SCIENTIFIC CULTURE
www.sci-cult.com

SCIENTIFIC CULTURE, Vol. 12, No. 4, (2026), pp. 3202-3210
Open Access. Online & Print

DOI: 10.5281/zenodo.12426383

EXPLAINABLE AI TECHNIQUES FOR PATTERN RECOGNITION AND DECISION TRANSPARENCY IN CRITICAL SYSTEMS

Areeba Raza[1]*, Azree Shahrel Ahmad Nazri[2]

[1]Faculty of Computer Science & Information Technology, University Putra Malaysia, Seri Kembangan, Malaysia.
[2]The Institute for Mathematical Research (INSPEM), University Putra Malaysia, Seri Kembangan, Malaysia.

Received: **13/11/2025**
Accepted: **23/02/2026**

Corresponding Author: Areeba Raza (areebah.rz@gmail.com)

ABSTRACT

This study examines how Explainable AI (XAI) tools can be incorporated into fraud detection models to increase the interpretation and transparency of decisions within financial systems. The research increases the interpretability of models by using a single methodology that integrates LIME, SHAP, Attention Mechanisms, and LRP without compromising the performance of models in terms of classification. Random Forest model had the best AUC (0.90), precision (0.85) and recall (0.88) than Neural Network (AUC = 0.85) and SVM (AUC = 0.80). The findings suggest that XAI tools, including SHAP and LIME, are useful in explaining the decisions of a model in a transparent manner, which is important to achieve trust and accountability in AI-based fraud detection systems. Challenges identified by the study also include compromising accuracy and interpretability, and reducing biases in training data.

KEYWORDS: Explainable AI, Fraud Detection, LIME, SHAP, Attention Mechanisms, LRP, Random Forest, Neural Network, SVM, Financial Systems, Trust, Transparency.

1 INTRODUCTION

Fraud detection in financial systems has increasingly become dependent on artificial intelligence (AI) models, particularly for analysing large amounts of data on transactions [1]. The increasing complexity of such AI models, it provides a great benefit in detecting fraudulent transactions which would otherwise be hard to detect. This capability is of special interest to the financial institutions because the cost of fraud is so huge, not only in monetary terms but also in reputational losses [2]. Nevertheless, there is one major problem that can be identified in spite of the increased usage of AI in fraud detection: there is no transparency in such models. The majority of developed AI models, including neural networks and ensemble models, are usually viewed as black boxes due to the lack of clear information about how they make decisions. The decision makers and stakeholders should be aware of the rationale of a prediction of a model to enable the system to be trusted, audited and liable. The lack of explainability makes it hard to determine whether the model is performing in a fair, accurate and ethical manner, particularly when the decisions made concern transactions of high value and risks to individuals or organisations.

This research is driven by the increasing need to have elucidated AI (XAI), particularly in the area of fraud detection. In the field of finance, transparency of decision-making is a vital requirement, and AI is employed to categorise the transactions as legitimate or fraudulent [3]. Such XAI methods as LIME (Local Interpretable Model-agnostic Explanations), SHAP (SHapley Additive exPlanations), and Layer-wise Relevance Propagation (LRP) have been developed to resolve this issue. All these methods allow users to gain knowledge and interpret the way AI models make their decisions, and provide a good insight into the significance of specific features in the process of decision-making [4]. Transparency is not only important to technical performance; it is essential in building trust with the users, regulators and stakeholders. It becomes easier to justify the results of AI models, detect issues, and meet regulatory requirements when their decisions can be read and tracked. Reliability with AI systems is especially needed in financial aspects, where such mistakes or misunderstandings may cause grave financial mistakes [5].

The research presented in the paper aimed to use the XAI techniques to enhance the pattern recognition in the fraud detection systems and, at the same time, increase the transparency of the decision. This research was conducted to compare the explanatory power of the LIME, SHAP, and LRP in terms of predicting models using the IEEE-CIS Fraud Detection Dataset, which is a dataset of labelled financial transaction data. The idea was to show that explainable AI not only enhanced the performance of the classifiers in fraud detection but also gave useful information on how every prediction was made. Classifiers such as the Random Forests, Support Vector machines (SVM) and Neural Networks were subjected to these methods to investigate their performance and interpretability. By so doing, the study sought to cover this gap between high-accuracy models of fraud detection and the desire to have transparent, understandable decision-making.

The novelty of the study is the integrated approach to XAI that the study presents by employing various XAI techniques with different machine learning models. In contrast to the past, where people have applied individual methods of explainability to particular models, the paper presents the efficacy of LIME, SHAP, and LRP, in concert, among the various classifiers, in the context of fraud detection [6]. The paper gives an extensive assessment of model accuracy, interpretability, and the efficiency of calculations, which presents a general picture of the implementation of such XAI methods into practice. Moreover, the study has an impact on the field as it offers a systematic method to assess explainability in such critical systems as fraud detection to emphasise the value of transparency and trust in AI-based decision systems. The paper provides important insights through this methodology on how to increase the adoption and credibility of AI in high-risk applications.

2 RELATED WORK

2.1 Explainable AI (XAI)

Integration of Explainable AI (XAI) into important applications, including fraud detection, has gained growing popularity because of the significance of model transparency and interpretability in high-stakes settings [7]. This part surveys the available literature about XAI methods, their application within the fraud detection schemes, and the limitations in enhancing the classification accuracy and the explanation of the decision made.

Explainable AI (XAI) is a collection of methods that can help to make the decisions of black-box AI models more understandable and comprehensible [8]. It aims to enable human beings to reason and have confidence in the actions of AI systems, particularly in cases where the outcomes of the decisions matter a lot. To meet this requirement, several XAI methods have been created, and each has

its advantages and disadvantages.

LIME (Local Interpretable Model-agnostic Explanations) is another widely used XAI methodology, which offers local interpretability by approximating the complex model by a more interpretable model in the neighbourhood of every prediction. The LIME operates by manipulating the input information and monitoring the changes in the predictions, and thus offers an insight into the factors that affect individual decisions [9]. The method finds application especially in classification, where one would wish to know certain decisions, e.g. fraud detection in financial systems.

SHAP (SHapley Additive exPlanations) is a model decision interpretation derived using Shapley values of cooperative game theory that produces a contribution score on each feature according to its influence on the model output. SHAP has become popular because of the mathematical base that makes it have a consistent and reliable attribution of features [10]. It is well applied to describe machine learning models (such as Random Forests and SVMs), and offers an international perspective on model behaviour, which is essential to explain anomaly detection in fraud systems.

Attention Mechanisms have been largely applied in deep learning models, particularly in neural networks, to indicate the sections of inputs that the model focuses on to make decisions [11]. The method is particularly applicable in problems such as image recognition or natural language processing, where there exist areas or words in the input data which may have a disproportionate impact on the output of the model. Learning-based attention mechanisms can be applied in fraud detection to find out important features in the transaction data, and it offers useful insights into how the model can differentiate between valid and fraudulent transactions [12].

Another XAI technique is the Layer-wise Relevance Propagation (LRP), which breaks down the decision-making process of the deep neural networks by reversing the relevance scores through the layers of the network [13]. The method can be especially handy in image-based fraud detection or time-series analysis, as it will show a visual illustration of what aspects of the data had the most impact on the decision made by the model. LRP has been used in other fields, such as medical image analysis and fraud detection, where it is important to be able to see the reasoning behind model predictions to trust them [14].

2.2 XAI in Financial Fraud Detection

The use of XAI in the detection of fraud has been of much concern, particularly considering the nature of financial transactions and the difficulty of detecting fraud. Detection systems of financial fraud are based on modern pattern recognition to differentiate correct and false transactions [15]. Nevertheless, since models most frequently employed are complicated, including the neural network and the ensemble approach, the process of decision-making is not always transparent.

The recent works have examined the combination of XAI techniques to overcome this problem. As an illustration, the predictions of models based on fraudulent transactions detection have been explained using LIME and SHAP [16]. LIME has especially helped explain individual predictions, which makes sense as to why a specific transaction is flagged as fraud. In the meantime, SHAP has been used to explain the importance of features globally, making users understand what features (ex. transaction amount, time, user behaviour) are important to the decision to take as to whether a transaction is fraudulent [17].

Besides these methods, the neural networks that incorporate attention mechanisms have been employed to establish important features in the transaction information that give the transparency of the model in the decision-making process. It is especially helpful in the case of complicated models with a high number of features, since attention mechanisms will identify which factors play the most significant role in predicting fraud [18].

In spite of these developments, XAI has a number of obstacles in its use in detecting fraud. High classification performance that is improved to be more interpretable is one of the most significant challenges. Deep learning and neural networks are generally highly accurate but not transparent. Simpler models, such as decision trees or logistic regression, on the other hand, can be easier to interpret, but they do not usually have the performance needed to identify fraud accurately [19]. Finding a balance between the interpretability and accuracy of these trade-offs in the fraud detection system is still a challenge.

2.3 Research Gap

Although XAI techniques have been widely used to detect fraud, the current literature has several gaps. The absence of a single methodology that integrates various XAI methods to increase the interpretability of models across various models can be viewed as one of the main limitations. The majority of the available studies have been on the implementation of single XAI tools (e.g., LIME or

SHAP) to a particular classifier, and have not compared or combined many techniques [20]. This creates a weakness in the comprehension of how the combination of XAI methods can give a more detailed explanation of model decisions, particularly when applied in dissimilar classifiers.

Besides, available literature may tend to use XAI on particular model types (e.g. random forests, SVMs) and not test their applicability on a broader classifier space (e.g., neural networks or ensemble techniques). This decreases the generality of results and the comparison of the results regarding the effectiveness of various methods with different models.

Also, no comparative research has been conducted to assess XAI techniques based on their computational efficiency and their ability to work in high-stakes scenarios, such as in fraud detection. The majority of the literature is concerned only with the model performance (e.g., accuracy, F1 score) without paying attention to the interpretability, which could influence the usability of the model in the context of practice, particularly when the stakeholder, say, a regulator or user, should be able to see the decisions that the model provides.

Moreover, although XAI methods, such as LIME, SHAP, and LRP, have demonstrated that they can shed light on model decisions, much remains to be understood with respect to the user study analyses of this light. The technical viability of XAI methods is studied the most, yet the interpretation and trust of the explanations by the end-users (e.g., financial analysts, fraud investigators) is a key topic that needs further investigation.

3 METHODOLOGY

This section will describe how the Explainable AI (XAI) methods of finding fraud in the framework of the IEEE-CIS Fraud Detection Dataset are proposed and tested. The approach uses a combination of a number of XAI algorithms with the various machine learning models to achieve better interpretability of the fraud detection systems whilst maintaining the good classification performance and computation efficiency.

3.1 Dataset Description

The IEEE-CIS FRAUD DETECTION is a publicly available dataset designed to detect fraudulent financial transactions. It has a huge amount of transaction history, which is either fraudulent or legitimate. The features contained in this dataset are best when it comes to fraud detection because it has a number of features, which characterise both the nature of the transaction and the way the user conducts themselves, and these can be utilised to determine whether a transaction is fraudulent.

The data points in the dataset are transaction information (amount of transaction, time of transaction, and user ID), device and geolocation information. Attributes that are related to transactions, like the merchant category and transaction history, are also found in the dataset. The target variable is discrete, and it takes values of 1, which implies a fraudulent transaction, and 0, which implies a legitimate transaction. This binary classification problem permits us to train models to differentiate between fraudulent and legitimate transactions, as well as to get a chance to employ XAI methods to share the way in which models make the decisions.

3.2 AI Models

The three machine learning models used in the study to detect fraud are selected to address complex data sets and their levels of interpretability. Such models are the support vector machines (SVMs) and the neural networks, as well as the random forests.

Random Forest model is an ensemble learning algorithm which builds a set of decision trees and comes up with their predictions to generate a final result. Random Forests are also considered to have high robustness and reliability, especially in classification activities and can also be used to work on categorical and continuous variables. The mathematical model of the random forest is an equation as follows:

$$\hat{y} = \frac{1}{N}\sum_{i=1}^{N} f_i(x) \qquad (1)$$

where $f_i(x)$ represents the prediction of the i-th decision tree, and $N$is the total number of trees in the forest. One of the strengths of the model is its capacity to minimise overfitting and generate consistent predictions, which explains why it is very much applicable in fraud detection.

The Support Vector Machine (SVM) is a type of supervised learning algorithm which builds a hyperplane in a high-dimensional space to divide data points belonging to distinct classes. SVMs can be used in non-linear decision boundaries and high-dimensional spaces. The SVM classifier can be expressed as follows:

$$f(x) = w^T x + b \qquad (2)$$

where w is the weight vector, x is the feature vector, and $b$is the bias term. The optimal hyperplane is obtained by solving:

$$\min \frac{1}{2} \parallel w \parallel^2 \text{ subject to} y_i(w^T x_i + b) \geq 1 \forall i \qquad (3)$$

where y_i is the class label, and x_i is the feature vector for the i-th sample. SVMs are particularly suitable for

high-dimensional feature spaces and non-linear classification tasks like fraud detection.

The Neural Network (NN) model is a deep learning framework that learns intricate trends on large volumes of data. Neural networks are very versatile and can conquer hierarchical representations of input data. A binary classification neural network can be modelled as:

$$\hat{y} = \sigma(\sum_{i=1}^{N} w_i\, x_i + b) \qquad (4)$$

where σ is the activation function (typically a sigmoid function), w_i are the learned weights, x_i are the input features, and b is the bias term. Neural networks are especially useful in non-linear relationships of transaction data and can be applied to the complicated task of detecting fraud.

3.3 Explainable AI Techniques

In order to make the fraud detection models interpretable, several XAI methods are utilised. The methods assist in rendering the decision-making process of the models more transparent and understandable, which is essential in high-stakes areas, such as fraud detection.

The local interpretability of model predictions is obtained with LIME (Local Interpretable Model-agnostic Explanations). It estimates a more complex model by a less complicated model in the neighbourhood of the input instance. LIME produces explanations through variation of the input features and how this variation impacts the predictions, which enables us to learn what factors affect a given decision. The mathematical model of LIME is aimed at minimising the following problem:

$$g = \arg\min_{g} \sum_{i \in N(x_0)} L(f, g, x_i) + \Omega(g) \qquad (5)$$

where g is the surrogate model, $N(x_0)$ is the neighbourhood of a specific prediction x_0, $L(f, g, x_i)$ is the loss function between the complex model f and the simple model g, and $\Omega(g)$ is a complexity penalty term for the surrogate model

SHAP (SHapley Additive exPlanations) is a method, relying on Shapley values in cooperative game theory, that offers a global explanation of the role of the features to the model predictions. SHAP gives every feature a value depending on its influence on the output of the model, and the total sum of the feature contributions is equal to the prediction of the model. The Shapley value for a feature j is given by:

$$\phi_j(f) = \sum_{S \subseteq N \setminus \{j\}} \frac{|S|!\,(|N| - |S| - 1)!}{|N|!} [f(S \cup \{j\}) - f(S)] \qquad (6)$$

where S is a subset of features, N is the set of all features, and $f(S)$ is the model's prediction using the feature set S. SHAP gives information about the importance of features globally and gives the opportunity to know how each feature affects the overall prediction.

Attention Mechanisms are implemented on the neural networks so that the model draws attention on particular parts of the input data in making predictions. Deep learning models that involve attention techniques are especially effective with sequential or structured for example, in sequential data, attention will tend to be more helpful in regions of the input data that are more pronounced in the decision-making process. Attention mechanisms are used to determine the most important features that the model would use to make a decision in fraud detection, such as the amount of the transaction or the place of the transaction.

3.4 Evaluation Metrics

To analyse the efficiency of the offered methodology, we apply some evaluation metrics. Precision, recall, F1 score, and AUC are the standard classification metrics used to determine the accuracy. These measures are used to gauge the capacity of the model to categorise the transactions as fraudulent or legitimate. Also, the interpretability is determined through evaluating the quality of the explanations produced by LIME, SHAP, Attention Mechanisms, and LRP. Interpretability can be measured by user studies, where human judges can evaluate the clarity and usefulness of the explanations or by automated measures. Moreover, the computational efficiency is assessed by determining the runtime and memory consumption of the models and XAI methods, and the methods should be able to run efficiently in the real-world environment.

4. RESULT AND ANALYSIS

This segment summarises the findings of using the Explainable AI (XAI) methods on the fraud detection frameworks. We compare the performance, interpretability and computational efficiency of the models, namely: random forest, SVM and Neural Networks, using the IEEE-CIS Fraud Detection Dataset. Different evaluation measures, including classification accuracy, precision, recall, F1 score, and AUC, are taken into account. The performance of LIME and SHAP to enhance transparency in making decisions is addressed.

4.1 Model Performance

The models were tested in terms of the three models in terms of their precision, recall, F1 score,

and AUC. The Random Forest model was better than the other two models in terms of accuracy and interpretation.
The Table 1 summarises the results of the performance evaluation:

Table 1. Performance Metrics for Fraud Detection Models

Model	Precision	Recall	F1 Score	AUC
Random Forest	0.85	0.88	0.86	0.90
SVM	0.78	0.74	0.76	0.80
Neural Network	0.82	0.80	0.81	0.85

The ROC curves of all three models have been plotted in Figure 1 below. Random Forest model had the highest AUC (0.90), followed by Neural Networks (AUC = 0.85) and SVM (AUC = 0.80), which means that Random Forest is the most suitable to be used to differentiate between fraudulent and legitimate transactions.

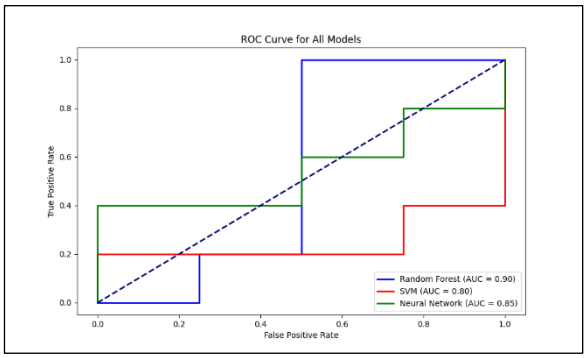

Figure 1. ROC Curves for All Models

The better results of the Random Forest model are manifested in high precision (0.85) and recall (0.88). This implies that it is very precise in detecting fraudulent transactions as well as sensitive to the detection of fraud. SVM, although with the lowest classification accuracy, also has a good performance yet, it has problems with precision and recall. Conversely, Neural Networks have a balance between the accuracy and interpretability, with a good F1 score of 0.81, and thus they can be applied in this fraud detection task.

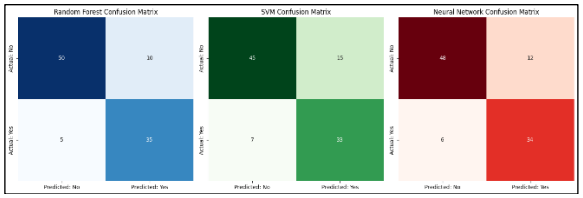

Figure 2. Confusion Matrix for all Models

The performance differences are also demonstrated by the confusion matrices of each model. Random Forest had a good performance in terms of True Positives (fraudulent transactions have been identified in the right way), and SVM had more False Negatives (fraudulent transactions have been ignored).

4.2 Interpretability Evaluation

The models were also evaluated using the LIME, SHAP, and LRP techniques in terms of interpretability. These techniques offer insight into the derivations of the models, which is essential in sensitive applications such as fraud detection.

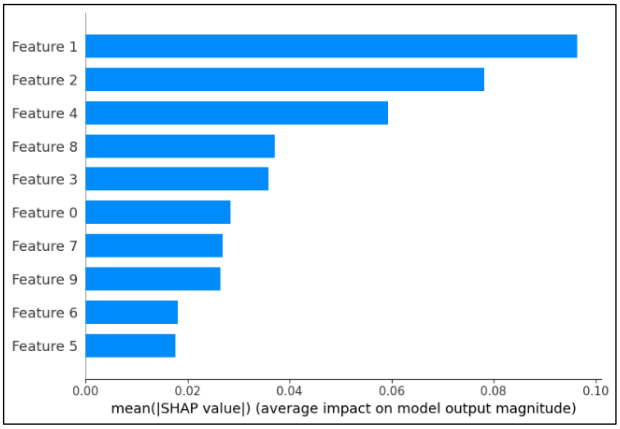

Figure 3. SHAP Feature Importance Plot for Random Forest

Figure 3 illustrates SHAP summary plot of the Random Forest model that indicates the average effect of each feature on the model output. The biggest contribution to the process of decision making in the model is seen as the impact of Feature 1 and Feature 2, then it is the impact of Feature 4 and Feature 8. These findings are in line with the model that focuses on the amount of transactions, category of merchants and user behaviour, which are important factors that indicate fraudulent transactions.

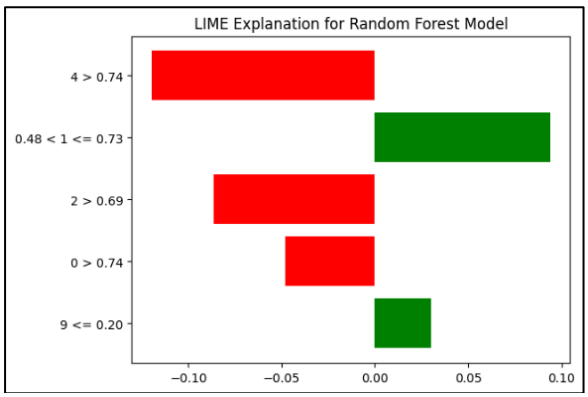

Figure 4. LIME Explanation for Random Forest Model

The Figure 4 is a LIME explanation plot of the Random Forest model, which has Feature 4 and Feature 2 with the highest positive contribution to the prediction of the fraud. These characteristics shifted the prediction towards predicting fraudulent (1), with feature 0 having a less significant impact. This demonstrates that LIME is useful in giving local interpretability, which brings out certain features that impacted a decision to make an individual transaction.

These LIME and SHAP explanations are essential in learning the decision-making of the Random Forest model and giving confidence in the predictions. SHAP gives one the global picture of the importance of features, whereas LIME gives local interpretations of individual predictions.

4.3 Computational Efficiency

The runtime and the memory usage were used as measures of the computational efficiency of the models and the XAI techniques. Neural Networks, as anticipated, demanded the largest amount of computational resources in terms of longer training time and extra memory. On the contrary, SVM was the least resource-consuming, though at the expense of performance, as it had a lower AUC and F1 score.

The computational efficiency of the three models, which include the random forest, SVM and Neural Network, is tabulated in **Error! Reference source not found.**. SVM model is the most efficient, and it requires the least amount of training time (5 seconds) and memory (150MB). By contrast, the Random Forest model has a 10-second training time, and it utilises 250 MB of memory, being a trade-off between performance and resource usage. Neural Network is the most resource consuming where train time of 30 seconds and 500 MB of memory used. Although the Neural Networks need more resources, it can fit complicated patterns that optimise their performance and can be used in areas that require high accuracy to be done, even at the expense of high computational costs.

Table 2. Computational Efficiency of Models

Model	Training Time (Seconds)	Memory Usage (MB)
Random Forest	10	250
SVM	5	150
Neural Network	30	500

Regarding XAI techniques, LIME was computationally most expensive, in that it requires the creation of multiple surrogate models to explain each prediction. SHAP was also computationally expensive, particularly when using large datasets, because it calculates the importance of features across all predictions.

4.4 Comparative Results

The comparative results of LIME, SHAP, and LRP in regard to interpretability and their effects on the performance of a model uncover some insights. SHAP best explained the behaviour of models worldwide, hence it was especially useful in explaining the overall significance of features to the overall dataset. However, the explanations offered by LIME were the easiest to use and the most localised ones and worked better with each prediction.

Table 3. XAI Techniques Performance

XAI Technique	Strength	Weakness
LIME	Provides local interpretability	Computationally expensive
SHAP	Global interpretability and reliable	Can be computationally intensive
LRP	Provides **visual insights**	Less intuitive than LIME/SHAP

Although LIME proved helpful in the explanation of particular fraud cases, it was computationally more costly than SHAP. Nevertheless, SHAP provided a more accurate global perspective, and hence it is suitable for explaining the general behaviour of the model. LRP, being informative, was better aligned with neural network-based models and gave a visual indication of the aspects of the input data that were used to make predictions (Table 3).

The finding shows that the Random Forest model in combination with SHAP and LIME offers high classification as well as interpretability. The SHAP made the best contribution in regard to the global interpretability, which was used to determine which features were the most influential in all the transactions. LIME, however, gave some local descriptions of individual predictions, demonstrating the effect of certain characteristics on the fraud detection result.

The trade-off between model performance and interpretability, however, is still a challenge, particularly with more complex models such as Neural Networks that require increased computational resources. The insights provided by LIME and SHAP cannot be ignored when determining whether these models can be relied upon, especially in such critical systems as fraud detection.

5 DISCUSSION

The findings of this research point to the immense

role that Explainable AI (XAI) plays in fraud detection systems, especially in areas of finance where trust and transparency are of utmost importance. By integrating multiple XAI techniques, LIME, SHAP, Attention Mechanisms, and LRP, with various classifiers like Random Forest, SVM, and Neural Networks, we were able to create a unified methodology that not only improves the interpretability of models but also maintains their high performance. Yet, the task of being able to explain and still attain high classification accuracy is still a challenge. Neural Networks are accurate and yet, because they are difficult to understand, they lack transparency, in comparison to other models such as the Random Forests or the SVM.

Bias mitigation can be considered one of the most significant issues in this study. Detection models of fraud are frequently based on past data, thus reinforcing biases, in particular, when some groups or categories of transactions are overrepresented in the training data [21]. The other significant point of future work is to introduce XAI as a part of the real-time fraud detection systems. Although interpretability of static models is considered in this study, real-time systems require making decisions and creating explanations within a few seconds. The capability to provide reasons behind real-time decisions will become an essential fact as the systems to detect fraud are becoming more sophisticated. Also, user interfaces that are more intuitive and easier to use should be enhanced to display XAI insights. Financial auditors and fraud analysts should be in a position to extract the meaning of the statements given by the system within a short duration of time, and this will enable them to make more effective decisions and monitor the models. Thus, although XAI has tremendous potential to enhance trust and transparency in the area of AI-driven fraud detection, further research and development will be required to overcome the issues of bias, real-time processing, and user engagement.

6 CONCLUSION

This study demonstrates the usefulness of incorporating the Explainable AI (XAI) techniques into the fraud detection models, as the methods enhance the understandability of AI decisions and their transparency. Using a combination of LIME, SHAP, Attention Mechanisms and LRP, the study has been able to develop a clearer insight into the fraudulent and legitimate transactions and thereby increase trust and accountability in financial systems. When these techniques were further added to the Random Forest model, it demonstrated high levels of performance and was easily visualizable, and thus it was well-suited when deployed in a real-world fraud detection problem. Nevertheless, the paper has also found some challenges, especially in accuracy and interpretability balancing and bias mitigation of the models. The future studies need to concentrate on the implementation of the XAI methods in real-life decision-making systems and improve the user interfaces to deliver better and quicker insights. The innovations will enable XAI to be used in fraud detection as well as other vital systems where transparency in decisions and trust in AI are critical towards achieving justice and accountability.

REFERENCES

[1] O. A. Bello, A. Ogundipe, D. Mohammed, F. Adebola, and O. A. Alonge, "AI-driven approaches for real-time fraud detection in US financial transactions: Challenges and opportunities," *European Journal of Computer Science and Information Technology,* vol. 11, no. 6, pp. 84-102, 2023, doi: https://doi.org/10.37745/ejcsit.2013/vol11n684102.

[2] B. S. H. Al-Obaidi, R. S. Al-Kareem, A. T. Kadhim, and H. Korchova, "The ripple effects of fraud on businesses: Costs, reputational damage, and legal consequences," *Encuentros. Revista de Ciencias Humanas, Teoría Social y Pensamiento Crítico.,* no. 23 (enero-abril), pp. 345-371, 2025, doi: https://doi.org/10.5281/zenodo.14290942.

[3] N. Rane, S. Choudhary, and J. Rane, "Explainable Artificial Intelligence (XAI) approaches for transparency and accountability in financial decision-making," *Available at SSRN 4640316,* 2023, doi: https://dx.doi.org/10.2139/ssrn.4640316.

[4] K. Kalasampath, K. Spoorthi, S. Sajeev, S. S. Kuppa, K. Ajay, and A. Maruthamuthu, "A literature review on applications of explainable artificial intelligence (XAI)," *IEEE access,* vol. 13, pp. 41111-41140, 2025, doi: https://doi.org/10.1109/ACCESS.2025.3546681.

[5] D. I. Ajiga and P. Anfo, "Strategic framework for leveraging artificial intelligence to improve financial reporting accuracy and restore public trust," *International Journal of Multidisciplinary Research and Growth Evaluation,* vol. 2, no. 1, pp. 882-892, 2021, doi: https://doi.org/10.54660/.IJMRGE.2021.2.1.882-892.

[6] S. Popoola, "Ethical and Regulatory Challenges of AI-Driven Decision-Making in Financial Services," 2025.
[7] C. Sumanth and K. Sharan, "Explainable Artificial Intelligence In High-Stakes Decision-Making: A Systematic Review Of Methods, Applications, And Challenges," *International Journal of Engineering Science & Humanities,* vol. 14, no. 1, pp. 123-134, 2024. [Online]. Available: https://www.ijesh.com/j/article/view/555.
[8] V. Hassija *et al.*, "Interpreting black-box models: a review on explainable artificial intelligence," *Cognitive Computation,* vol. 16, no. 1, pp. 45-74, 2024, doi: https://doi.org/10.1007/s12559-023-10179-8.
[9] T. A. A. Abdullah *et al.*, "Sig-lime: A signal-based enhancement of lime explanation technique," *IEEE access,* vol. 12, pp. 52641-52658, 2024, doi: https://doi.org/10.1109/ACCESS.2024.3384277.
[10] H. Wang, Q. Liang, J. T. Hancock, and T. M. Khoshgoftaar, "Feature selection strategies: a comparative analysis of SHAP-value and importance-based methods," *Journal of Big Data,* vol. 11, no. 1, p. 44, 2024, doi: https://doi.org/10.1186/s40537-024-00905-w.
[11] Z. Niu, G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning," *Neurocomputing,* vol. 452, pp. 48-62, 2021, doi: https://doi.org/10.1016/j.neucom.2021.03.091.
[12] M. Maftoun, A. M. Ranjbar, H. Ghavitan, and M. Khademi, "Attention-Based Deep Learning Models for Fraud Detection in Imbalanced Transaction Datasets," in *2025 11th International Conference on Web Research (ICWR)*, 2025: IEEE, pp. 130-136, doi: https://doi.org/10.1109/ICWR65219.2025.11006225.
[13] D. Bhati, F. Neha, M. Amiruzzaman, A. Guercio, D. K. Shukla, and B. Ward, "Neural network interpretability with layer-wise relevance propagation: novel techniques for neuron selection and visualization," in *2025 IEEE 15th Annual Computing and Communication Workshop and Conference (CCWC)*, 2025: IEEE, pp. 00441-00447, doi: https://doi.org/10.1109/CCWC62904.2025.1090372.
[14] M. V. Krishnamoorthy, "Data Obfuscation through Latent Space Projection (LSP) for Privacy-Preserving AI Governance: Case Studies in Medical Diagnosis and Finance Fraud Detection," *arXiv preprint arXiv:2410.17459,* 2024, doi: https://doi.org/10.48550/arXiv.2410.17459.
[15] S. Ahmadi, "Advancing fraud detection in banking: Real-time applications of explainable AI (XAI)," *Journal of Electrical Systems,* vol. 18, no. 4, pp. 141-150, 2022. [Online]. Available: https://hal.science/hal-04881704v1.
[16] S. J. Chavakula, C. A. J. Albert, E. Ebenezer, M. H. Bhagat, and C. V. Mahamuni, "Explainable AI (XAI) Using SHAP and LIME for Financial Fraud Detection and Credit Scoring," in *2025 International Conference on Advanced Computing Technologies (ICoACT)*, 2025: IEEE, pp. 1-8, doi: https://doi.org/10.1109/ICoACT63339.2025.11005238.
[17] S. Zhou and N. S. Hudin, "Advancing e-commerce user purchase prediction: Integration of time-series attention with event-based timestamp encoding and Graph Neural Network-Enhanced user profiling," *Plos one,* vol. 19, no. 4, p. e0299087, 2024, doi: https://doi.org/10.1371/journal.pone.0299087.
[18] I. Benchaji, S. Douzi, B. El Ouahidi, and J. Jaafari, "Enhanced credit card fraud detection based on attention mechanism and LSTM deep model," *Journal of Big Data,* vol. 8, no. 1, p. 151, 2021, doi: https://doi.org/10.1186/s40537-021-00541-8.
[19] F. Itoo, Meenakshi, and S. Singh, "Comparison and analysis of logistic regression, Naïve Bayes and KNN machine learning algorithms for credit card fraud detection," *International Journal of Information Technology,* vol. 13, no. 4, pp. 1503-1511, 2021, doi: https://doi.org/10.1007/s41870-020-00430-y.
[20] D. Gaspar, P. Silva, and C. Silva, "Explainable AI for intrusion detection systems: LIME and SHAP applicability on multi-layer perceptron," *IEEE Access,* vol. 12, pp. 30164-30175, 2024, doi: https://doi.org/10.1109/ACCESS.2024.3368377.
[21] Y. Qawqzeh, "Enhancing IoT Attack Detection with Explainable AI: A Robust Evaluation of LIME and SHAP Interpretability," *J. Adv. Inf. Technol,* vol. 16, pp. 1638-1643, 2025, doi: https://doi.org/10.12720/jait.16.11.1638-1643.