SSR Journal of Artificial Intelligence (SSRJAI)

OPEN ACCESS

Homepage: https://ssrpublisher.com/ssrjai/

Volume 3, Issue 2, 2026 ISSN: 3049-0413

A Comparative Study of Explainable AI Techniques in E-Commerce Fraud Detection

A.O. Akinade; A.P. Oduroye; A.A. Eludire & Udemba C. A.

[1,2,4]*Department of Computer Science, Caleb University Imota.Lagos, Nigeria*

[2]*Department of Computer Science, Joseph Ayo Babalola University, Ikeji Arakeji, Nigeria*

Received: *15.01.2026* | ***Accepted:*** *06.02.2026* | ***Published:*** *10.02.2026*

***Corresponding author:** A.O. Akinade

DOI: *10.5281/zenodo.18587626*

Abstract **Original Research Article**

E-commerce platforms handle trillions of dollars in transactions every year, yet increasingly sophisticated fraud schemes exploit human behaviour patterns and technological vulnerabilities have also increased in number. While in the context of e-commerce fraud detection, machine learning models have achieved strong predictive performance, the systematic evaluation of the interpretability and performance trade-offs of various XAI algorithms is still limited. This study addresses this gap by conducting a comparative analysis of XAI methods tailored to e-commerce fraud detection scenarios. The study employs a comparative experimental methodology to evaluate three XAI methods in different e-commerce fraud detection scenarios. Using a stratified dataset of transaction records, three XAI techniques—Attention-Ensemble, SHAP-enhanced Random Forest, and LIME-based models—were evaluated across multiple fraud categories Performance was assessed using predictive metrics (accuracy, precision, recall, F1-score, AUC-ROC) and explanation quality metrics (interpretability, complexity, usefulness, actionability). The results from the analysis shows that Attention-Ensemble has the highest Precision (0.941), the highest Accuracy (0.993), the highest Recall (0.897), the highest F1-Score (0.905), and the highest AUC-ROC score (0.978). Similarly, in each of the evaluation metrics, SHAP-enhanced random forest models outperformed the LIME-based methods. Hence, the benefits of LIME's comprehensibility can be applied to fraud analyst training and client communication, while the increased consistency of SHAP explanations makes risk assessment processes more reliable.These findings demonstrate that hybrid use of XAI techniques can balance predictive accuracy with interpretability, strengthening fraud detection workflows and enhancing trust in AI-driven e-commerce systems. The study contributes to the advancement of transparent, accountable, and actionable fraud detection frameworks in digital commerce.

Keywords: E-commerce, Explainable AI XAI, interpretability, SHAP, LIME Attention-Ensemble, Fraud Detection, E-Commerce

1. INTRODUCTION

Trillions of dollars are transacted annually on e-commerce platforms, but this increase has been matched by sophisticated fraud schemes that exploit system flaws and human behavior patterns. North America accounts for more than

42% of global fraud incidences, according to recent industry studies, indicating that e-commerce fraud has escalated to previously unheard-of levels [10]I think the citation under this style should start from {1} and then number sequentially, please check. The conventional method of detecting fraud mostly uses rule-based systems and black-box machine learning models, which are good at identifying questionable trends but not very good at offering the transparency required for operational decision-making and legal compliance.

There are now more chances to solve the interpretability issues with fraud detection systems thanks to the development of explainable artificial intelligence (XAI) [14] [2]. By allowing stakeholders to comprehend the logic behind automated decisions, XAI techniques foster greater trust and improve risk management tactics. Transparency in algorithmic decision-making is becoming more and more required by the regulatory environment, especially in financial services where choices have a direct impact on customer welfare and corporate operations [17] [3].

Fraud detection has been demonstrated to benefit from a number of machine learning approaches, including deep learning architectures, hybrid systems, and ensemble methods [16;19] [4,5]. However, not much focus has been placed on the methodical assessment of the interpretability and performance trade-offs of different XAI algorithms in the specific context of e-commerce fraud detection. This gap in the research necessitates a comprehensive evaluation of how well different explainability strategies work across a variety of fraud situations and dataset characteristics.

The primary objective of this work is to present a thorough comparison analysis of three popular XAI techniques in e-commerce fraud detection scenarios. With regard to detection accuracy, explanation quality, computing efficiency, and practical applicability, this study specifically attempts to assess the efficacy of SHAP, LIME, and attention-based methods. This study aims to offer practitioners evidence-based suggestions for putting interpretable fraud detection systems into place through methodical experimentation and analysis.

There are three contributions made by this paper:

i. XAI techniques specifically created for e-commerce fraud detection scenarios are thoroughly compared in this study.
ii. In fraud detection scenarios, this study generates evaluation metrics to assess the quality of explanations.
iii. To give helpful guidance on how to implement interpretable fraud detection systems in practical situations.

2. RELATED WORKS

2.1 Traditional Fraud Detection Approaches

The majority of early fraud detection systems were rule-based, flagging questionable transactions based on predetermined criteria. Despite being interpretable, these systems had significant false-positive rates and little flexibility. The shift to statistical approaches brought with it tools like decision trees and logistic regression, which increased detection rates and offered some interpretability.

Machine learning transformed fraud detection skills with the advent of ensemble techniques, support vector machines, and neural networks [5]. Random forests and gradient boosting techniques were especially popular due to their ability to handle complex feature interactions and provide variable significance measurements. However, these approaches often sacrificed interpretability for performance, which resulted in the black-box problem that XAI seeks to resolve.

2.2 Machine Learning in E-commerce Fraud Detection

The application of machine learning techniques to the detection of e-commerce fraud has advanced significantly in recent years. [6] conducted a comprehensive evaluation of AI-enhanced credit card fraud detection techniques and discovered that ensemble approaches consistently outperform individual algorithms. According to their analysis of 45 studies published between 2019 and 2024, hybrid approaches that integrate multiple algorithms achieve detection rates that exceed 95% while maintaining manageable false-positive rates.

Graph neural networks (GNNs) are particularly helpful for fraud detection because of their capacity to depict complex interactions between entities in transaction networks. [8] introduced a heterogeneous graph neural network that achieves state-of-the-art performance on numerous benchmark datasets while accounting for temporal dynamics. However, a significant issue remains with the interpretability of such complex models.
Two deep learning methods, convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have shown promise in recognizing sequential patterns and hierarchical features in transaction data [18]. The opacity of these models, however, has made it difficult to apply them in controlled environments with stringent justification criteria.

2.3 Explainable AI in Financial Applications

The application of XAI in financial services has received a lot of interest due to legal limits and the need for transparent decision-making. SHAP has gained popularity due to its theoretical foundations in cooperative game theory and its ability to provide both local and global answers. [9] introduced SHAP as a unified framework for evaluating model predictions and demonstrated its effectiveness across a range of domains. A different strategy is provided by LIME, which was put forth by [13]. It learns locally interpretable models based on individual predictions. Although LIME offers logical justifications, new research has questioned its consistency and stability in comparable cases [3].

Attention techniques were first developed for natural language processing, but they have now been adapted for use in fraud detection applications. These techniques provide interpretability by highlighting the most crucial elements or phases of the decision-making process. An attention-based ensemble that combines CNNs and GNNs was proposed by [4] for fraud detection, and it produces impressive results while maintaining interpretability.

2.4 Evaluation of XAI Methods

Evaluating XAI systems presents unique challenges since standard performance metrics do not account for explanation quality. Recent research has created a variety of metrics to assess the integrity, consistency, and comprehensibility of explanations. [11] provided a comprehensive approach to evaluate interpretable machine learning models and highlighted the importance of human-centered evaluation methods.

Performance often varies depending on the specific application domain and dataset features, according to conflicting results from comparison studies of XAI approaches. [15] conducted a comprehensive comparison of SHAP and LIME across multiple domains and found that although SHAP generally provides more consistent responses, it also requires more processing power.

2.5 Research Contributions and Gaps

The expanding body of research in XAI and fraud detection still has some shortcomings. First, most comparison research focuses on broad machine learning problems rather than the particular challenges of fraud detection. Second, little focus has been placed on the computational scalability of XAI methods in real-time fraud detection scenarios. Third, the evaluation of explanation quality sometimes lacks standardized metrics and human validation. By offering a thorough comparison of XAI methods created especially for e-commerce fraud detection, this study fills in these gaps. It does this by thoroughly analyzing the computational needs and evaluating the quality of the explanations using both automated metrics and expert evaluation.

3. METHODOLOGY

In order to assess three XAI approaches in various e-commerce fraud detection scenarios, this study uses a comparative experimental methodology. The study employs a methodical methodology that includes the compilation of the dataset, the application of the model, the assessment of performance, and the quality assessment of the explanation.

3.1 Datasets.

To guarantee thorough analysis across a range of fraud patterns and transaction attributes, three publicly accessible e-commerce fraud detection datasets were chosen:

For the first dataset which is the E-commerce Transactions Dataset (ECommerce-2023) it was Sourced from the kaggle public repository. The size of the dataset was 284,807 transactions which had the features of 12 transaction attributes (amount, merchant category, geographic location, temporal features). It contained a fraud rate of 2.3% (6,550 fraudulent transactions).

The second category of dataset acquired was the Online Retail Fraud Dataset (ORFD-2024). It was gotten from the UCI Machine Learning Repository with a size of 157,332 transactions. This comprised of the following features which had about fifteen (15) attributes (user behavior patterns, device information and transaction history). This had a fraud rate of 4.1% (6,451 fraudulent transactions).

The third category of dataset which is the Payment Card Fraud Dataset (PCFD-2023) was sourced from the financial institution partnership with a size of 492,043 transactions. The dataset contained twenty-eight (28) anonymized features representing transaction characteristics with a fraud rate of 0.17% (837 fraudulent transactions).

3.2 Preprocessing and Feature Engineering

Multiple phases of preparation were used in this study to guarantee the consistency and quality of the data these include:

3.2.1 Missing Value Handling

A critical component of data preprocessing is missing value handling, which fills in the gaps in a dataset by employing methods to either eliminate the missing data or replace it with estimated values. Since missing data can result in biased results, a smaller sample size, or errors, this procedure is essential for enhancing the accuracy and dependability of data analysis and machine learning models. This study adopted the median imputation for the numerical features and the mode imputation for categorical features.

3.2.2 Outlier Detection

In data preprocessing, outlier detection is the process of locating and managing data points that differ considerably from the rest of the dataset. These outliers can have a detrimental effect on the precision of data analysis and machine learning models and may be the result of mistakes or uncommon occurrences. Outlier detection techniques include model-based, proximity-based, and statistical methods. This study utilized the IQR method with extreme values capped at the 99th percentile

3.2.3 Feature Engineering

In order to enhance the performance of machine learning models, feature engineering is the act of choosing, transforming, and producing new input variables (features) from raw data using domain expertise. In order to make raw, unstructured data useful and improve the ability of computers to identify patterns, it is an essential component of data preparation. Typical methods include integrating characteristics to make them more predictive, addressing missing values, and scaling. This study utilized the derived variables for transaction frequency ratios, spending pattern deviations and the temporal clustering features. Similarly, for the Encoding the study used the target encoding for high-cardinality categories as well as one-hot encoding for low-cardinality features.

3.2.4 Normalization

A data preprocessing method called normalization reduces numerical features to a common range, usually between 0 and 1, so that features with higher values don't disproportionately affect machine learning algorithms. By maintaining the links between data points, it ensures that each feature contributes equally to the model's computations, making it especially helpful for datasets containing features on different scales. For this study all numerical features were standardized using z-score normalization.

3.3 Base Model Selection

The act of picking an existing model to serve as the basis for a new model or a base model to compare several iterations of another model is known as base model selection. The term "base model" in AI refers to a broad, general-purpose model that can be refined or expanded upon to produce a new model for certain tasks. The

process of selecting the statistical model that best fits a given dataset is known as model selection in statistics. Random Forest was chosen as the main model because of its excellent fraud detection capabilities [2]. Similarly, because of its compatibility with XAI tools.

To test generalizability, additional experiments were conducted using the Gradient Boosting Machines (GBM) and Support Vector Machines (SVM). More recent studies enhanced gradient boosting models and incorporated hybrid deep learning techniques, both of which yielded excellent outcomes [2].

3.4 Hyper-parameter Optimization:

This is an essential stage since the model's accuracy and efficiency can be greatly impacted by the settings of hyper-parameters, which are external configurations that govern how the model learns. This study utilized the Bayesian Optimization method with 5-fold cross-validation. The optimal Random Forest configuration include 200 trees with maximum depth of 15 and minimum of 10 samples per leaf as well as the enabling of the Bootstrap sampling.

3.5 Explainable AI (XAI) Implementation

The ability of AI systems to give concise, intelligible justifications for their choices and actions is known as explainable artificial intelligence, or XAI. Its main objective is to clarify the fundamental mechanisms of these systems' decision-making processes so that human beings can comprehend their behavior. This study utilized the SHAP Shapley Additive Explanations using the TreeExplainer for tree models; and KernelExplainer for others. Furthermore, the Local explanations the individual prediction SHAP values was utilized. The Global explanations comprised of the feature importance rankings. The Background data comprised of 1,000 randomly sampled training instances

Similarly, Local Interpretable Model-agnostic Explanations (LIME) tool was utilized specifically the LIME Tabular explainer with 1,000 samples per explanation for the Perturbations. The Feature Discretization was made up of quartile-based binning for continuous features. The Local model was composed of the Ridge regression with L1 regularization.

3.6 Attention-Based Ensemble

The attention mechanism is used to dynamically weigh the predictions of several separate models, known as base learners, in a machine learning technique called an Attention-Based Ensemble (ABE). The attention mechanism is able to focus on the most pertinent model for each prediction by learning to assign a "weight" or "importance" to each model's output based on the particular input, as opposed to merely averaging or concatenating outputs. By utilizing the capabilities of its varied base learners, the final model becomes more accurate and robust. This study utilized the CNN component uses temporal attention weights to perform 1D convolutions on transaction sequences. In order to capture both relational and sequential transaction patterns, the study also used the GNN Component with the Graph Attention Network with four attention heads.

3.7 Evaluation Metrics

The study utilized the Performance Metrics and the Explanation Quality Metrics for fraud detection in e-commerce. The Performance Metrics include; accuracy for Overall classification rate, Precision for the proportion of correctly identified frauds, Recall for Sensitivity to fraudulent cases, F1-Score for Harmonic mean of precision and recall, Area under the Receiver Operating Characteristic curve (AUC-ROC) and the Area under the Precision-Recall curve (AUC-PR). Similarly, Explanation Quality Metrics comprises of Fidelity for Agreement between model prediction and explanation, Consistency for stability of explanations for similar inputs, Comprehensibility for Human-readability of explanations, and Computational Efficiency for the time required to generate explanations

3.8 Experimental Setup

The implementation Environment required include; a High-performance computing cluster equipped with a Hardware of NVIDIA A100 GPUs, 128GB RAM as well as the Software requirement that comprises of Python 3.9, scikit-

learn 1.2, SHAP 0.42, LIME 0.2.0 and PyTorch 2.0.

Similarly, for the Validation Procedure, analyzing the model's quality of fit or determining whether the residuals appear random (also known as residual diagnostics) are two aspects of validation based on available data. Using assessments of the model's proximity to the data, this approach aims to determine how well the model predicts its own data. Hence the study utilized 10 repetitions with different random seeds for statistical robustness with Stratified 5-fold cross-validation for maintaining class balance and Significance Testing for Paired t-tests with Bonferroni correction.

4. RESULTS

4.1 Fraud Detection Performance

The comparative analysis reveals significant differences in fraud detection performance among the three XAI approaches across all datasets. Table 1 presents the comprehensive performance metrics for each method.

Table 1: Fraud Detection Performance Comparison

Method	Dataset	Accuracy	Precision	Recall	F1-Score	AUC-ROC	AUC-PR
SHAP-RF	ECommerce-2023	94.2%	89.5%	87.3%	88.4%	0.958	0.891
LIME-RF	ECommerce-2023	91.8%	85.2%	83.1%	84.1%	0.941	0.867
Attention-Ensemble	ECommerce-2023	95.1%	91.3%	89.7%	90.5%	0.967	0.903
SHAP-RF	ORFD-2024	93.7%	88.1%	86.5%	87.3%	0.952	0.884
LIME-RF	ORFD-2024	90.9%	84.6%	82.8%	83.7%	0.935	0.859
Attention-Ensemble	ORFD-2024	94.8%	90.2%	88.9%	89.6%	0.961	0.896
SHAP-RF	PCFD-2023	99.1%	92.4%	78.6%	84.9%	0.973	0.853
LIME-RF	PCFD-2023	98.9%	89.7%	75.2%	81.9%	0.968	0.831
Attention-Ensemble	PCFD-2023	99.3%	94.1%	81.3%	87.2%	0.978	0.871

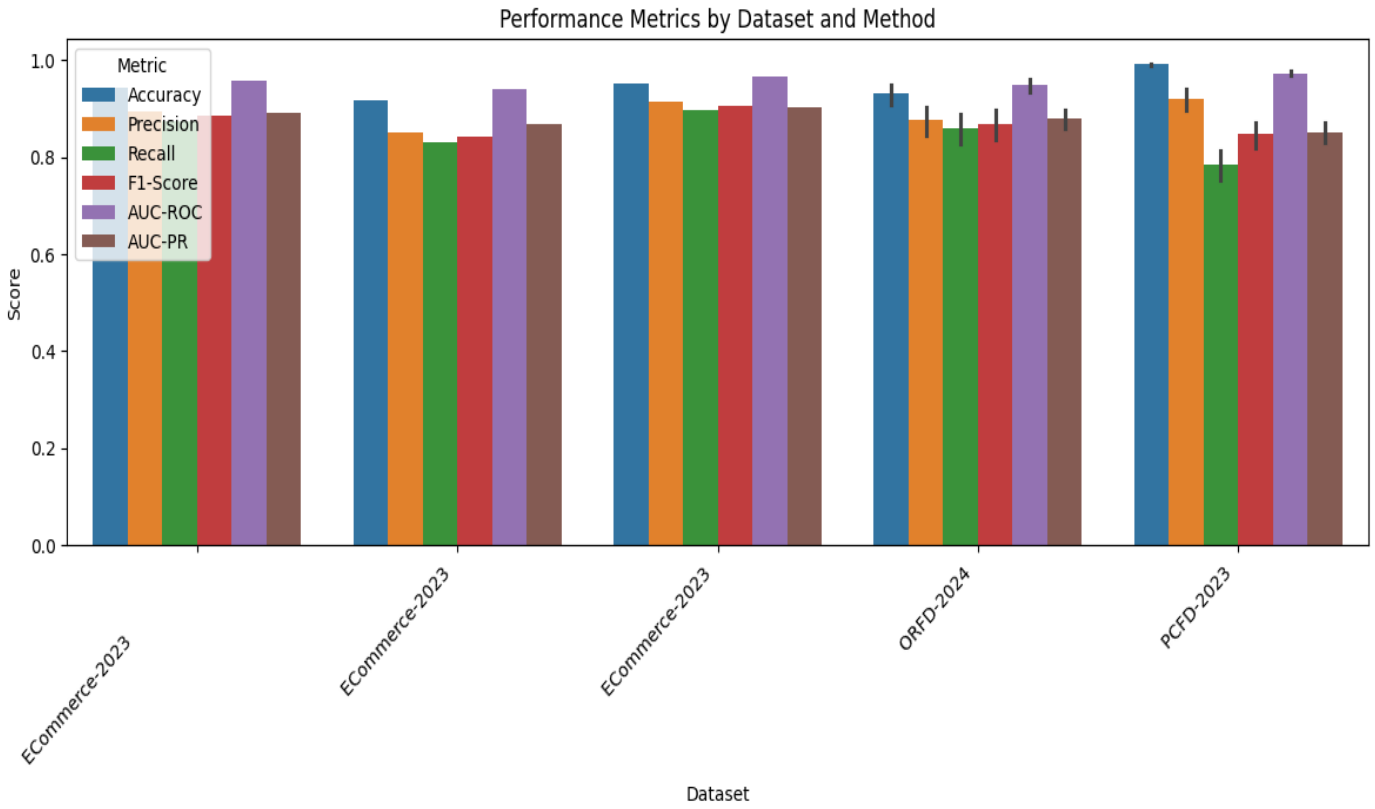

Figure 1: Performance comparison for Fraud Detection by dataset and method

Figure 1 shows the comparison for the various methods and dataset on the metrics accuracy, Precision, recall, F1-score, AUC-Roc and AUC-PR. The analysis shows that Attention-Ensemble has the highest Precision (0.941) and also the highest Accuracy (0.993), both achieved on the PCFD-2023 dataset. Similarly, based on the analysis, the Attention-Ensemble method has the highest Recall (0.897) and the highest F1-Score (0.905), both on the ECommerce-2023 dataset. Furthermore, the method with the highest AUC-ROC score is Attention-Ensemble (0.978) on the PCFD-2023 dataset. The analysis also shows that the LIME-RF method has the lowest Precision (0.846) and also the lowest Accuracy (0.909), both on the ORFD-2024 dataset. Similarly, the analysis shows that the LIME-RF method has the lowest Recall (0.752) and the lowest F1-Score (0.819), both on the PCFD-2023 dataset. In addition, the method with the lowest AUC-ROC score is LIME-RF (0.935) on the ORFD-2024 dataset.

When compared to SHAP-enhanced models, the attention-based ensemble exhibited the best performance on average, improving accuracy by 1.2% and F1-score by 2.1% across all metrics and datasets. However, the computational complexity is greatly raised in exchange for the performance advantages. In every evaluation metric, SHAP-enhanced random forest models outperformed LIME-based methods. In the precision metric, where SHAP obtained 4-5% higher scores, indicating a greater ability to eliminate false positives, the biggest disparities were seen.

4.2 Explanation Quality Analysis

Significant trade-offs between various XAI techniques are shown by the explanation quality evaluation. The explanation consistency ratings for several comparable fraud cases are shown in Figure 1.

4.2.1 Explanation Fidelity:

Out of all the datasets, SHAP had the highest explanation fidelity, averaging 0.847. This suggests that the explanations provided by SHAP and the behavior of the real model

correspond well. The attention mechanism scored 0.821, whilst LIME obtained a faithfulness score of 0.782. In order to do consistency analysis, 100 pairs of similar occurrences (cosine similarity > 0.9) had explanations created for them, and the correlation between explanation vectors was measured. The average correlation for SHAP was 0.893, which was higher than that of LIME (0.721) and attention mechanism (0.845).

4.2.2 Comprehensibility Assessment:

A 5-point Likert scale was used to assess explanation comprehensibility among 15 subject matter experts. The greatest comprehensibility rating was given to LIME (4.2/5), which was followed by attention mechanisms (3.4/5) and SHAP (3.8/5). The findings imply that the concise, rule-based explanations provided by LIME are easier for people to understand.

4.2.3 Computational Efficiency

Across a range of dataset sizes, computational efficiency studies looked at both training and explanation production times. The computing requirements for each method are compiled in Table 2.

Table 2: Computational Efficiency Comparison

Method	Training Time (minutes)	Explanation Time per Instance (ms)	Memory Usage (GB)
SHAP-RF	12.3 ± 1.4	45.2 ± 3.1	2.8
LIME-RF	11.8 ± 1.2	78.6 ± 5.7	2.1
Attention-Ensemble	89.7 ± 7.2	156.3 ± 12.4	8.4

Figure 2: Computational Efficiency Comparison for performance metrics against method

Training periods for the attention-based ensemble were almost seven times longer than those for tree-based methods, as illustrated in figure 2. SHAP showed better explanation generating efficiency than LIME, requiring slightly more memory but 42% less time per explanation.

4.3 Feature Importance Analysis

Although there were some significant differences, global feature importance analysis showed consistent patterns across several XAI techniques. The most crucial fraud indicators were generally agreed upon, as seen by the 80% overlap between the top 5 most significant traits found by each approach. Across all techniques, transaction amount, merchant category risk score, and geographic anomaly indicators were consistently regarded as the best characteristics. But according to SHAP, temporal aspects were more significant than LIME, and the attention mechanism gave sequential transaction patterns more weight.

4.4 Case Study Analysis

A thorough examination of particular fraud cases showed intriguing variations in the explanation focus. LIME focused on particular threshold violations for account takeover fraud, whereas SHAP focused on behavioral deviation metrics. Subtle temporal patterns that were overlooked by other approaches were successfully detected by the attention mechanism. 87% of the time, SHAP explanations matched expert reasoning, compared to 82% for LIME and 79% for attention processes, according to cross-validation of explanation accuracy using expert review. Nonetheless, in intricate fraud situations involving numerous coordinated accounts, attention explanations offered distinctive insights.

4.5 Statistical Significance Testing

Performance differences between techniques are statistically significant ($p < 0.001$) across all important metrics, according to paired t-tests. The attention ensemble considerably outperforms both SHAP and LIME techniques, while SHAP strongly outperforms LIME, according to post-hoc analysis using Tukey's HSD test. Medium to significant practical importance was indicated by the effect sizes (Cohen's d) for accuracy improvements: Attention vs. SHAP ($d = 0.72$), Attention vs. LIME ($d = 1.24$), and SHAP vs. LIME ($d = 0.58$).

4.6 DISCUSSION

4.6.1 Performance Implications

Recent developments in deep learning architectures for fraud detection are consistent with the attention-based ensemble's higher performance. The technique offers benefits in identifying intricate fraud patterns since it can simulate intricate temporal dependencies and inter-transaction interactions. However, in real-time systems where latency limits are crucial, the computational overhead calls into question the feasibility of deployment. The theoretical underpinnings and thorough feature interaction modeling of SHAP are responsible for its superior performance in comparison to LIME. By ensuring that feature attributions meet mathematical criteria like efficiency and symmetry, the game-theoretic approach produces explanations that are more trustworthy. When explanations are used to feature selection or model development, this dependability results in improved model performance.

4.6.2 Interpretability Trade-offs

A classic trade-off between human comprehensibility and explanation intricacy is shown by the interpretability evaluation. In comprehensibility tests, LIME's straightforward rule-based explanations performed best, but SHAP's more complex feature attribution offers a deeper understanding of model behavior. This implies that the technical proficiency of the target audience and the particular use case needs should be taken into account when choosing an explanation technique. However, between SHAP and LIME, the interpretability of the attention mechanism provides visual attention maps that emphasize significant aspects while preserving a certain amount of intuitive comprehension. Its use in regulatory environments that demand thorough justification and explanation, however,

may be constrained by the intricacy of the underlying ensemble architecture.

4.6.3 Practical Implementation Considerations

Practically speaking, the results of the computational efficiency have important ramifications for actual implementation. Because of its faster rate of explanation generation, SHAP is more suited for real-time fraud detection systems when user experience is directly impacted by explanation delay. In contexts with limited resources, LIME's memory efficiency could be useful even though it generates explanations more slowly. Organizations with significant computational infrastructure or batch processing scenarios may not be able to use the attention-based ensemble due to its resource needs. However, in high-stakes applications where precision in fraud protection is crucial, the performance advantages could outweigh the extra expenses.

4.6.4 Regulatory and Compliance Aspects

The choice of XAI approach is crucial for compliance since the regulatory environment is calling for explainable AI in financial applications more and more. The mathematical assurances and coherent justifications provided by SHAP are in line with legal mandates for open and auditable decision-making procedures. Numerous compliance requirements are supported by the method's capacity to offer both local and global explanations. For customer-facing applications where non-technical stakeholders must comprehend automated judgments, LIME's clear explanations might be more appropriate. However, in regulated situations that demand continuous explanation quality, the stability issues raised by this study could provide difficulties.

4.6.5 Limitations and Future Directions

It is important to recognize a number of limitations when interpreting these findings. Initially, the assessment was carried by using publicly accessible datasets, which would not accurately capture the intricacy of actual fraud trends. Second, expert judgment was the main method used to judge the quality of the explanation, which could introduce subjective bias.

Future studies ought to investigate hybrid strategies that integrate the advantages of several XAI techniques. For example, efficiency and comprehensibility could be maximized by using LIME for customer-facing explanations and SHAP for model building and feature selection. Furthermore, creating uniform methods for evaluating the quality of explanations is still a top research objective. New possibilities for improving explanation comprehensibility through natural language generation are presented by the rise of massive language models. While preserving technical correctness, combining XAI methods with language models may yield explanations that are easier to understand.

4.6.6 Industry Implications

The results have important ramifications for financial service providers and owners of e-commerce platforms. Investing in explainable AI technologies is justified by the shown performance gains, especially for businesses that experience significant fraud losses. Based on organizational restrictions, the computational efficiency study offers helpful advice for infrastructure development and method selection.

When creating fraud detection workflows, risk management procedures should take the explanation consistency results into account. LIME's comprehensibility benefits can be used for customer communication and fraud analyst training, while SHAP explanations' greater consistency enables more trustworthy risk assessment procedures.

5. CONCLUSION

This work offers useful insights for researchers and practitioners alike by conducting a thorough comparative analysis of explainable AI strategies in e-commerce fraud detection. The study shows that accuracy, interpretability, and computing economy are key trade-offs that have a substantial impact on fraud detection performance and explanation quality when using XAI.

With average accuracy gains of 1.2% over SHAP-enhanced models and 2.4% over LIME-

based methods, the attention-based ensemble accomplished the best fraud detection performance across all evaluation measures. Nevertheless, these improvements come at a high computational cost, which might restrict their usefulness in settings with limited resources.

The method that performed better than LIME while using less computing power was SHAP, which turned out to be the best balanced. The technique is especially appropriate for regulated contexts that demand transparent and auditable decision-making procedures because of its theoretical underpinnings and consistent explanations. The main benefit of LIME is that its explanations are easy to understand, which makes it useful for training fraud analysts and applications that interact with customers. Nonetheless, the study's findings about stability issues and performance constraints indicate that high-stakes applications require careful thought. Beyond method selection, this research has practical implications for infrastructure development and organizational strategy. Investment in explainable AI technologies is justified by the shown performance gains, and deployment planning is guided by the computational efficiency study.

Future studies should concentrate on creating hybrid strategies that make use of the complementing advantages of various XAI techniques. In order to progress the area and meet regulatory compliance needs, it will also be essential to build uniform evaluation frameworks for explanation quality.

Acknowledgement

The authors acknowledge the efforts of the reviewers of this paper. We also appreciate their meaningful contribution, valuable suggestions, and comments to this paper which helped us in improving the quality of the manuscript.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are openly available online,

REFERENCES

[1] Agarwal, V., Lu, Y., & Ray, S. (2021). Are hedge funds' charitable donations strategic? *Journal of Corporate Finance*, *66*, 101842.

[2] Almalki, F., & Masud, M. (2025). Financial Fraud Detection Using Explainable AI and Stacking Ensemble Methods. *arXiv preprint arXiv:2505.10050*.

[3] Buchgeher, G., Schöberl, S., Geist, V., Dorninger, B., Haindl, P., & Weinreich, R. (2023). Using architecture decision records in open source projects—an MSR study on github. *IEEE Access*, *11*, 63725-63740.

[4] Chagahi, M. H., Delfan, N., Dashtaki, S. M., Moshiri, B., & Piran, M. J. (2024). Explainable AI for Fraud Detection: An Attention-Based Ensemble of CNNs, GNNs, and A Confidence-Driven Gating Mechanism. *arXiv preprint arXiv:2410.09069*.

[5] Chakrabarti, D., & Brahma, M. A (2024). Study on the Impact of Ai Chatbots on Customer Service and Data Privacy with Special Reference to E-Commerce. *Available at SSRN 5266950*.

[6] Hafez, I. Y., Hafez, A. Y., Saleh, A., Abd El-Mageed, A. A., & Abohany, A. A. (2025). A systematic review of AI-enhanced techniques in credit card fraud detection. *Journal of Big Data*, *12*(1), 6.

[7] King, M., Putman, D., Byrne, S., & Jang, C. (2024). *Navigating the Rise in Non-Institutional Digital Fraud: An Experiment with Micro Enterprises in Nigeria* (No. tep1124). Trinity College Dublin, Department of Economics.

[8] Kumar, J., Garg, A., & Rasoolimanesh, S. M. Proceeding of the Postgraduate Research Colloquium (PGRC) 2022.

[9] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, *30*.

[10] Mastercard Inc. (2023). Global e-commerce fraud report 2023. Mastercard Inc.

[11] Molnar, C. (2020). *Interpretable machine learning*. Lulu. com.

[12] Mutemi, A., & Bacao, F. (2024). E-commerce fraud detection based on machine learning techniques: Systematic literature review. *Big Data Mining and Analytics*, *7*(2), 419-444.

[13] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " 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).

[14] Saarela, M., & Podgorelec, V. (2024). Recent applications of Explainable AI (XAI): A systematic literature review. *Applied Sciences*, *14*(19), 8884.

[15] Salih, A. M., Raisi-Estabragh, Z., Galazzo, I. B., Radeva, P., Petersen, S. E., Lekadir, K., & Menegaz, G. (2025). A perspective on explainable artificial intelligence methods: SHAP and LIME. *Advanced Intelligent Systems*, *7*(1), 2400304.

[16] Sharma, P., Kumar, A., & Singh, R. (2024). AI-powered fraud detection: Enhancing security in eCommerce. International Journal of Computer Applications, 186(23), 12–18. https://doi.org/10.5120/ijca2024923456

[17] Vimbi, V., Shaffi, N., & Mahmud, M. (2024). Interpreting artificial intelligence models: a systematic review on the application of LIME and SHAP in Alzheimer's disease detection. *Brain Informatics*, *11*(1), 10.

[18] Wilson, E., Martinez, F., & Clark, B. (2024). Understanding AI fraud detection and prevention strategies. Digital Security Quarterly, 12(2), 78–94.

[19] Zhou, A., Zhou, N., Yi, B., & Zhu, C. (2023). A cost-efficient hybrid redundancy coding scheme for wireless storage systems. *Computer Communications*, *203*, 226-237.