Power System Technology
ISSN:1000-3673

Received: 06-11-2024 *Revised: 15-12-2024* *Accepted: 05-01-2025*

Federated Learning and Explainable AI: A Transparent and Secure Approach to Fraud Detection

Mrs. C. Leelavathi[1] K. Malathi[2] C. Nithya[3] A.Brinda Gowd[4] K.Gajendra[5]

[1,2,3,4,5] Department of Computer Science and Engineering & Business Systems Rajeev Gandhi Memorial College of Engineering & Technology (Autonomous) Nandyal, India.

[1]leelavathicse@rgmcet.edu.in [2]malathikondasani1813@gmail.com
[3]challanithya244@gmail.com [4]avulabrinda2003@gmail.com
[5]gajendrabunnygps143@gmail.com

Abstract:-

In the realm of financial fraud detection, the integration of Explainable AI (XAI) and Federated Learning (FL) is a new way of boosting transparency and privacy protection. The project uses the Paysim1 dataset from Kaggle to determine the authenticity of transactions and detect fraudulent transactions with greater accuracy. The current system mainly uses Deep Neural Networks (DNN), Recurrent Neural Networks (RNN), and Stochastic Gradient Descent (SGD) for detecting fraud. Although these approaches provide valuable insights, they tend to be non-interpretable and need centralized processing of data, which is a privacy and transparency concern. Conversely, our method makes advantage of Gradient Boosting Machines (GBMs), Random Forests, and Decision Trees. These are chosen due to their stability, interpretability, and better performance in dealing with intricate datasets. In addition to this, Federated Learning is also used in order to promote privacy by permitting models to get trained on decentralized devices without actually sharing raw data. This way not only confidentiality of data gets preserved but there is also development of a collective learning environment through which fraud detection models can improve accuracy. This blend of Explainable AI and Federated Learning seeks to mitigate the pressing issue of transparent yet privacy- protected solutions in fraud detection in financial institutions.

Keywords*:* Financial Fraud Detection, Decision Tree, Random Forest, Gradient Boosting Machines, (GBMs), Deep Neural Networks (DNN), RNN, Stochastic Gradient Descent.

1. Introduction

Financial fraud is a growing problem that evolves with the modern digital era. As online transactions and electronic payment systems proliferate, so do the innovative ways to bypass those securities used by fraudsters. Traditional methods for fraud detection heavily depend upon the extraction of patterns from the data using machine learning models, e.g., Deep Neural Network (DNN), Recurrent Neural Network (RNN), etc. Although these models show

Power System Technology
ISSN:1000-3673
Received: 06-11-2024 Revised: 15-12-2024 Accepted: 05-01-2025

great power in predictive applications, they exhibit critical limitation in some aspects, including opacity of decision and dependence on centralized data processing environment. And because of these limitations, the challenges of trust, interpretability, and data "hygiene" are so significant that alternatives that meet these challenges must be explored.

With this, we have a powerful combination of Explainable AI (XAI) and Federated Learning (FL) that can redefine how financial fraud is detected, focusing on interpretability and privacy at the same time. Also, Explainable AI will give a better understanding of how the fraud detection models make the decisions, which will help financial institutions to understand why a certain transaction is classified as fraudulent. This transparency is not just critical for compliance with regulations but also for establishing trust in customers and creating effective strategies to prevent fraud. On the other hand, Federated Learning allows for training machine learning models on numerous decentralized data sources without revealing raw financial data and providing strong privacy guarantees. This method mitigates the risk of this class of access, as well as data breaches, by decentralizing data storage.

So in this spreadsheet we applied and compared many different fraud detection methods on the well-known Paysim1 dataset from Kaggle, which is a framework for the simulation of mobile money transactions. Where as deep learning-based efforts rely on very large amounts of data in a centralized fashion, we use Decision Trees, Random Forests and (GBM) Gradient Boosting Machines. We assume these kind of machine learning models as they are interpretable and provide a good accuracy on predicting the fraud transactions. Moreover, the inclusion of Federated Learning further enriches the design by ensuring that financial brokers are able to jointly find models without the necessity of sharing sensitive knowledge of their users with a multitude of individuals and work with stringent regulations.

The main purpose of this study is to create a fraud detection model, which achieves high accuracy, while improving the interpretability of results and maintaining data privacy. This work, therefore, proposes a novel approach that uses Explainable AI coupled with Federated Learning to enhance the fraud detection landscape within financial institutions while simultaneously meeting the demand for effective, explainable and privacy-preserving fraud detection mechanisms. This project can set a new standard for financial security, allowing banks and payment processers to detect wrongdoing with greater accuracy, without compromising data security.

2. LITERATURE SURVEY

Rajesh et al. (2023) introduced a new solution capable of addressing the critical security, privacy, and performance challenges posed by federated transfer learning in the context of network intrusion detection in the IIoT environment. This work works a approach to reveal the design of deep neural network for bonded application where multiple client devices performs

Power System Technology
ISSN:1000-3673
Received: 06-11-2024 Revised: 15-12-2024 Accepted: 05-01-2025

the model training and blend via knowledge transfer technique to get ideal in learning perspective. This enables to jointly train without sharing sensitive data, preserving privacy and improving the accuracy of detection. The matter offloads computational tasks from clients or central servers, enhancing intrusion detection in the mobile and cloud domain. Thus, these findings highlight the potential that federated learning frameworks hold in safeguarding IIoT networks, while enabling effective anomaly detection performance in the process without degrading the confidentiality of data that is transmitted over wireless communication links.

Van Driel (2019) argued that a literature review on financial fraud, scandals, and regulatory responses, and a more comprehensive theoretical framework were needed, and accordingly incorporated historical and contemporary cases across all sectors in the analysis . You are skilled in identifying and explaining the principles relevant to financial fraud and their implications, and you can differentiate between ethical and unethical business behaviours. Additionally, it assesses the development of legal mechanisms and the impact of preventative measures in reducing fraudulent behaviours. The study underscores the need for adaptive regulatory policies by synthesizing insights across several domains, emphasizing the interconnected nature of financial fraud. The major takeaways are the need for ongoing regulatory reforms and early warning monitoring frameworks to protect financial systems from emerging threats and vulnerabilities.

Kamei, T., & Taghipour, K. (2023). Research: Centralized vs decentralized federated learning to estimate remaining useful life using transformer architecture: A comparative study — The study explores both architectures in the context of the predictive maintenance pipeline while discussing how both architectures help achieve improved model performance, communication efficiency, and data privacy. The findings show that, while centralized federated learning converges faster to a common model by using pooled global information, decentralised methods further scale-up by reducing the single point of failure through removing the dependence on a central server. The proposed methods boost the prediction performance based on the transformer architecture, indicating the feasibility of applying federated learning to industrial prognostics and health management. The results thus improve our understanding of federated learning frameworks as well as better align their systematical parameters to be more reflective of use in real-world applications pertinent to reliability engineering and system safety.

Comparative study on some machine learning techniques in credit card fraud detection L.

K. Tharakunnel, J. C. Westland. In this article authors compare Support Vector Machines (SVM), Random Forests and Logistic Regression on transaction data from the real world both for fraud detection and for the purpose of controlling and prosecuting the fraudulent act. The paper reviews how well these machine learning models were able to detect fraudulent transactions, with a low level of false positives. They also talk about the corner cases in which

Power System Technology
ISSN:1000-3673
Received: 06-11-2024 Revised: 15-12-2024 Accepted: 05-01-2025

such features could be employed in real world, as in real time fraud detection systems where the balance lies between being accurate, low computational cost and interpretable model. This could be considered a contribution which is part of these emerging fields of financial security with a focus on credit card fraud detection and it would discuss the advantages as well as their drawbacks.

3. METHODOLOGY

Proposed Work

In order to overcome the drawbacks in current methods, the proposed system to detect financial fraud makes use of Explainable AI (XAI) and Federated Learning (FL). It relies on algorithms like Random Forests, Decision Trees, and Gradient Boosting Machines (GBMs), that had been chosen for their simplicity, effectiveness and clarity for identifying fraud. Models trained on distributed data using federated learning help to prevent centralising private financial information. XAI techniques make the system more transparent by allowing clear, explainable descriptions of fraud detection decision-making. The system provides numerous advantages, including higher interpretability in the form of more transparent model decisions, higher privacy due to decentralized training of data, reduced computational complexity in comparison with deep learning models, and greater adaptability in which the system can adapt to emerging fraud trends with minimal retraining.

System Architecture

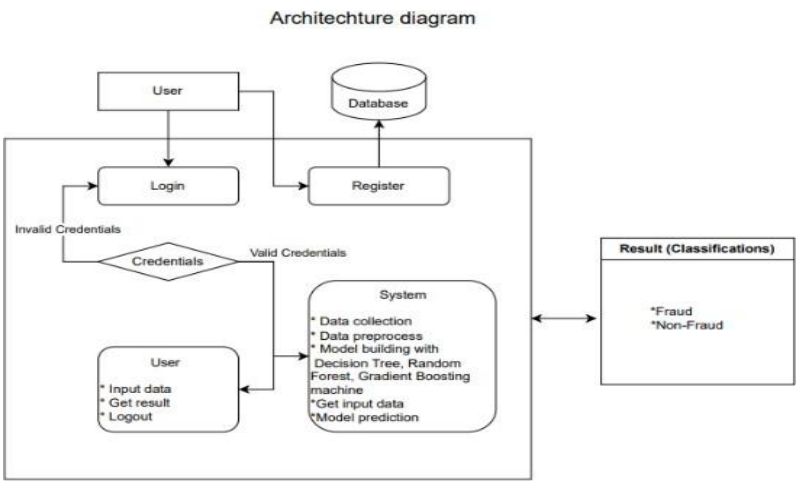

Fig 1 System Architecture

a) User Interaction

This was the initial flow that users will go through to register or log in with their credentials. New users register to set up an account. Current users would log in and enter the data to be used for fraud detection, view the results, and log out when done.

Power System Technology
ISSN:1000-3673
Received: 06-11-2024 Revised: 15-12-2024 Accepted: 05-01-2025

b) Credential Validation

The system validates user credentials to ensure security. The user is not allowed access if their credentials are invalid. The user is given access to the system's fraud detection features if they are legitimate.

c) Data Collection

The PaySim1 dataset provides simulated mobile financial transactions derived from actual data, readily available on Kaggle. Designed to research fraud detection within financial services, it portrays an assortment of transaction kinds like cash-ins, cash- outs, payments, transfers and account debits. Each event in the extensive set is designated as either fraudulent or legitimate, rendering the information appropriate for machine learning models requiring supervision. Crucial characteristics involved transaction categories, amounts, account balances prior to and following actions and whether deceit was included. PaySim1 is extensively applied to gauge the performance of algorithms meant to find fraud and offers a realistic benchmark for evolving and testing machine learning models intended to uncover financial deception. The diversity within the complex catalog provides researchers and analysts alike with an exceptionally robust resource with which to advance protections.

d) Data Preprocessing

This involves cleaning and preparing the data before you actually get into predictive modeling. It involves many steps such as managing missing data by filling in values or removing incomplete records to avoid bias or error in the model. This step corrects inconsistencies like those from duplicate entries, formats and outliers that might reduce model performance interestingly. We are also performing feature extraction in order to choose or develop relevant features from raw data that can enhance the accuracy and interpretability of the model. In summary, proper preprocessing improves the dataset and makes it more robust for creating reliable and preformat predictive models.

e) Algorithms Decision Trees

The decision tree is a non-parametric supervised learning technique . It is a hierarchical tree with a root node, branches, internal nodes, and leaf nodes. Numerous traditional machine learning techniques, such as Random Forests, Bagging, and Boosted Decision Trees, are based on decision trees. With each leaf node (terminal node) carrying a class label, each inside node signifying a test on an attribute (basically a condition), and each branch expressing the test's outcome, he came up with the idea to display data as a tree.

Random Forest

A supervised machine learning method for classification and regression problems is called random forest. It makes use of ensemble learning, which combines several classifiers to solve

Power System Technology
ISSN:1000-3673
Received: 06-11-2024 Revised: 15-12-2024 Accepted: 05-01-2025

challenging problems. The random forest approach consists of many decision trees. A 'forest' is created by bagging or bootstrap aggregating the random forest method. An ensemble meta-algorithm called bagging is frequently used to increase machine learning algorithms' accuracy.

In the random forest algorithm, the outcome is decided according to the voting of decision trees. calculating the average production of the forest's different trees. The more trees, the more accurate the outcome. A decision tree algorithm suffers from such drawbacks which a random forest overcomes. It increases the accuracy and avoids the over fitting of input datasets. It requires no tuning, and has predictive power in packages.

XGBoost

A scalable and efficient gradient-boosted decision tree (GBTD) machine learning library implementation is called XGBoost Extreme Gradient Boosting, or simply XGBoost. With parallel tree boosting, it is the quickest machine learning library for ranking, classification, and regression.

To understand what XGBoost is doing, it is helpful to first understand the machine learning concepts and techniques on which it is based: supervised machine learning, decision trees, ensemble learning, and gradient boosting.

An algorithm is used to train a model in supervised machine learning in order to find patterns in a dataset that contains features and labels . The labels are then predicted by the model using a fresh dataset's features.

Gradient Boosting Classifier

The most significantly learning idea we introduced for last twenty year was boosting algorithm. Gradient boosting is one type of supervised machine learning technique that can be used for both classification and regression problems.

This ensemble technique, which combines bagging and boosting, uses multiple weak learners to create a powerful model for regression and classification. This method, known as gradient boosting, is based on the idea that the best feasible future model, when added to the existing ones, minimizes the total prediction errors. That is, the target outputs of the previous models are used as inputs to the next models in a way that minimizes the errors. Another boosting algorithm is this one.

4. EXPERIMENTAL RESULTS

Accuracy: Accuracy cannot be applied in scenarios of imbalanced cases and it measures how many predictions were correct overall.

TP + TN

Power System Technology
ISSN:1000-3673
Received: 06-11-2024 Revised: 15-12-2024 Accepted: 05-01-2025

$$Accuracy = \frac{}{TP + FP + TN + FN} \quad (1)$$

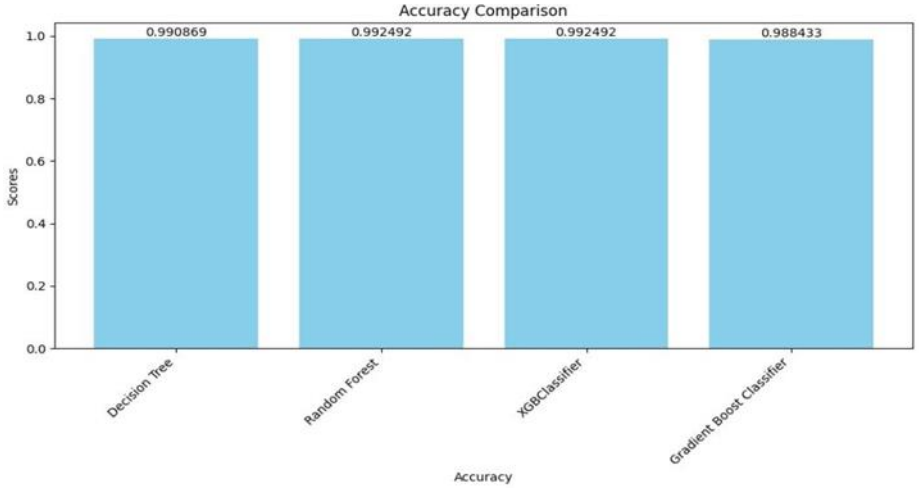

Fig 2 Accuracy Comparison GrapPrecision: It is a value of TP/(TP+FP) which describes the correctly predicted positive cases to the total number of positive instances predicted.

$$Precision = \frac{\text{True Positive}}{\text{True Positive +False Positive}} \quad (2)$$

Power System Technology
ISSN:1000-3673
Received: 06-11-2024 Revised: 15-12-2024 Accepted: 05-01-2025

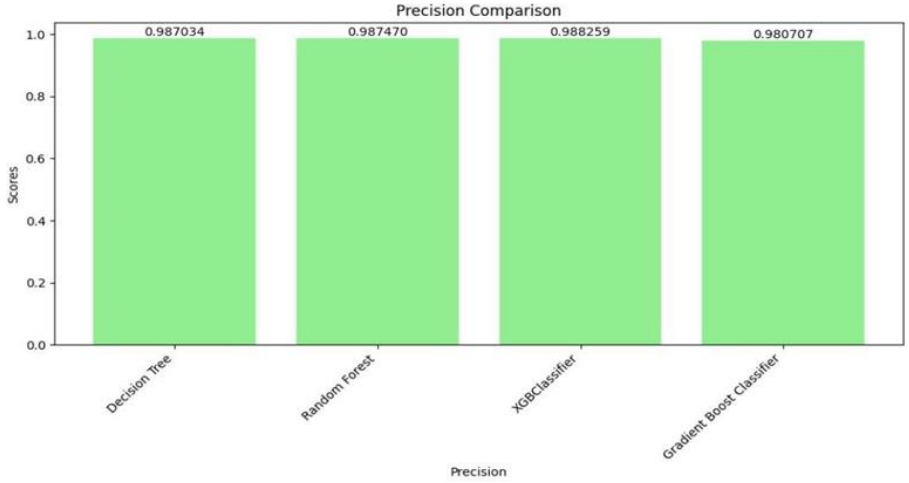

Fig 3 Precision Comparison Graph

Recall: Shows the proportion of actual positives that are correctly identified as such.

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

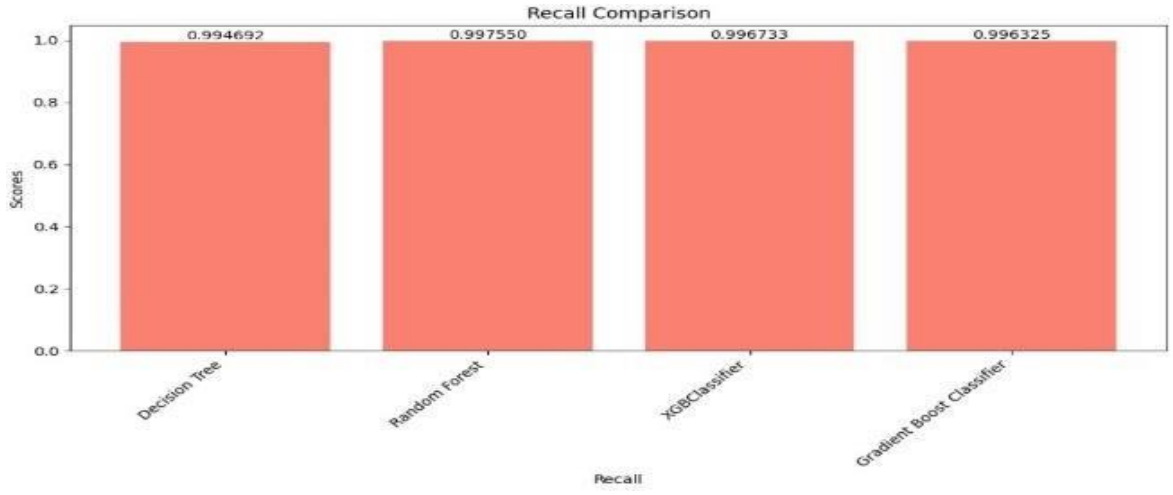

Fig 4 Recall Comparison Graph

F1-Score: To do this, we go back to taking the harmonic mean of Precision and Recall which is a balanced metric.

$$F1\ Score = 2 * \frac{Recall\ X\ Precision}{Recall + Precision} * 100 \quad (4)$$

Power System Technology
ISSN:1000-3673
Received: 06-11-2024 *Revised: 15-12-2024* *Accepted: 05-01-2025*

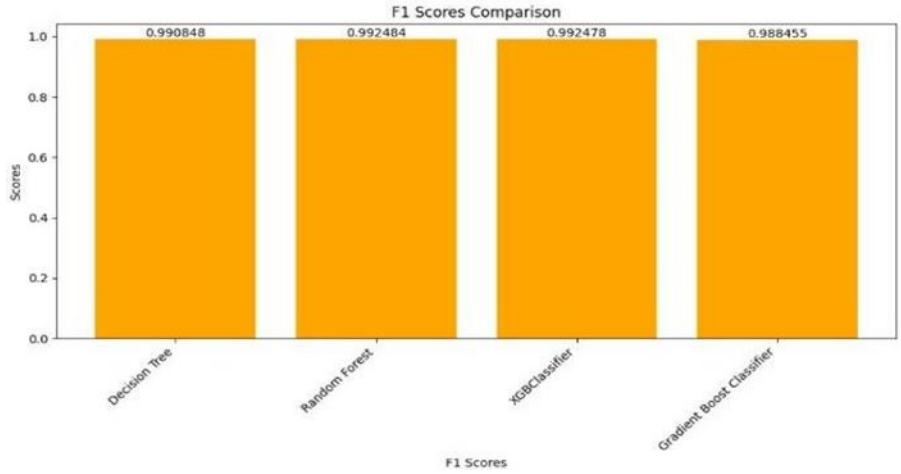

Fig 5 F1 Score Comparison Graph

ML Model	Accuracy	Precision	Recall	F1-Score
Decision Trees	0.990	0.987	0.994	0.990
Random Trees	0.992	0.987	0.997	0.992
XGBoost	0.992	0.988	0.996	0.992
Gradient Boosting Classifier	0.988	0.980	0.996	0.988

5. CONCLUSION

In short, the interplay between Explainable AI(XAI) for heterogeneous perspectives and Federated Learning (FL) to maintain privacy, tackles fundamental problems in most applications where privacy and interpretability are key factors. The core models in this project - Decision Trees, Random Forests, and Gradient Boosting Machines (GBM) — improve upon the transparency and performance of traditional Deep Neural Networks (DNN) and Recurrent Neural Networks (RNN) models which have non-informative black box characteristics. Federated Learning enhances privacy by enabling models to be trained in a decentralized way, which allows financial data to stay private. This fusion of two methods not only leads to a more precise fraud detection process, but also a transparent, privacy-preserving framework for optimization. Therefore, the proposed hybrid framework integrates multi-layered XAI and FL architecture to ensure privacy-preserving solutions that report how to comply under the high- performance environment. This shift in paradigm offers the way towards a future of more secure and accountable financial systems.

Power System Technology
ISSN:1000-3673
Received: 06-11-2024 Revised: 15-12-2024 Accepted: 05-01-2025

6. FUTURE SCOPE

XAI and FL have the potential to advance financial fraud detection by enabling greater accountability, regulatory compliance, and personalization while reducing the risk of adversarial attacks or intrusion of sensitive information. For example, that could include the adoption of hybrid models combining Decision Trees and Random Forests interpretability with generated patterns from deep learning architectures like Transformer models. Other strategies — such as creating more complex cross-device communication within Federated Learning frameworks not only can add to the increase in global model accuracy, but overall lead to being able to still keep efficiency in model training. Thus, applying differential privacy technology in FL can further improve the security of the training process, protecting sensitive data. Finally, systems that detect and prevent fraud in real-time through streaming data and recursive learning algorithms may also be implemented to provide prompt responses and significantly lower the risk of fraudulent activity and enhance the overall robustness of the system.

REFERENCES

[1] L. T. Rajesh, T. Das, R. M. Shukla, and S. Sengupta, ‘‘Give and take: Federated transfer learning for industrial IoT network intrusion detection,’’ 2023, arXiv:2310.07354.

[2] A. Abdallah, M. A. Maarof, and A. Zainal, ‘‘Fraud detection system: A survey,’’ J. Netw. Comput. Appl., vol. 68, pp. 90–113, Jun. 2016.

[3] A. Pascual, K. Marchini, and S. Miller. (2017). 2017 Identity Fraud: Securing the Connected Life. Javelin. [Online].

[4] 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.

[5] 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.

[6] S. Vyas, A. N. Patra, and R. M. Shukla, ‘‘Histopathological image classification and vulnerability analysis using federated learning,’’ 2023, arXiv:2306.05980.

[7] R. J. Bolton and D. J. Hand, ‘‘Statistical fraud detection: A review,’’ Stat. Sci., vol. 17, no. 3, pp. 235–255, Aug. 2002.

[8] H. van Driel, ‘‘Financial fraud, scandals, and regulation: A conceptual framework and literature review,’’ Bus. Hist., vol. 61, no. 8, pp. 1259–1299, Nov. 2019.

[9] G. M. Trompeter, T. D. Carpenter, N. Desai, K. L. Jones, and R. A. Riley, ‘‘A synthesis of

Power System Technology
ISSN:1000-3673
Received: 06-11-2024 Revised: 15-12-2024 Accepted: 05-01-2025

fraud-related research,'' AUDITING, A J. Pract. Theory, vol. 32, no. Supplement 1, pp. 287–321, May 2013.

[10] P. Raghavan and N. E. Gayar, ''Fraud detection using machine learning and deep learning,'' in Proc. Int. Conf. Comput. Intell. Knowl. Economy (ICCIKE), Dec. 2019, pp. 334–339.

[11] M. Zareapoor and P. Shamsonmoali, "Application of credit card fraud detection: Based on bagging ensemble classifier, "Proc, Comput Sci., vol. 48, pp. 679-685, 2015