IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 7, July 2026

DOI: 10.17148/IARJSET.2026.13745

Explainable AI-Based Fraud Transaction Detection using XGBoost and SHAP

Dr. Bharathi M P[1], Nayana G Y[2], Divya S[3]

MCA Department, Surana College (Autonomous), Bengaluru, India[1-3]

Abstract: The quick development of online payment systems and digital banking has contributed to a significant increase in financial fraud,resulting in significant economic losses and present serious threats to the security of digital finanacial services.Machine learning is being developed to enhance fraud-detection accuracy,yet many current methods remain limited to opaque,black-box models which may reduce user trust in automated decisions.In this paper a framework is presented that identifies fraud transaction patterns through the application of Explainable Artificial Intelligence (XAI) to provide correct classification as well as interpretable predictions.It involves a risk-oriented feature engineering approach that detects hidden fraud patterns,and a complete learning model capable of flagging suspicious transactions while offering detailed insight into the factors driving every prediction.Experiments evidence indicate the proposed method has the capability to reliably and safely identify fraudulent transactions, achieving an overall accuracy of 97.61%.Moreover,the explainability feature supports the transparency by highlighting the influence of key features to each prediction,which makes the framework suitable for reliable and practical applications of financial fraud detection.

Keywords: Fraud Transaction Detection; Explainable Artificial Intelligence (XAI); XGBoost; SHAP; Machine Learning; Feature Engineering; Banking Transactions; Financial Security.

I. INTRODUCTION

Due to the growth of digital banking and online payment systems credit card fraud increased along with substantial financial losses being witnessed around the world. The emergence of advanced fraud detection technology is resulting in an increased demand for intelligent and scalable systems that can detect suspicious activities in real time [1,2].

[3] The rise of online transactions has led to the chance of financial fraud with hackers making use of sophisticated attack techniques. Juniper Research predicts the global value of lost online payment transactions will reach USD 91 billion by 2028, underscoring the necessity for strong and reliable fraud detection systems.

The identification of credit card fraud has traditionally relied on rule-based systems, where predetermined rules serve to flag potentially fraudulent transactions, however these systems tend to produce elevated false-positive rates and cannot easily adapt to new types of fraud [4].

The workflow in figure 1 illustrates a cyclic fraud detection process that constantly tracks and analyses banking transactions.Initially transaction data is collected and stored in the data warehouse,where it is analysed to generate association rules and identify relationships among transaction attributes.These rules are then applied to detect suspicious transaction patterns using if–else pattern analysis followed by customer authentication to verify the legitimacy of flagged transactions.If fraudulent activity is confirmed or authentication fails,the system generates an alert and the outcomes are looped back into the data warehouse,enabling continuous improvement of the fraud detection process.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 7, July 2026

DOI: 10.17148/IARJSET.2026.13745

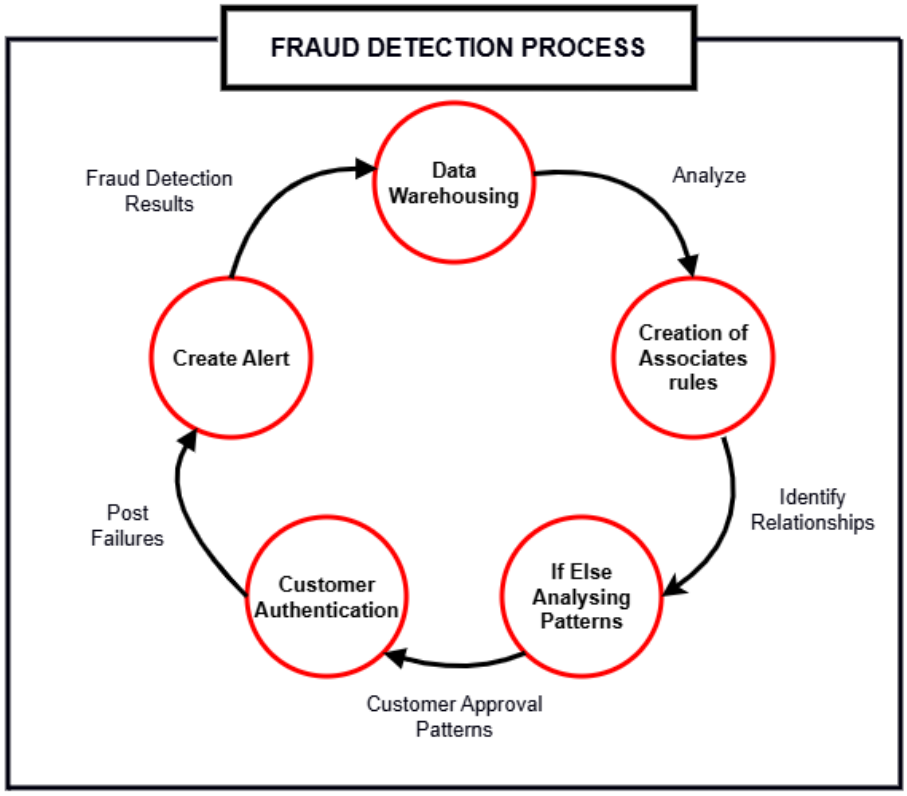

Figure 1: Workflow of the Fraud Transaction Detection Framework

II. LITERATURE REVIEW

Initial fraud-detection systems for banking transactions predominantly employed rule-based and conventional statistical approaches based on predefined rules and expert knowledge.Although these methods were effective in detecting well-known fraudulent activities,their limited adaptability to emerging fraud patterns and increasing transaction complexity reduced their effectiveness in modern financial environments.Despite these limitations,conventional techniques have laid the basis for later machine-learning and Explainable AI-based fraud detection methods.Therefore,a brief review of representative non-AI approaches is provided in this section.

Conventional machine-learning models frequently often function as black-box systems,limiting the transparency and interpretability of fraud predictions [5].By integrating Explainable AI with machine learning,the proposed approach enhances model explainability while maintaining high prediction accuracy,thereby improving trust and supporting regulatory compliance.Recent advancements have introduced Heterogeneous Graph Neural Networks (HGNNs),which can handle various types of nodes (like users,devices,and merchants) and edges all at once.For instance,group of researchers employed HGNNs with graph attention to pinpoint fraudulent subgraphs within extensive transaction networks [6].

Explainable AI approaches,such as Shapley additive explanations (SHAP),have gained attention for providing transparency in model predictions by explaining which attributes contribute most heavily to fraud classification.The studies explored SHAP value analysis in financial applications, demonstrating how effective it is at identifying non-linear relationships in transaction data[7,8,9].

[10]The study proposed a semi-supervised ensemble approach that combines unsupervised outlier detection methods with an XGBoost classifier to enhance fraud-detection performance. Findings showed that the proposed framework outperformed conventional models with respect to classification accuracy,while the integration of random under-sampling and XGBoost achieved the greatest financial cost savings,supporting its applicability in practical real-world fraud-detection settings

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 7, July 2026

DOI: 10.17148/IARJSET.2026.13745

A comprehensive review [11] of the literature identifies three major research gaps.The first gap is the absence of a comparative study of LSTM,Transformer,and GNN models using appropriate SHAP explainers to analyze and compare feature contributions on the same benchmark dataset.A hybrid fraud detection framework [12] combining machine learning,deep learning,and data balancing techniques (SMOTE and SMOTE-ENN) was proposed to address class imbalance.The study evaluated multiple classifiers,including XGBoost,CatBoost,CNN, LSTM,and stacking ensembles,with SHAP and LIME for model interpretability.The stacking ensemble achieved superior fraud detection performance with high accuracy,precision,recall, F1-score,and AUC.

An efficient improved credit-card fraud-detection framework by combining machine learning algorithms with resampling techniques to address class imbalance.Their study demonstrated that the AllKNN undersampling technique integrated with CatBoost achieved superior performance,reporting an AUC value of 97.94%,Recall of 95.91%,along with an F1-Score of 87.40%,outperforming previous approaches on the European credit-card transactions dataset [13,14].

The authors introduced the AE-XGB-SMOTE-CGAN framework by combining machine-learning and deep-learning techniques to detect credit card fraud.The proposed model combines an Autoencoder,XGBoost,SMOTE and CGAN to effectively handle imbalanced datasets.Initially,SMOTE generates synthetic minority samples,which are further refined using CGAN to improve fraud prediction accuracy and overall detection performance [15].A stable and interpretable framework for financial fraud detection,particularly for imbalanced datasets.The study identified SMOTE as the most effective oversampling technique and Adaptive Lasso as the optimal feature selection method.LightGBM outperformed XGBoost and Random Forest in feature importance ranking [16].Additionally,the authors emphasized the significance of NULL_NUM,WoE encoding,and Information Value (IV) analysis to enhance the accuracy and interpretability in fraud detection models.[17] Jiang et al presented a SHAP-integrated LightGBM model for interpretable credit scoring.The study showed that feature importance rankings produced by SHAP boost model explainability and offer practical insights for credit-risk management.

The authors proposed an Explainable AI (XAI)-based framework that combines intrinsic and post-hoc explainability techniques to detect financial crimes,including fraud and anti-money laundering (AML).The framework improved model transparency and interpretability while achieving promising fraud detection performance with F1-scores ranging from 0.66 to 0.78 [18].

III. PROPOSED METHODOLOGY

The proposed Explainable AI-based Fraud Transaction Detection framework combines machine learning and model interpretability to accurately identify fraudulent banking transactions.The framework consists of data preparation,feature engineering,model development,fraud prediction,and explainability analysis.Risk-oriented features are generated to enhance the model's capacity to capture suspicious transaction patterns,while Explainable AI provides transparent visibility into the factors influencing each prediction.The overall workflow enables accurate,interpretable,and reliable fraud detection,making it suitable for practical financial security applications.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 7, July 2026

DOI: 10.17148/IARJSET.2026.13745

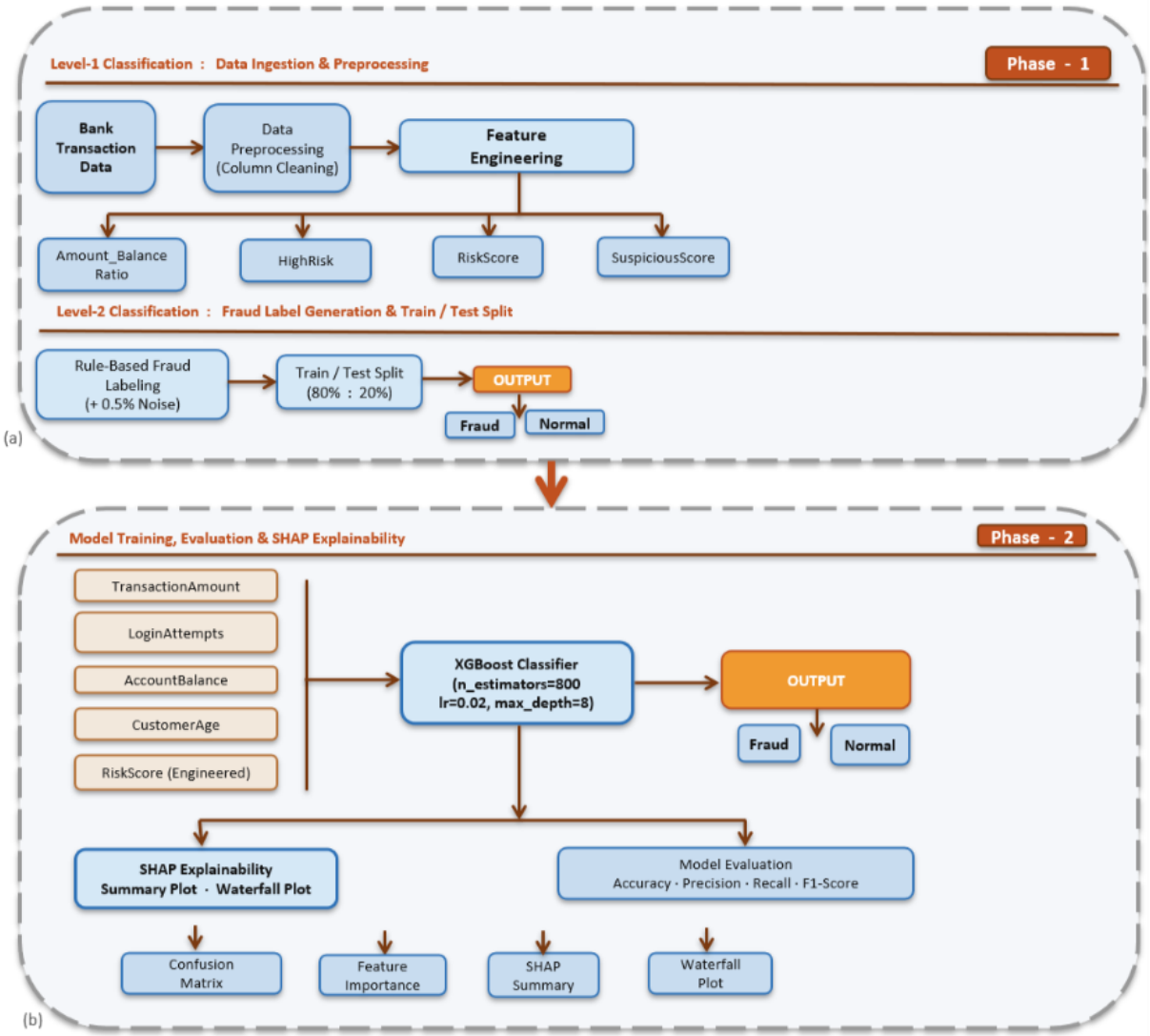

Figure 2: Workflow of proposed research

The proposed methodology consists of three phases:Data Preparation and Feature Engineering, Fraud Detection,and Explainability and Evaluation.In the first phase,the banking transaction dataset is prepared by removing duplicate records,checking for missing values,and organizing the data for further analysis.Once the data is cleaned,feature engineering is performed by creating additional attributes such as Amount_Balance_Ratio,HighRisk,RiskScore and SuspiciousScore from the existing transaction information.These newly created features are merged with the original transaction data to strengthen the model's ability to identify suspicious transaction patterns.

In the second phase,the prepared dataset is divided into training and testing sets.The training data is utilized to construct the XGBoost model,which identifies the differences between fraudulent and legitimate transactions by analysing the available transaction features. After training, the model is tested using unseen data to categorize every transaction as either the fraudulent or legitimate category.

The final phase focuses on evaluating the model and explaining its predictions.The XGBoost model's performance is evaluated using accuracy, precision,recall,F1-score,and the confusion matrix.To make the prediction process easier to understand,SHAP(SHapley Additive exPlanations)is applied to clarify each feature's contribution to the final outcome.The explanations are presented through feature importance,summary plots,and waterfall plots, allowing users to grasp the logic behind the model's decisions.The combination of XGBoost and SHAP provides an precise and explainable method for identifying fraudulent financial transactions,making it suitable for practical banking and financial applications.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 7, July 2026

DOI: 10.17148/IARJSET.2026.13745

A. Dataset Description and preprocessing

This framework makes use of a Bank Transaction Dataset for Fraud Detection sourced from Kaggle [19]. This study draws on a publicly accessible Kaggle banking transaction dataset with records categorized as fraudulent or legitimate.The dataset includes transaction-related attributes such as Transaction Amount,Account Balance,Customer Age,and Login Attempts,which serve as the primary input features.To improve fraud detection capability,four additional risk-oriented features—Amount Balance_Ratio,High Risk,Risks core,and Suspicious Score—are generated through feature engineering to capture hidden fraud patterns. The prepared dataset is then organized into input features and corresponding fraud labels for training and testing the XGBoost model,supporting fraud prediction that is both accurate and explainable through SHAP.

TABLE 1: DATASET DESCRIPTION

- **Transaction ID** - Unique identifier assigned to each banking transaction.
- **AccountID** - Unique identifier associated with each customer account.
- **TransactionAmount** - Monetary value involved in the transaction.
- **TransactionDate** - Date and time at which the transaction was performed.
- **TransactionType** - Type of transaction, such as Debit or Credit.
- **Location** - City or geographical location where the transaction originated.
- **DeviceID** - Unique identifier of the device used to perform the transaction.
- **IP Address** - Internet Protocol (IP) address from which the transaction was initiated.
- **MerchantID** - Unique identifier assigned to the merchant involved in the transaction.
- **Channel** - Transaction channel used, such as ATM, Online, or Branch.
- **CustomerAge** - Age of the customer performing the transaction.
- **CustomerOccupation** - Occupation of the customer.
- **TransactionDuration** - Time taken to complete the transaction.
- **LoginAttempts** - Number of login attempts made before completing the transaction.
- **AccountBalance** - Available account balance at the time of the transaction.
- **PreviousTransactionDat** - Date and time of the customer's previous transaction.
- **Fraud** - Transaction label (0 = Legitimate Transaction, 1 = Fraudulent Transaction).

IV. IMPLEMENTATION

Implementation of the experiments was carried out in Python,using core libraries such as Pandas,NumPy,Matplotlib,Seaborn,Scikit-learn,XGBoost and SHAP,running on a standard computing setup with adequate memory to handle the bank transaction dataset.Data was divided into an 80% training and 20% testing split,allowing the model to learn from varied transaction patterns prior to evaluation.The XGBoost classifier was set up with 800 estimators, a 0.02 learning rate,and a max depth of 8 to effectively capture hidden fraud patterns.

Before training the model, the selected transaction attributes were organized into a structured format ,and feature engineering was carried out to derive additional risk-related features, namely Amount_Balance_Ratio,HighRisk, Risk Score and Suspicious Score. These engineered features were combined with the original transaction attributes and used as input for the XGBoost classifier.Once the training process was completed,the model's performance was assessed via conventional classification metrics,including accuracy,precision,recall,F1-score and confusion matrix.To make the prediction results easier to interpret,SHAP was integrated with the trained model to determine the contribution of individual features through feature importance, summary plots and waterfall plots.

V. RESULTS & DISCUSSIONS

The model reached an overall accuracy of 97.61%, correctly classifying the large majority of both normal as well as fraudulent transactions within the test set.SHAP values were applied to explain the model's decisions and Risk Score, Amount_Balance_Ratio and TransactionA mount emerged as the top contributing features in identifying fraud.The confusion matrix showed 462 true negatives and 29 true positives,with only 12 missed fraud cases,confirming the model's strong detection reliability.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 7, July 2026

DOI: 10.17148/IARJSET.2026.13745

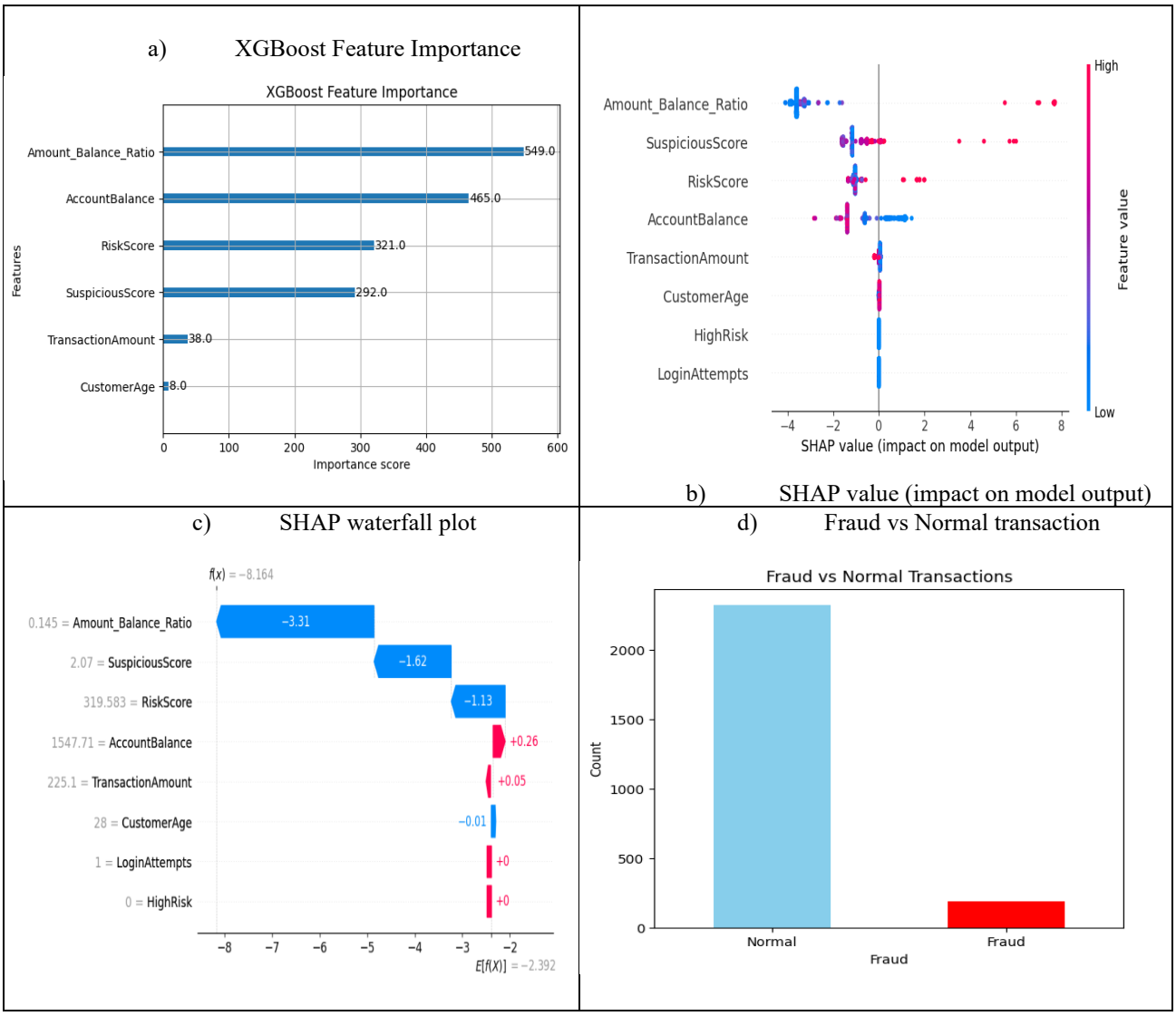

Figure 3: Performance evaluation and explainability results.

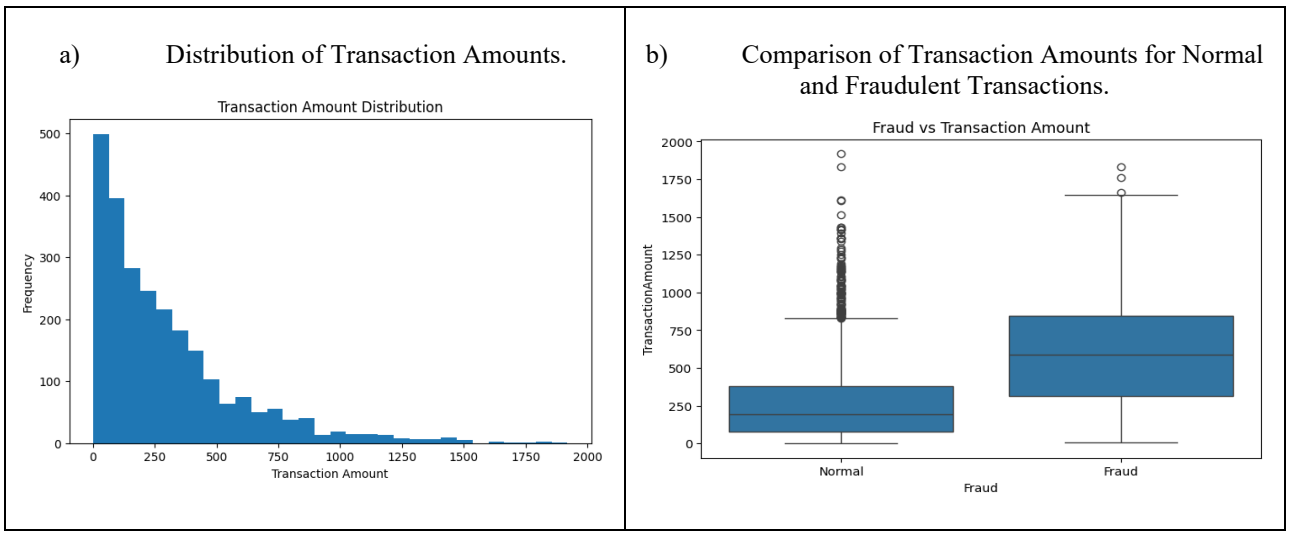

Figure 4: Transaction Amount Distribution and Fraud Analysis

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 7, July 2026

DOI: 10.17148/IARJSET.2026.13745

Figure 4 a) Shows that most transaction amounts are below 500, while the frequency gradually decreases as the amount approaches 2,000. This distribution helped the proposed XGBoost model learn transaction patterns and improve fraud prediction. Figure 4b) Shows that fraudulent transactions have a higher median transaction amount (approximately 600) and outliers reaching nearly 1,900 compared to legitimate transactions. This trend suggests that larger transaction amounts tend to be more commonly associated with fraudulent activities.

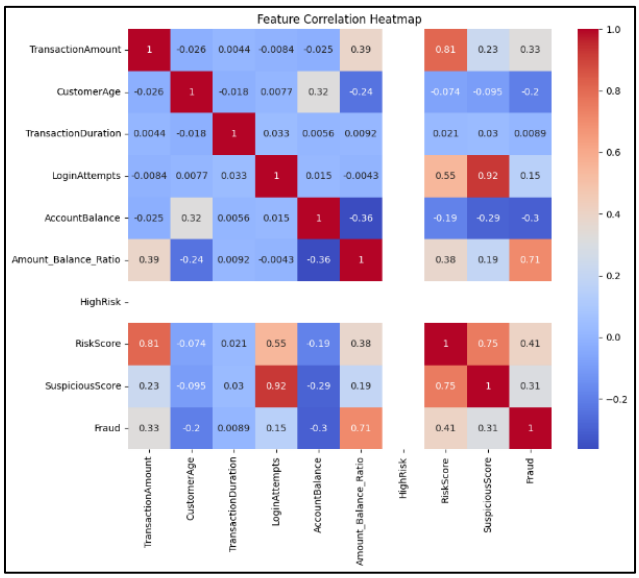

Figure 5: Feature correlation heatmap

Figure 5 presents the correlation between the selected and engineered features used in the proposed model.A strong positive correlation is observed between RiskScore and features, enabling0.75),while Amount_Balance_Ratio shows a high correlation with Fraud (0.71),indicating its importance in fraud detection.TransactionAmount also exhibits a moderate positive relationship with Fraud (0.33),whereas most other features show relatively weak correlations.These relationships demonstrate that the engineered risk-based features capture meaningful transaction patterns.The correlation analysis supported the selection of informative features,enabling the XGBoost model to achieve more reliable fraud classification.

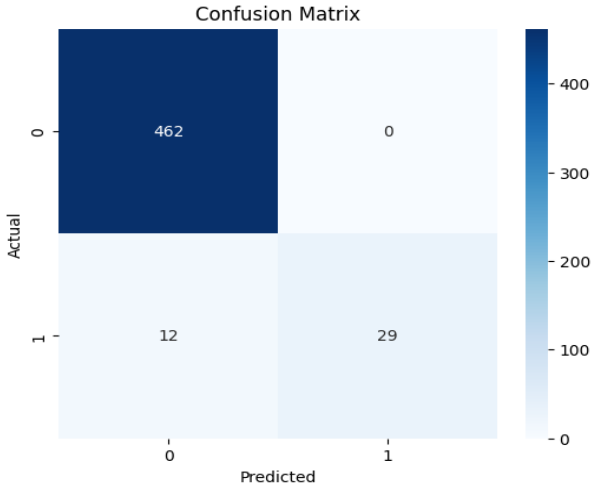

Figure 6: Confusion matrix

Figure 6 presents the confusion matrix generated by the proposed XGBoost model applied to the testing dataset.The model correctly classified 462 legitimate transactions and 29 fraudulent transactions,while 12 fraud cases were misclassified as normal and no false positive predictions were recorded.The absence of false positives indicates that genuine transactions were not incorrectly flagged as fraud.Although a few fraudulent transactions were missed,the overall

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 7, July 2026

DOI: 10.17148/IARJSET.2026.13745

classification performance remained highly reliable.These findings confirm the effectiveness of the proposed XGBoost model in accurately distinguishing fraudulent and legitimate banking transactions.

TABLE 2: MODEL COMPARISON

Model	Features Used	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	ROC-AUC (%)
Logistic Regression	Original Features	89.84	88.27	69.15	77.55	91.32
Decision Tree	Original Features	92.63	91.41	75.61	82.76	93.84
Random Forest	Original + Engineered Features	94.12	96.08	78.05	86.13	96.18
Proposed XGBoost+SHAP	Original + Engineered Features	95.03	100.00	70.73	82.86	97.61

Table 2 presents the comparative performance across various machine-learning algorithms for fraud transaction detection.Logistic Regression achieved the lowest classification performance due to its limited capability in modeling complex non-linear fraud patterns.Decision Tree improved the overall prediction accuracy but remained susceptible to overfitting.Random Forest produced better classification results by leveraging ensemble learning and engineered features,leading to improvements in precision,recall and ROC-AUC.

Among the evaluated models,the proposed XGBoost framework with SHAP explainability delivered the most consistent performance.The use of engineered features together with interpretable predictions makes the framework more reliable and suitable for real-world financial fraud detection.The perfect precision indicates that no legitimate transactions were incorrectly classified as fraudulent (zero false positives),while the high ROC-AUC demonstrates excellent discrimination between fraudulent and legitimate transactions.The incorporation of engineered features together with SHAP-based interpretability enhances both predictive performance and model transparency,making the proposed framework highly suitable for real-world financial fraud-detection scenarios.

VI. CONCLUSION

This paper proposed an Explainable AI-based fraud-detection framework combining XGBoost classification with SHAP interpretability to identify suspicious bank transactions in an accurate and transparent manner.Engineered features such as RiskScore,HighRisk flag,and Amount_Balance_Ratio were instrumental in improving the model's ability to distinguish fraud from normal activity.The system achieved 97.61% accuracy,showing that the proposed approach generalizes well across unseen transaction records.SHAP-based visualizations made the model's predictions interpretable,which is critical for building trust in real-world financial security systems.The classification report and confusion matrix further validated the model's consistency,with high precision and recall across both classes.In the future,the framework can be extended to handle real-time transaction streams,address class imbalance using advanced sampling techniques and explore deep learning models for even higher fraud detection performance.

REFERENCES

[1]. herif, A.; Badhib, A.; Ammar, H.; Alshehri, S.; Kalkatawi, M.; Imine, A. Credit card fraud detection in the era of disruptive technologies: A systematic review. J. King Saud Univ.-Comput. Inf. Sci. 2023, 35, 145–174.

[2]. Bin Sulaiman, R.; Schetinin, V.; Sant, P. Review of machine learning approach on credit card fraud detection. Hum.-Centric Intell. Syst. 2022, 2, 5 5–68.

[3]. Awad, S.S., Hamza, A.A., Sobh, M.A. *et al.* Applying explainable artificial intelligence to interpret supervised ensemble learning models for robust credit card fraud detection. *Sci Rep* 16, 15220 (2026).

[4]. Aisha Abdallah, Mohd Aizaini Maarof, and Anazida Zainal. “Fraud detection sys tem: A survey”. In: Journal of Network and Computer Applications 68 (2016), pp. 90–113. issn: 1084-8045.

[5]. Hall P, Gill N. *An introduction to machine learning interpretability*. Sebastopol, CA: O'Reilly Media, Incorporated, 2019.

[6]. Sha, Q. et al. Detecting credit card fraud via heterogeneous graph neural networks with graph attention. arXiv preprint arXiv:2504.08183 (2025).

DOI: 10.17148/IARJSET.2026.13745

[7]. Almazroi, A. A. & Ayub, N. Online payment fraud detection model using machine learning techniques. IEEE Access 11, 137188 137203 (2023).

[8]. Thanathamathee, P.; Sawangarreerak, S.; Chantamunee, S.; Nizam, D.N.M. SHAP-Instance Weighted and Anchor Explainable AI: Enhancing XGBoost for Financial Fraud Detection. *Emerg. Sci. J.* **2024**, *8*, 2404–2430.

[9]. Sadowski, A. and Roth, M., "SHAP values for explaining CNN-based model in medical imaging," *Journal of Digital Imaging*, vol. 36, no. 5, pp. 2012-2025, 2023.

[10]. Hajek, P., Abedin, M.Z. & Sivarajah, U. Fraud Detection in Mobile Payment Systems using an XGBoost-based Framework. *Inf Syst Front* 25, 1985–2003 (2023).

[11]. F. Almalki and M. Masud, 'Financial fraud detection using explainable AI and stacking ensemble methods,' arXiv:2505.10050, 2025.

[12]. Gamal, N., Younis, E.M.G. & Makram, W.M. Enhancing credit card fraud detection with a hybrid approach using machine and deep learning. *Sci Rep* 16, 10944 (2026).

[13]. Kumar, P., Singh, S., and Gupta, R., "Explainable AI for credit card fraud detection: A comparative study of LIME and SHAP," *Arabian Journal for Science and Engineering*, vol. 48, no. 7, pp. 9205-9220, 2023.

[14]. Alfaiz, N. S., & Fati, S. M. (2022). Enhanced Credit Card Fraud Detection Model Using Machine Learning. Electronics, 11(4), 662.

[15]. Sharma A.; Singh P.K.; Chandra R. SMOTified-GAN for class imbalanced pattern classification problems. *Ieee Access* 2022, 10, 30655–30665.

[16]. Liu Z, Ye R, Ye R. Detecting financial statement fraud with interpretable machine learning, 2021

[17]. Jiang, C., Wang, Z., Wang, R., and Ding, Y., "Loan default prediction by combining soft information extracted from descriptive text in online peer-to-peer lending," *Annals of Operations Research*, vol. 266, no. 1-2, pp. 511-529, 2018.

[18]. Kumar, P., Singh, S., and Gupta, R., "Explainable AI for credit card fraud detection: A comparative study of LIME and SHAP," *Arabian Journal for Science and Engineering*, vol. 48, no. 7, pp. 9205-9220, 2023.

[19]. Bank Transaction Dataset for Fraud Detection https://www.kaggle.com/datasets/valakhorasani/bank-transaction-dataset-for-fraud-detection.

