iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** XII **Month of publication:** December 2025

DOI: https://doi.org/10.22214/ijraset.2025.76120

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XII Dec 2025- Available at www.ijraset.com

An Explainable and Privacy-Preserved Machine Learning Framework for Financial Fraud Detection

Amjad Khan Patan[1], Dr. Ch. D.V. Subba Rao[2]
[1]M. Tech Student, [2]Professor, Department of Computer Science and Engineering, Sri Venkateswara University College of Engineering, Tirupati, A.P

Abstract: ***Banking and online financial service providers face significant challenges due to financial fraud. Traditional fraud-detection methods are often inadequate because of imbalanced datasets, limited interpretability, and privacy concerns involving confidential customer information. This paper presents an explainable AI–based system for financial fraud detection designed to address these issues. The system employs the Light Gradient Boosting Machine (LightGBM) as the primary model, combined with SMOTE oversampling to mitigate class imbalance. Privacy is maintained by anonymizing sensitive features, including Personally Identifiable Information (PII), by temporarily adding and later removing attributes such as name_email_similarity before model training. Model transparency is achieved through SHAP (Shapley Additive Explanations), which offers feature-level interpretability for fraud predictions. The system is implemented as a web-based interactive dashboard using the Flask framework, enabling users to upload datasets, perform fraud detection, adjust detection sensitivity (via threshold tuning), and download a detailed fraud report. When evaluated on a real-world dataset, the system achieved an overall accuracy of 98.5%, an ROC-AUC of 0.89, improved privacy preservation, and enhanced interpretability through SHAP. The proposed solution provides a practical end-to-end framework that balances accuracy, transparency, and privacy protection, making it suitable for banking and fintech fraud-detection applications.***

Keywords: Financial Fraud Detection, Explainable AI, LightGBM, SMOTE, Privacy Preservation

I. INTRODUCTION

In general, fraud can be described as a deliberate act of misleading or lying consciously, and with ill intent, to gain unlawful and usually personal benefit. Frauds are not new, and they continue to develop and change over time [1]. The growing forces of globalization, electronic finance, and electronic communication have increased the capability of fraudsters, making fraud more sophisticated, prevalent, effective and costly to financial institutions [2]. Industry reports show that financial institutions lose billions of dollars every year to fraud, ranging from the fraudulent use of credit cards to large-scale cybercrimes[3]. Digital banking has escalated the challenge of fighting fraud due to the modern paradigm shift from traditional banking to digital banking. The use of credit cards, digital wallets, internet banking, and UPI-based payment systems provides convenience and access to financial services but simultaneously increases risk [4]. Frauds are committed by cybercriminals by exploiting weaknesses in user authentication, user behavior and transaction activities [5]. Furthermore, fraud is not limited to loss of money - it also includes money laundering, telecommunications fraud, identity fraud, and computer systems intrusion. Fraud prevention and fraud detection are two different aspects that should be developed. Some of the efforts put in place with regard to fraud prevention include PINs, encryption, biometrics, full disclosure, and anti-money laundering laws. However, no perfect or inviolable system of prevention exists [6]. Fraudsters keep finding methods of evading the fraud prevention systems. In that regard, fraud detection is also crucial; detection controls warn about possible malfeasance that may lead to serious losses in complex cases by detecting suspicious or unusual transactions in time [7].

II. PROBLEM DEFINITION

Detecting fraud may not be easy. It is linked to several challenges connected to the nature of the problem. Human Privacy: The PII in transaction data can be regulated by privacy laws such as GDPR, which restricts the opportunities to disclose raw data between institutions [8]. Class Imbalance: Fraudulent transactions are infrequent (generally less than 1 per cent of all records) and hence a standard classifier cannot easily identify fraud [9]. Actually, even a typical classifier can be correct on 99 per cent of the instances by merely assuming that all the transactions are of type Not Fraud.

Black Box Models: High-accuracy models, such as deep learning and gradient boosting solutions, will have high accuracy but have no explainable decision-making processes [10]. To regulators, auditors and the banks, they would wish to know the reason why a transaction has been flagged. Dynamic Nature of Fraud: Fraud quickly adapts and alters like techniques that they use. Historical models will lose relevance soon when they are learned models that fail to adjust to the change in patterns. Scalability: Large institutions usually handle millions of transactions daily. The detection algorithms of the fraud should be effective and computationally feasible

III. OBJECTIVES

The major objectives of this thesis are to develop a centralized machine learning model that is utilized to detect fraud. Assist in privacy protection with feature anonymization. Use the Synthetic Minority Oversampling Technique (SMOTE) to overcome the imbalance of classes [11]. Create a fast and high-performing LightGBM model on tabular financial data [12]. Make it explainable with SHAP so that stakeholders can make sense of the logic behind predictions [12]. Create a web-based dashboard to enable users to insert transactional data, do fraud detection and explanations in an interactive fashion. The project will focus on a centralised training perspective of detecting fraudulent transactions based on anonymised data. Even though federated learning was a possibility, the system adopted centralised training due to practical issues regarding the instability of the method and also due to reduced recall. Assessment of the system on all assessment measures, such as Accuracy, Precision, Recall, F1-score, and ROC-AUC. Illustrating how to solve in Flask as a web application. Detection of fraudulent transactions and explanation of the uploaded transactions on a nearly real-time basis.

IV. RELATED WORK

*[13] Tomisin Awosika, Raj Mani Shukla, And Bernardi Pranggono." Transparency And Privacy: The Role of Explainable AI And Federated Learning in Financial Fraud Detection"*Detection and fraudulent transactions are still a major issue for financial institutions in the world. Some of the main challenges are very skewed datasets of transactions and the need to comply with the laws of privacy of data that do not allow them to share information about their customers. The current study proposes a new method based on Federated Learning (FL) and Explainable AI (XAI), which can solve such problems. FL allows financial institutions to jointly train a model to identify fraudulent transactions without a direct stake in customer data, hence preserving privacy. In the meantime, the implementation of XAI is the guarantee that the human experts will be able to understand and interpret the predictions offered by the model and make the system transparent and trustworthy. As per the results of the experiment, which are grounded in realistic dataset of transactions, it has been shown that the FL-based fraud detection system is able to consistently record high performance metrics which form the basis of the potential of FL as an effective and privacy-preserving tool in combating fraud.

*[14] Kuldeep Randhawa 1, Chu Kiong Loo 1, (Senior Member, Ieee), Manjeevan Seera 2,3, (Senior Member, Ieee), Chee Peng Lim4, And Asoke K. Nandi5,6, (Fellow, Ieee) "Credit Card Fraud Detection Using Adaboost and Majority Voting"*Credit card fraud has been a menace in the financial service sector, and billions of dollars are lost annually. This research employs twelve machine learning models, including traditional neural networks and deep learning models, to identify credit card fraud. The explored methodologies are standard models and hybrid ones, i.e. the application of AdaBoost and majority voting. The model effectiveness is tested on a publicly available credit card data set as well as an actual credit card data set of a financial organization, with noise injected into samples to test robustness. The most important contribution is the consideration of a diversity of machine learning models using a real credit card data set obtained based on real transactions. The positive findings of the experiment are that most of the methods of voting have high accuracy rates in fraud cases in credit cards.

*[15] Sahil Somaji Kamble, Suyog Sudhir Pawar, Tejas Babasaheb Veer, Digambar M. Padulkar- "Fraud Detection using Quantum Machine Learning (QML)"*The article proposes an alternative approach to finding fraud that involves quantum-based feature engineering with neural network modelling, and the classification accuracy is 92%. The model defines the QuantumCircuit, part of the Qiskit, to encode the input features to quantum states by angle encoding, which enhances the representation of data and the ability to capture the effect of complex feature interaction within the system, like entanglement between qubits. The technology makes use of a Quantum Neural Network (QNN) to categorise records and detect fraudulent activity. With experiments on real-world datasets, the method achieved both a precision of 0.65 and a recall of 0.68 on fraudulent transactions and a false positive rate of less than 10%. The article concludes that quantum-inspired techniques have the potential to improve fraud detection systems tremendously.

*[16] Seunghyeok Oh, Jaeho Choi, and ◦Joongheon Kim "A Tutorial on Quantum Convolutional Neural Networks (QCNN)"*Convolutional Neural Network (CNN) is a popular computer vision model, but it has difficulties in performing an efficient learning process when the dimension of data or the model is too large. The new solution available is a quantum Convolutional Neural Network (QCNN) that utilises a quantum computing environment. The initial proposed study suggests that to solve the problem of classification in quantum physics and chemistry, it is effectively solved using the CNN structure by applying it to the quantum computing environment to allow calculations with O (log (n)) depth using the Multi-scale Entanglement Renormalisation Ansatz (MERA). The second study presents a hybrid model approach as a way of enhancing performance through the addition of a layer to the CNN learning model with quantum computing. This quantum hybrid model may utilise small quantum computers. The article confirms the hypothesis that the QCNN model can learn effectively compared to CNN by training on the MNIST dataset on the TensorFlow Quantum platform.

*[17] Siddhartha Bhattacharyya, a, Sanjeev Jha, b,1, Kurian Tharakunnel c, J. Christopher Westland "Data mining for credit card fraud: A comparative study"*Credit card fraud is a major and increasing menace. In this paper, the three methods of credit card fraud detection are evaluated: the famous Logistic Regression (LR) and two more sophisticated methods of data mining, Support Vector Machines (SVM) and Random Forests (RF). It uses real-life transaction data of an international credit card operation on which the study is based. The research questions, such as the issue of unbalanced class sizes, with random undersampling to change the train data fraction of fraud cases (15%, 10% 5% 2% fraud cases). It is measured by several metrics, such as the fraud capture rates at different levels of data. The overall performance of the Random Forests showed improvement in terms of the metrics, with a high rate of fraud capture at deeper file depths, which is of great importance in practice in managing *fraud.*

V. METODOLOGY

The methodology of the project relates to the procedures that were followed to develop and operate the Explainable AI-Based Financial Fraud Detection System. It describes the method of the preparation of raw data, how it guarantees privacy, how SMOTE handles the issue of the unbalanced classes, how LightGBM is used to train the model, and how SHAP assists in clarifying the model decisions.

A. System Architecture

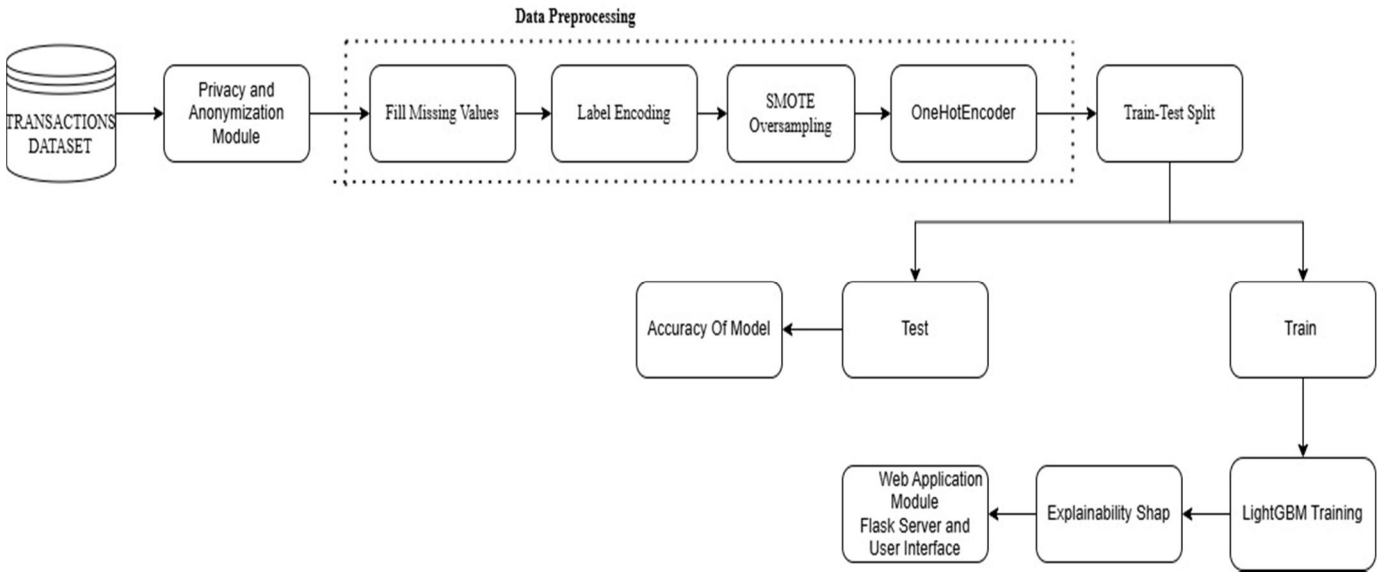


Fig. 1. System Architecture

The main purpose of this Figure 1 is to illustrate the overall workflow of the Explainable AI-Based Financial Fraud Detection System and show how different modules interact within the system architecture. The architecture deploys a central Explainable AI model and integrates a number of modules.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XII Dec 2025- Available at www.ijraset.com

Data Ingestion Module: This module loaded the file FilteredBase.csv into the system. Privacy Layer: It eliminates features that can expose individual information, like nameemailsimilarity. Preprocessing Module: Cleanup of data, features transformation and encoding of features.
Model Training Module: The model is a system that trains a model of fraud detection using LightGBM. Evaluation Module: It computes key metrics, such as accuracy, precision, recall, ROC-AUC, and F1-score. Explainability Module: SHAP visualisations are provided that can explain the process by which the model arrives at its decision. Web Deployment Layer: This is a Flask-based application that allows the user to upload new datasets and view the predictions on a web interface.

B. Proposed LightGBM Model

LightGBM (Light Gradient Boosting Machine) is scalable and simple to interpret, and it is suitable when one has structured or tabular data. It is based on the Gradient Boosting Decision Tree (GBDT) algorithm, but it is optimised to be able to run faster and consume less memory.

C. Explainable AI Integration

SHAP applies cooperative game theory to provide each feature with a score of its contribution. It offers answers to general patterns of the model and single projections.

VI. IMPLEMENTATION DETAILS

A. Data Set

The project employs a dataset (illustrated in Fig. 2: Data Set) containing information on the transactions of financial customers and demographics. All the records in the dataset are single transactions, which have such attributes as income, age, address history, how frequently it is doing transactions over time, and the device specifications. The variable of interest in this case is fraudbool. It shows: 1 - Fraudulent Transaction 0 - Legitimate Transaction[18]

```
data.head(10)
```

	Unnamed: 0	fraud_bool	income	name_email_similarity	prev_address_months_count	current_address_months_count	customer_age	days_since_request	intended_balcon_amount	payment_type	...	has_other_cards	proposed_credit_limit	foreign_request	source	session_length_in_minutes	device_os	keep_alive_se
0	0	0	0.3	0.986506	-1	25	40	0.006735	102.453711	AA	...	0.0	1500.0	0.0	INTERNET	16.224843	linux	
1	1	0	0.8	0.617426	-1	89	20	0.010095	-0.849551	AD	...	0.0	1500.0	0.0	INTERNET	3.363854	other	
2	2	0	0.8	0.996707	9	14	40	0.012316	-1.490386	AB	...	0.0	200.0	0.0	INTERNET	22.730559	windows	
3	3	0	0.6	0.475100	11	14	30	0.006991	-1.863101	AB	...	0.0	200.0	0.0	INTERNET	15.215816	linux	
4	4	0	0.9	0.842307	-1	29	40	5.742626	47.152498	AA	...	0.0	200.0	0.0	INTERNET	3.743048	other	
5	5	0	0.6	0.294840	-1	369	30	0.024232	-1.232556	AD	...	0.0	200.0	0.0	INTERNET	6.987316	linux	
6	6	0	0.2	0.773085	22	4	40	0.006919	-0.544676	AB	...	0.0	200.0	0.0	INTERNET	28.199923	x11	
7	7	0	0.8	0.153880	-1	103	40	0.045122	-1.101184	AB	...	1.0	200.0	0.0	INTERNET	11.234264	other	
8	8	0	0.3	0.523655	21	2	30	0.035206	-0.955737	AB	...	0.0	200.0	0.0	INTERNET	5.329387	other	
9	9	0	0.8	0.834475	-1	134	20	0.017245	-1.356393	AD	...	0.0	1500.0	0.0	INTERNET	4.103970	other	

10 rows × 33 columns

Fig.2 Data Set

B. Data Balancing

The datasets on fraud detection are skewed in nature, and consequently, the models are biased towards the majority (non-fraud) category. In response to the same, the Synthetic Minority Over-sampling Technique (SMOTE) was used on the training set.

C. Data Pre-Processing Anonymization:

The preprocessing is used to ensure that the dataset is clean, structured and is ingestible by the model. The transformation of the following was conducted through a scikit-learn Pipeline:
Missing Value Imputation: Numerical features - substituted with the average of the column. Categorical features - substituted by the most common one. Encoding and Scaling: One-Hot Encoding was used to transform non-numeric labels into binary vectors. Standards of the numerical features were Scaled Standard to have an equal features range which enhances convergences.
Anonymization: Sensitive data, such as name, email, similarity is eliminated to provide privacy of data.

D. Lightgbm_Model

The LightGBM was chosen because it is efficient and effective on large-scale and tabular data. LightGBM is grounded on the Gradient Boosting Decision Tree (GBDT) algorithm, which forms an ensemble of weak learners (decision trees) in a stage-wise fashion to reduce the prediction error.

E. Training and Validation

The centralised model had been trained on 80 per cent of the data (after anonymisation and SMOTE balancing) and tested on the other 20 percent of test data.

F. Performance Matrix and Confusion Matrix

The model scored the following scores on evaluation:

METRIC	SCORE
Accuracy	0.9852
ROC-AUC	0.8910
Precision	0.2982
Recall	0.1789
F1-Score	0.2237

The confusion matrix of the proposed model is demonstrated in Fig 3: Confusion Matrix, which gives a more detailed analysis of correct and incorrect predictions on both classes

G. Interpretation

The accuracy (98.5) is high, which means that it has a good prediction reliability when dealing with normal transactions. From Fig.4, the ROC curve shows the ROC-AUC of 0.89 which indicates that there is high model discrimination between fraud and non-fraud cases. Though this is moderate, recall can be enhanced by increasing or decreasing the decision threshold (i.e., a 0.5 to 0.1) to find more fraud at the expense of a few false positives

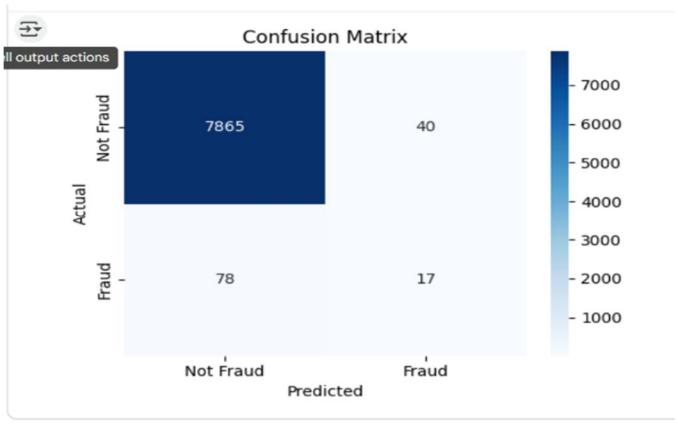


Fig 3: Confusion Matrix

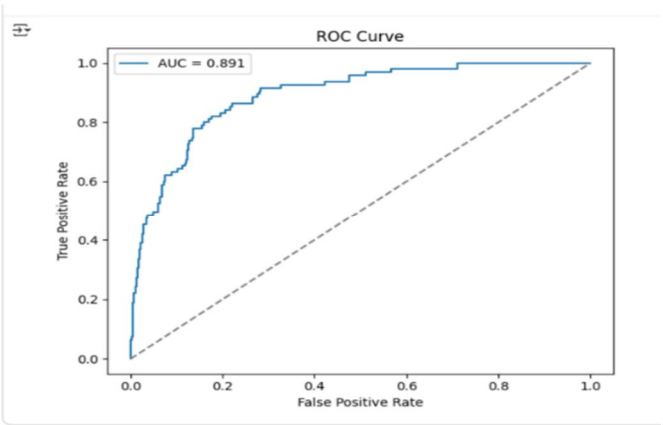


Fig 4: ROC Curve

H. Explainable Ai Integration

SHAP (Shapley Additive explanations) was added to address the black box problem of ML models due to the need to have feature-level interpretability. SHAP is used to put a contribution value of each feature to a prediction. Positive SHAP values indicate features that push the prediction toward fraud, whereas the SHAP values that are negative signify the non-fraud influence.

VII. RESULTS

The ROC curve indicates a good trade-off between the True Positive Rate and False Positive rate, which confirms model performance. SHAP explainability plots demonstrate such important characteristics as velocity_6h, intended_balcon_amount, and creditrisk_score to be strong predictors of fraud. The Flask dashboard provides the opportunity to detect fraud in real-time as shown in the Fig:5. Fraud Detection Dashboard, with the option to upload the transaction information and immediately visualize the flagged transactions and SHAP-based explanations. The results of detected fraudulent transactions can also be downloaded, as demonstrated in Fig. 6: Results of Detected Fraud Transactions.

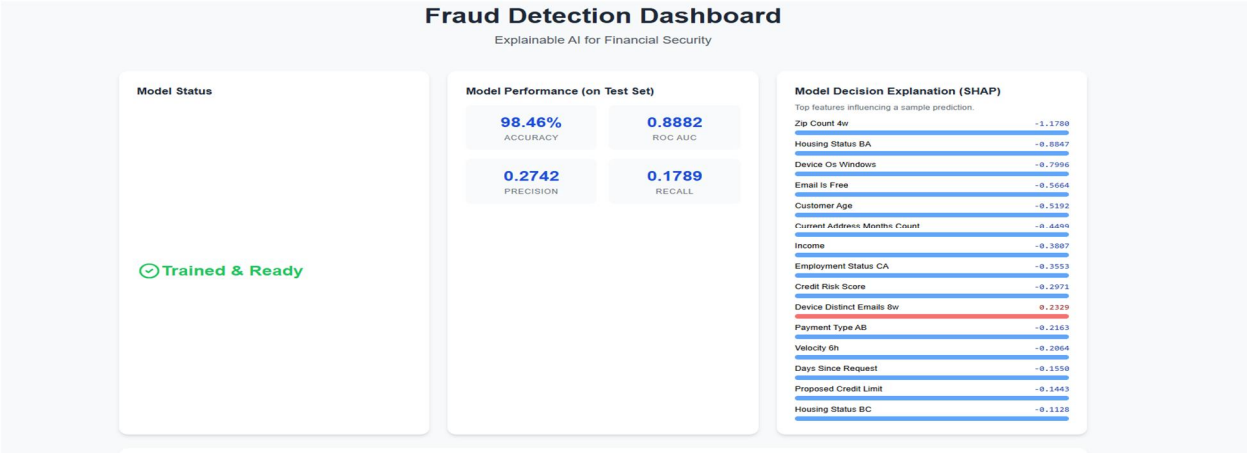
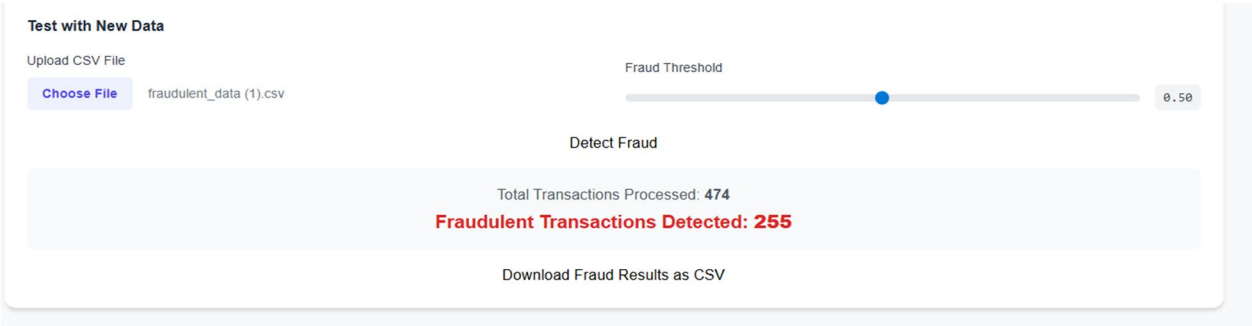


Fig 5:. Fraud Detection Dashboard



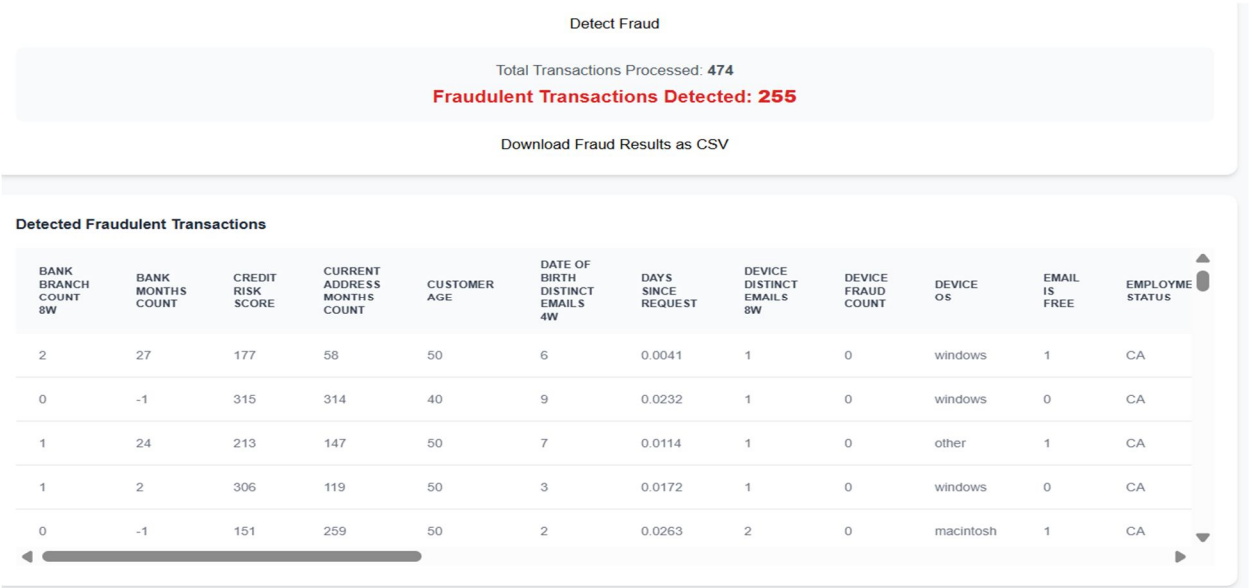


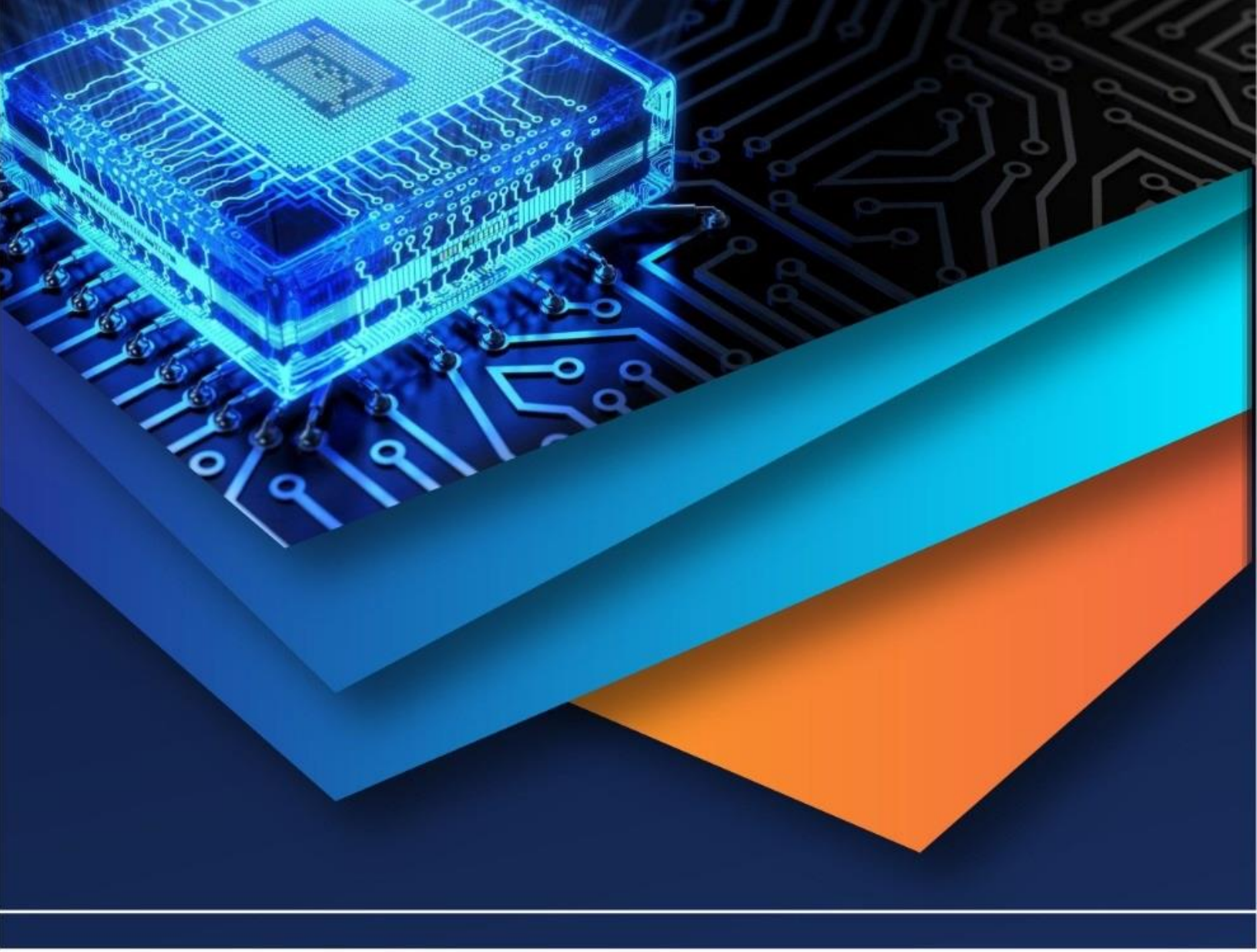
Fig 6: Results of Detected Fraud Transactions

VIII. CONCLUSION

This study demonstrates that explainability and privacy can be effectively integrated into AI-based fraud detection systems to achieve accurate, transparent, and ethically responsible outcomes. The proposed system successfully meets the key objectives of high accuracy, strong interpretability, and robust privacy preservation, while also maintaining scalability for deployment in real-world financial environments.

REFERENCES

[1] Tomisin Awosika, Raj Mani Shukla, And Bernardi Pranggono." Transparency And Privacy: The Role Of Explainable AI And Federated Learning In Financial Fraud Detection", 10.1109/ACCESS.2024.3394528.
[2] A. Pascual, K. Marchini, and S. Miller. (2017). 2017 Identity Fraud: Securing the Connected Life. Javelin. [Online]. Available: http://www. javelinstrategy.com/coverage-area/2017-identity-fraud.
[3] UKFinance. (2022). Annual Fraud Report 2022. [Online]. Available: https://www.ukfinance.org.uk/policy-and-guidance/reports-andpublications/annual-fraud-report-2022
[4] A. O. Adewumi and A. A. Akinyelu, ''A survey of machine-learning and nature-inspired based credit card fraud detection techniques,'' Int. J. Syst. Assurance Eng. Manage., vol. 8, no. 2, pp. 937–953, 2017.
[5] A. Srivastava, A. Kundu, S. Sural, and A. Majumdar, ''Credit card fraud detection using hidden Markov model,'' IEEE Trans. Depend. Sec. Comput., vol. 5, no. 1, pp. 37–48, Jan. 2008
[6] The Nilson Report. (Oct. 2016). [Online]. Available: https://www.nilson report.com/upload/content_promo/The_Nilson_Report_10-17-2016.pdf
[7] S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, ''Data mining for credit card fraud: A comparative study,'' Decis. Support Syst., vol. 50, no. 3, pp. 602–613, Feb. 2011.
[8] R. J. Bolton and D. J. Hand, ''Statistical fraud detection: A review,'' Stat. Sci., vol. 17, no. 3, pp. 235–255, Aug. 2002.
[9] S. Kamei and S. Taghipour, ''A comparison study of centralized and decentralized federated learning approaches utilizing the transformer architecture for estimating remaining useful life,'' Rel. Eng. Syst. Saf., vol. 233, May 2023, Art. no. 109130
[10] A. Pumsirirat and L. Yan, ''Credit card fraud detection using deep learning based on auto-encoder and restricted Boltzmann machine,'' Int. J. Adv. Comput. Sci. Appl., vol. 9, no. 1, 2018..
[11] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, ''SMOTE: Synthetic minority over-sampling technique,'' J. Artif. Intell. Res., vol. 16, pp. 321–357, Jun. 2002.
[12] Dingling Ge, Shunyu Chang, " Credit Card Fraud Detection Using Lightgbm Model". Rel..,2020 International Conference on E-Commerce and Internet Technology(ECIT).
[13] Tomisin Awosika, Raj Mani Shukla, And Bernardi Pranggono." Transparency And Privacy: The Role Of Explainable AI And Federated Learning In Financial Fraud Detection"
[14] Kuldeep Randhawa, Chu Kiong Loo (Senior Member, Ieee), Manjeevan Seera 2,3 (Senior Member, Ieee), Chee Peng Lim4, And Asoke K. Nandi5,6 (Fellow, Ieee) "Credit Card Fraud Detection Using Adaboost And Majority Voting"
[15] Sahil Somaji Kamble, Suyog Sudhir Pawar, Tejas Babasaheb Veer, Digambar M. Padulkar- "Fraud Detection using Quantum Machine Learning (QML)"
[16] Seunghyeok Oh, †Jaeho Choi, and ◦Joongheon Kim "A Tutorial on Quantum Convolutional Neural Networks (QCNN)"
[17] Siddhartha Bhattacharyya, Sanjeev Jha,1, Kurian Tharakunnel, J. Christopher Westland "Data mining for credit card fraud: A comparative study"
[18] S. Jesus, J. Pombal, D. Alves, A. Cruz, P. Saleiro, R. P. Ribeiro, J. Gama, and P. Bizarro, ''Turning the tables: Biased, imbalanced, dynamic tabular datasets for ML evaluation,'' in Proc. Adv. Neural Inf. Process. Syst., 2022.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY