JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Enhancing Transparency in AI Models for Credit Scoring and Fraud Detection : A review

Rekha Dangi[1], Charu Kavadia[2], Shourya Dev Singh Rathore[3], Shriya Ameta[4], Yakshi Choubisa[5]

UG Scholars[1,3,4,5] , Assistant Professor[2]

Department of Computer Science and Engineering[1,2,3,4,5]

Geetanjali Institute of Technical Studies, Dabok, Udaipur, Rajasthan, India[1,2,3,4,5]

Abstract : AI is becoming a big force in the financial world today in areas such as credit scoring, fraud detection, etc. But since these machines have become increasingly intelligent, they have been increasingly opaque, and so it has become hard for human beings to grasp how they manage to arrive at a specific conclusion.

Legacy systems are not able to handle huge and dynamic financial data sets and keeping up with emerging fraud techniques. Deep networks and other cutting-edge AI and ML techniques can produce the optimal outcomes but are sure to be viewed as "black boxes" and reduce the degree of accountability and transparency.

This paper outlines the manner in which XAI tools like decision trees, LIME and SHAP facilitate regulators and users to gain confidence while opening the systems to the underlying complexity. The most effective methods of judging credit and fraud detection are support vector machines, Random Forests, and neural networks, says this report.

It should be noted that the combination of XAI perspectives can greatly enhance the explanation of the decision-making process and have high prediction accuracy for ensemble or hybrid models.

In general, transparency is the balancing of regulatory compliance and minimizing bias and leads to a more dependable automation of financial decisions.

Keywords : Explainable Artificial Intelligence (XAI), Artificial Intelligence (AI), Machine Learning (ML), Credit Risk Assessment, Predictive Analytics, Fraud Detection, and Bias Mitigation.

I. INTRODUCTION

Machine learning (ML) and artificial intelligence (AI) advances have enabled financial security systems to better detect forgeries. Yet, the "black box" problem with most AI models remains a chronic and very troubling issue, especially for financial institutions and regulators who desire transparent, accountable, and transparent decision-making. Explainable AI (XAI) has arisen as a necessary solution to this problem , assisting in illuminating opaque model behavior, leading to true stakeholder trust, and providing human-understandable, informative explanations of AI-driven choices. As Jovanovic et al. [11] point out, the combination of advanced blockchain technology with federated learning and XAI is an exciting, secure, and future-oriented method for constructing transparent and accountable credit scoring systems. This paper explores the critical and evolving role of Explainable Artificial Intelligence (XAI) in fraud detection, particularly its influence on clear model interpretability, user trust, and responsible regulatory compliance within financial security systems. Thoughtfully designed interpretable AI models developed by Hasan, Gazi, and Gurung [9] promote transparent and understandable decision-making, while also boosting confidence in credit card fraud detection systems. The goal of this study is to emphasize the importance of ethical transparency and introduce practical methods to improve the readability and trustworthiness of AI-powered fraud detection , all while preserving robust identity verification accuracy through a range of XAI techniques grounded in real-world applications.

II. LITERATURE REVIEW

Talaat et al. [1] have examined the application of deep learning and Explainable Artificial Intelligence (XAI) in order to alleviate the "black box" character of credit scoring models, in the form of an interpretable model as a remedy for predicting credit card defaults. Their contribution showcases the inherent trade-off between the predictability of high-end models and interpretability of financial risk conclusions. Employing SHAP (SHapley Additive exPlanations) within a deep neural network-based model enables the authors to maintain the high accuracy of the model, while also deriving interpretable insights into features in determining default, such as missed payments and unpaid bills.

Using this model, they achieved 83.5% accuracy, 88.2% sensitivity, and 98.8% specificity an example of the potential of interpretable AI to be applied toward trust and knowledge-based decision-making in credit risk analysis..

Vihurskyi [2] suggested a system in applied XAI-augmented capable fraud detection,which was based on LIME (Local Interpretable Model-agnostic Explanations) to explain multiple machine learning models prediction.

The framework, which is for transparency and trust, especially between stakeholders, such as financial analysts and regulators, is applied with the use of balanced version of the Kaggle credit card fraud dataset. In this investigation, it was noted that the decision tree and random forest models were exceptional and had 100% accuracy during experimentation.

This is how XAI is disposed to describe the ML decisions, allowing stakeholders to validate and understand fraud detection outcomes an important prerequisite for both regulatory compliance as well as user trust.

Baisholan et al. [3] introduced FraudX AI, an explainable ensemble-based credit card fraud detection system for credit card imbalanced datasets. To address real-world limitations, their system avoids synthetic oversampling but optimizes class weights and thresholds in Random Forest and XGBoost classifiers. What sets their system apart is the use of SHAP to explain model predictions so that financial analysts and auditors can look beyond complicated ML decisions. With 95% recall and 97% AUC-PR, FraudX AI shows how high-quality detection and transparency can go together, offering a scalable fraud prevention system for financial institutions.

Faruk et al. [4] presents an entire gamut of the increasing application of Explainable AI in financial fraud detection on the basis of the need for transparency for compliance, stakeholder trust, and ethical AI adoption. They mention some of the methods - SHAP, LIME, and counterfactual explanations, describing their application in enabling model interpretability without compromising performance to a great extent. The paper describes the trade-offs in using explainability at the expense of predictive performance and presents a real-time XAI-integrated pipeline for fraud detection. The work adds a healthy conceptual framework to the use of XAI in high-risk financial settings where explainability cannot be compromised.

Nallakaruppan et al. [5], introduced a XAI-framework to gain more transparency in AI loan approval decisions to combat cases where no explanation is provided to customers when being denied a loan. The model provided local as well as global interpretability in credit evaluation with the help of a Random Forest model and LIME and SHAPLEY explainers. The proposed method was validated using a Kaggle dataset and attained a high level of accuracy, sensitivity and specificity of ~99.8%, potentially beating classical approaches, indicating the promise of implementing XAI to enhance trust in financial service applications.

Wang [6] also investigated the effect of Artificial Intelligence and Machine Learning on credit risk prediction, noting that traditional methods such as logistic regression are insufficient. A number of models were being considered, i.e., Random Forests, Neural Networks, and Support Vector Machines (SVMs), with the best performance being noted as that of Random Forest with an accuracy rate of 93%. The research highlighted the utilization of alternative sources of data—namely social media and payment histories—

to increase predictive power. Additionally, use of Explainable AI methods such as Shapley values and LIME was crucial in solving model transparency and fairness concerns, ensuring compliance and confidence among stakeholders.

Nwafor et al. [7] proposed a hybrid machine learning model based on 1DCNN, XGBoost, and logistic regression to improve credit scoring performance and transparency. The model outperformed the single methods with a classification accuracy of 96%. Using SHAP, they gave more local explainability and feature importance insights, solving fairness issues. Removing discriminatory features such as age and gender had a negligible effect on performance, indicating the potential for developing fair, unbiased models without compromising accuracy. The study prioritizes predictive performance as well as ethical considerations in credit risk assessment.

De Lange et al. [8] proposed the use of Explainable Artificial Intelligence (XAI) in predicting bank credit defaults. In this case, push the airport ontology to the real pipeline and process a lot of the raw data from a Norwegian bank for which we integrated LightGBM models with SHAP values to enhance how predictably we could explain the results. We obtained a ROC AUC value of 0.96 which is better than the logistic regression model used by the bank. The tests confirmed that XAI improves predictive accuracy and explainability, encouraging compliance with laws such as the GDPR, as well as more transparent credit decision-making information.

Hasan et al. [9] explored machine learning algorithm application in detecting credit card fraud. To avoid the shortcomings in the application of accuracy alone, this paper gave emphasis to three other measures: precision, recall, and F-measure. The research achieved extremely high performance using SVM, Logistic Regression, Random Forest, and ANN on a Kaggled dataset of more than 284,000 transactions. While detecting actual cases of fraud, SVM achieved the highest recall rate at 89.5, and ANN achieved the highest precision at 79.4, reflecting high accuracy in positive prediction. This observation highlighted the capacity of machine learning models in detecting fraud, especially in regard to the significance of precision and interpretability.

Rane et al. [10] talk about the role of Explainable Artificial Intelligence (XAI) here, providing transparency and accountability in financial decision-making. Among the problems they had to address in their paper is transparency or the lack of transparency in MS due to extremely complex AI models which can pull the transparency directly from the stakeholders of whether AI-based decisions actually work or not. This paper employs different XAI techniques such as decision trees, SHAP, LIME, etc. to highlight the significance of interpretability. The findings show that explainable AI can definitely enhance transparency in decision-making, reduce potential biases, and support the ethical application of AI in important financial scenarios.

Jovanovic et al. [11] investigated the combination of Blockchain technology and the Explainable Federated Learning (XAI) in order to achieve automated credit score with the intention of improving the credit assessment process with the potential of secure sharing of data with various parts while preserving privacy.

Their research has demonstrated that the combination enhanced credit score accuracy and transparency, mitigating the deficiency of traditional methods based on previous data. The proposed framework not only guaranteed the explanation of the model with utmost significance to regulatory compliance but also promoted trust from the stakeholders of the financial sector.

Bücker et al. [12] experimented with incorporating Blockchain technology and interpretable federated learning to automate credit score, highlighting increasing the transparency of models and adherence to explanation-based regulatory requirements. Its suggested architecture successfully blended Blockchain's secure data management capabilities with collaborative learning learning method towards federated learning, achieving a predictive and more accurate credit score model. The research showed substantial improvements in model performance, providing a clear audit trail for decision-making processes, hence building confidence among stakeholders in the banking industry. Lastly, the research identifies the potential to integrate these new technologies to address the evolutionary needs of regulatory frameworks and consumer expectations in credit score uses.

Al Shiam et al. [13] discussed the use of explainable artificial intelligence to enhance credit default prediction. Traditional methods like logistic regression remain popular due to being simple and transparent but will ultimately fail when presented with complex financial trends. In trying to overcome this limitation, the authors experimented with a range of machine learning models like Decision Trees, Random Forests, LightGBM, and XGBoost. Among them, XGBoost was highest on accuracy, precision, recall, and AUC. To gain a better insight into the model's decision-making process, they employed SHAP values, which also extracted the significant factors influencing the predictions. Although their method can be criticized in the context of imbalanced data, the outstanding performance coupled with transparency of their method makes it feasible, unbiased, and implementable in practice for finance applications.

C. Oko-Odion [14] has highlighted that AI-driven risk assessment models have greatly improved predictive accuracy and fraud detection in financial markets by utilizing machine learning and deep learning techniques. These models outperform traditional methods by analyzing large datasets and identifying patterns in real-time. Even with their benefits, there are challenges such as algorithmic bias and model opacity. To offset such drawbacks, Explainable AI (XAI) methods, including SHAP and LIME, have been proposed to improve transparency and justice in AI-driven financial decision-making.

Table 1 Given below summarizes the researchers' work on XAI with Credit Scoring and Fraud Detection

Researcher(s) & Year	Model/Approach Used	Purpose	Result/Key Findings	Transparency & Future Insights
Baisholan et al., 2025	FraudX AI (RF + XGBoost) with SHAP	Credit card fraud detection in imbalanced datasets	Recall: 95%; AUC-PR: 97%; outperformed 8 baselines	SHAP identifies influential features; scalable in the real world with optimized thresholds
Talaat et al., 2023	CreditNetXAI (Deep Learning + SHAP)	Accurately pre dict credit card defaults and in terpret them	Accuracy: 0.8350; Sensitivity: 0.8823; Specificity: 0.9879; PAY_0, PAY_3, BILL_AMT2 are the important features	Very high transparency with SHAP;model ca n be adjusted to wider credit risk issues
Vihurskyi [2]	Decision Tree & Random Forest + LIME (XAI)	Detect fraud with transparency to stakeholders	Models achieved 100% accuracy on balanced Kaggle dataset	LIME makes ML outputs understandable; allows for compliance and stakeholder trust
Nwafor et al., 2024	Hybrid ML + XAI (e.g., SHAP, LIME)	Improve transparency/fairness in automated credit decisions	Achieved high performance and clear decision rationales.	Focuses on fairness; future work includes deployment in live banking
Wang, 2024	AI and ML models (Random Forest, Neural Networks, SVM) Fairness-aware techniques along with explainability tools (Shapley values, LIME	The research aims to improve credit risk assessment accuracy compared to traditional statistical methods and balance predictive performance with fairness.	Found that AI models, notably Random Forest, can achieve high accuracy (93% accuracy, AUC-ROC 0.94) while illustrating the potential trade-offs when incorporating fairness-aware approaches	Recommends developing hybrid models that integrate performance with transparency and fairness-aware methods, ensuring ethical, explainable, and regulatory-compliant credit risk assessment
Faruk et al., 2025	ML with SHAP, LIME, Counterfactuals	Increase transp arency of AI-fraud detection	Increases interpretability without reducing information on fraud detection accuracy	Focuses on stakeholder trust and regulatory adherence; upcoming e mphasis on explainability in real-time
Rane et al., 2023	XAI techniques overview (LIME, SHAP, Counterfactuals)	Enhance accountability and transparency in finance	Reviewed efficacy of multiple XAI techniques	Urges standardization and explainability benchmarks
Jovanovic et al., 2024	Blockchain + Federated Learning + XAI	Automated, decentralized credit scoring	Robust, privacy-preserving, explainable credit decisions	High scalability; suggests cross-institution collaborations.
Hashemi et al., 2025	Voting Ensemble of CatBoost, XGBoost, LightGBM	Enhance fraud detection throu gh sophisticate d ML	AUC-ROC: 0.95; Recall: 0.80; F1-score: 0.81	Strong performing yet computationally ex haustive; explainable through SH AP
Oko-Odion, 2025	AI-driven predictive models	Financial risk assessment and fraud detection	Improved predictive accuracy in financial markets	Suggests real-time monitoring; use of adaptive AI models

III. CONCLUSION

This study highlights the important role of AI (XAI) explain to increase fraud detection systems, explain lecturers, transparency and regulatory compliance. By integrating XAI techniques such as shape and lime, financial institutions can promote confidence among stakeholders and ensure accountability. Research credit reflects the ability to detect AI-managed fraud in changing credit risk modeling and financial decision making. While challenges such as fairness, transparency and accountability remain, this study reflects the benefits of combining AI and XAI in the creation of reliable and clear credit scoring processes. Future research should focus on addressing these obstacles and discovery of the application of XAI in diverse financial sets

REFERENCES

[1] F. M. Talaat, A. Aljadani, M. Badawy, and M. Elhosseini, “Toward interpretable credit scoring: integrating explainable artificial intelligence with deep learning for credit card default prediction,” Neural Comput. Appl., vol. 36, no. 9, pp. 4847–4865, Mar. 2024, doi: 10.1007/s00521-023-09232-2. https://doi.org/10.1007/s00521-023-09232-2

[2] B. Vihurskyi, “Credit Card Fraud Detection with XAI: Improving Interpretability and Trust,” in 2024 Third International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE), Ballari, India: IEEE, Apr. 2024, pp. 1–6. doi: 10.1109/ICDCECE60827.2024.10548159. https://doi.org/10.1109/ICDCECE60827.2024.10548159

[3] N. Baisholan, J. E. Dietz, S. Gnatyuk, M. Turdalyuly, E. T. Matson, and K. Baisholanova, “FraudX AI: An Interpretable Machine Learning Framework for Credit Card Fraud Detection on Imbalanced Datasets,” Computers, vol. 14, no. 4, p. 120, Mar. 2025, doi: 10.3390/computers14040120. https://doi.org/10.3390/computers14040120

[4] Faruk, Nayab & Tariq, Ahmad & Oladele, Sunday & Gok, Mooale. (2025). Explainable AI (XAI) for Fraud Detection: Building Trust and Transparency in AI-Driven Financial Security Systems. https://www.researchgate.net/publication/390235753_Explainable_AI_XAI_for_Fraud_Detection_Building_Trust_and_Transparency_in_AI-Driven_Financial_Security_Systems

[5] M. K. Nallakaruppan, B. Balusamy, M. L. Shri, V. Malathi, and S. Bhattacharyya, “An Explainable AI framework for credit evaluation and analysis,” *Appl. Soft Comput.*, vol. 153, p. 111307, Mar. 2024, doi: 10.1016/j.asoc.2024.111307. https://doi.org/10.1016/j.asoc.2024.111307

[6] Z. Wang, “Artificial Intelligence and Machine Learning in Credit Risk Assessment: Enhancing Accuracy and Ensuring Fairness,” *Open J. Soc. Sci.*, vol. 12, no. 11, pp. 19–34, 2024, doi: 10.4236/jss.2024.1211002. https://doi.org/10.4236/jss.2024.1211002

[7] C. N. Nwafor, O. Nwafor, and S. Brahma, “Enhancing transparency and fairness in automated credit decisions: an explainable novel hybrid machine learning approach,” *Sci. Rep.*, vol. 14, no. 1, p. 25174, Oct. 2024, doi: 10.1038/s41598-024-75026-8. https://doi.org/10.1038/s41598-024-75026-8

[8] P. E. De Lange, B. Melsom, C. B. Vennerød, and S. Westgaard, “Explainable AI for Credit Assessment in Banks,” *J. Risk Financ. Manag.*, vol. 15, no. 12, p. 556, Nov. 2022, doi: 10.3390/jrfm15120556. https://doi.org/10.3390/jrfm15120556

[9] Md Rokibul Hasan, Md Sumon Gazi, and Nisha Gurung, “Explainable AI in Credit Card Fraud Detection: Interpretable Models and Transparent Decision-making for Enhanced Trust and Compliance in the USA,” *J. Comput. Sci. Technol. Stud.*, vol. 6, no. 2, pp. 01–12, Apr. 2024, doi: 10.32996/jcsts.2024.6.2.1. https://doi.org/10.32996/jcsts.2024.6.2.1

[10] N. Rane, S. Choudhary, and J. Rane, “Explainable Artificial Intelligence (XAI) Approaches for Transparency and Accountability in Financial Decision-Making,” *SSRN Electron. J.*, 2023, doi: 10.2139/ssrn.4640316. https://doi.org/10.2139/ssrn.4640316

[11] Z. Jovanovic, Z. Hou, K. Biswas, and V. Muthukkumarasamy, “Robust integration of blockchain and explainable federated learning for automated credit scoring,” *Comput. Netw.*, vol. 243, p. 110303, Apr. 2024, doi: 10.1016/j.comnet.2024.110303. https://doi.org/10.1016/j.comnet.2024.110303

[12] M. Bücker, G. Szepannek, A. Gosiewska, and P. Biecek, “Transparency, auditability, and explainability of machine learning models in credit scoring,” *J. Oper. Res. Soc.*, vol. 73, no. 1, pp. 70–90, Jan. 2022, doi: 10.1080/01605682.2021.1922098. https://doi.org/10.1080/01605682.2021.1922098

[13] Sarder Abdulla Al Shiam *et al.*, “Credit Risk Prediction Using Explainable AI,” *J. Bus. Manag. Stud.*, vol. 6, no. 2, pp. 61–66, Mar. 2024, doi: 10.32996/jbms.2024.6.2.6. https://doi.org/10.32996/jbms.2024.6.2.6

[14] C. Oko-Odion, *"AI-Driven Risk Assessment Models for Financial Markets: Enhancing Predictive Accuracy and Fraud Detection," Int. J. Comput. Appl. Technol. Res.*, Mar. 2025, doi: 10.7753/IJCATR1404.1007 . https://doi.org/10.7753/IJCATR1404.1007