the number of failed attempts. The local interpretation of SHAP also illustrated how individual predictions could be supported by specific features.

The research demonstrated that explainable ensemble learning has the potential to provide accurate and transparent fraud analytics. The design of the framework has the potential to assist financial services in the monitoring of transactions, fraud, and the management of risk. Further development of the framework with real financial transaction data is warranted to assess model drift over time and expand the framework to real-time fraud detection.

Data Availability Statement
The dataset used in this study was synthetically generated for experimental validation. The feature structure, class distribution, and evaluation setting are described in the methodology section. The data generation script and model evaluation code may be made available by the corresponding author upon reasonable request.

Conflict of Interest Statement
The authors declare no conflict of interest.

Ethical Statement
This study used a controlled synthetic transaction dataset and did not involve direct human participants or personally identifiable financial information.

Acknowledgements

The researcher did not receive any funding for this study, and the results have not been published in any other sources.

References

Al-dahasi, E. M., Alsheikh, R. K., Khan, F. A.,&Jeon, G. (2025). Optimizing fraud detection in financial transactions with machine learning and imbalance mitigation. Expert Systems, 42(2), Article e13682. https://doi.org/10.1111/exsy.13682

Ali, A., Abd Razak, S., Othman, S. H., Eisa, T. A. E., Al-Dhaqm, A., Nasser, M., Elhassan, T., Elshafie, H.,&Saif, A. (2022). Financial fraud detection based on machine learning: A systematic literature review. Applied Sciences, 12(19), Article 9637. https://doi.org/10.3390/app12199637

Ashfaq, S.,&Chowdhury, T. K. (2023). Explainable artificial intelligence (XAI) approaches for cyber risk assessment in financial services. American Journal of Interdisciplinary Studies, 4(3), 96–135. https://doi.org/10.63125/3gjcb322

Chen, J., Liu, J., Liang, Y.,&Zhou, M. (2026). KE-MLLM: A knowledge-enhanced multi-sensor learning framework for explainable fake review detection. Applied Sciences, 16(6), Article 2909. https://doi.org/10.3390/app16062909

Hashemi, S. K., Mirtaheri, S. L.,&Greco, S. (2023). Fraud detection in banking data by machine learning techniques. IEEE Access, 11, 3034–3043. https://doi.org/10.1109/ACCESS.2022.3232287

Khalid, A. R., Owoh, N., Uthmani, O., Ashawa, M., Osamor, J.,&Adejoh, J. (2024). Enhancing credit card fraud detection: An ensemble machine learning approach. Big Data and Cognitive Computing, 8(1), Article 6. https://doi.org/10.3390/bdcc8010006

Leong, W. Y., Leong, Y. Z., & Leong, W. S. (2026). Artificial intelligence-driven fraud detection: Enhancing security in digital age. In T.-H. Meen, C.-F. Yang, & C.-Y. Chang (Eds.), *Proceedings of the 7th International Conference on Knowledge Innovation and Invention, Volume 1* (pp. 65–78). Springer. https://doi.org/10.1007/978-981-95-2113-5_7

Liu, S. (2026). AI and machine learning for e-commerce risk management. Springer Singapore. https://doi.org/10.1007/978-981-95-6184-1

Mathew, D. E., Ebem, D. U., Ikegwu, A. C., Ukeoma, P. E.,&Dibiaezue, N. F. (2025). Recent emerging techniques in explainable artificial intelligence to enhance the interpretable and understanding of AI models for human. Neural Processing Letters, 57(1), Article 16. https://doi.org/10.1007/s11063-025-11732-2

Nahar, J., Hossain, M. S., Rahman, M. M.,&Hossain, M. A. (2024). Advanced predictive analytics for comprehensive risk assessment in financial markets: Strategic applications and sector-wide implications. Global Mainstream Journal of Business, Economics, Development&Project Management, 3(4), 39–53. https://doi.org/10.62304/jbedpm.v3i4.148

Xie, Y., Liu, G., Yan, C., Jiang, C., Zhou, M.,&Li, M. (2024). Learning transactional behavioral representations for credit card fraud detection. IEEE Transactions on Neural Networks and Learning Systems, 35(4), 5735–5748. https://doi.org/10.1109/TNNLS.2022.3208967

Zafar, U.,&Wu, F. (2026). Methodological challenges in explainable AI for fraud detection: A systematic literature review. Artificial Intelligence Review, 59(4), Article 115. https://doi.org/10.1007/s10462-026-11516-7