International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(15s) 2026 pp. 378-393
https://doi.org/10.70917/ijcisim-2026-4409
Article

Artificial Intelligence-Based Empirical Study for Fraud Detection in Digital Financial Transactions

Manoj P. K.[1*], Akshay Dilip Homkar[2], Raj Kumar Singh[3], Naresh Konduri[4], L Devi[5], Atowar ul Islam[6]

[1*]Department of Applied Economics, Cochin University of Science and Technology (CUSAT), Kochi, Kerala - 682 022, India, Orcid Id: 0000-0002-5710-1086, Email Id: manoj_p_k2004@yahoo.co.in
[2]Department of Computer Science and Engineering, Specialization in Cyber Security, Web Development, Digital Forensics, Network and System Administration, Kasegaon Education Society's Rajarambapu Institute of Technology, affiliated to Shivaji University, Sakharale, MS-415414, India, Orcid Id: 0009-0007-9086-6869, Email Id: akshayhomkar@gmail.com
[3]Department of Management and Commerce, Specialization in Marketing and Entrepreneurship, School of Management Sciences, Varanasi, UP-221011, India, Email Id: rksingh@smsvaranasi.com
[4]Department of CSE (IoT & CS including Block Chain Technology), Specialization in Computer Science and Engineering, Sasi Institute of Technology & Engineering (Autonomous), Directorate of Research & Development, Jawaharlal Nehru Technological University Kakinada (JNTUK), Orcid Id: 0009-0000-1363-8773, Email Id: naresh.konduri@gmail.com
[5]Department of Commerce, Specialization in Marketing, Krupanidhi College of Commerce and Management, Bengaluru-560034, India, Orcid Id: 0000-0001-9368-1308, Email Id: ldevi933@gmail.com
[6]Department of Computer Science, University of Science and Technology Meghalaya, Ri- bhoi, Meghalaya, India, Orcid Id: 0000-0001-5345-0556, Email Id: atowar91626@gmail.com

Abstract: Fraud in digital financial transactions presents a critical challenge due to increasing transaction volumes and complex behavioral patterns, particularly under highly imbalanced conditions where fraudulent cases are extremely rare. This study aims to develop artificial intelligence-based models for fraud detection, compare multiple machine learning algorithms, and evaluate their effectiveness using robust performance metrics. A large-scale transaction dataset comprising over 1.32 million records with a fraud rate of approximately 0.13% was analyzed. Data preprocessing, feature engineering, and leakage-aware modelling were performed, and a chronological split (70–15–15) was adopted for training, validation, and testing. Logistic Regression, Random Forest, and XGBoost models were implemented, with threshold optimization conducted on the validation set. Performance was evaluated using accuracy, precision, recall, F1-score, ROC-AUC, and PR-AUC. The results indicate that XGBoost achieved the best performance with precision = 0.99, recall = 0.96, F1-score = 0.98, ROC-AUC = 0.99, and PR-AUC = 0.97, outperforming Random Forest and Logistic Regression. Feature importance analysis revealed that transaction-to-balance ratios, sender balance characteristics, and behavioral indicators were the most influential predictors. Overall, the findings demonstrate that ensemble-based artificial intelligence models, particularly XGBoost, provide accurate and reliable fraud detection under imbalanced conditions, supporting their applicability in real-world financial security systems.

Keywords: Artificial Intelligence, Fraud Detection, XGBoost, Imbalanced Data, Financial Transactions

1. Introduction

The rapid technological advancement in the financial world, with the integration of various digital technologies, has revolutionized the global financial system, creating an interconnected financial world with the help of data. The integration of big data, AI, and various digital platforms has significantly improved the efficiency of financial services with the help of real-time transactions and decision-making processes [1]. This technological advancement has also helped create value co-creation in the financial world, especially in the fintech world [2]. Nonetheless, in parallel with these developments, there are new risks, such as those associated with financial fraud and cyber risks, which have emerged in line with the increased level of digitization in financial transactions. The emergence of digital financial systems, such as peer-to-peer lending and other transactional systems, has led to complex regulatory and security issues, which are associated with these systems due to their decentralized nature and their potential for fraud, as

discussed in [3]. Additionally, with increased reliance on digital infrastructure, there are growing concerns over issues of cybersecurity and protection of critical financial systems, which calls for robust and adaptive security measures [4]. In this regard, financial institutions are increasingly using fintech solutions to enhance their operational efficiency and minimize risks associated with digital transactions [5].

In the same way, the development of financial systems can also be discussed in the framework of the technical shifts when digital innovation brings structural shifts in the economical systems and institutions [6]. The access to high volumes of financial information has also been enabling the higher uptake of data-oriented systems so that organizations could use predictive analytics and machine-learning tools in decision-making [7]. Nonetheless, fraud, money laundering, and unauthorized transactions remain an obstacle in the most digitized systems [8] of fraud detection systems in developing financial ecosystems.

The conventional rule-based fraud detection systems are not effective in dealing with the complexity of financial fraud. The reason is that such systems are rule-based, which restricts the flexibility of the system to evolving fraud patterns and advanced attack schemes [9]. This calls for the need to develop intelligent systems that are capable of processing large amounts of transactional data to detect unusual patterns in real time. The use of artificial intelligence and machine learning methods has been proposed as the way to overcome the shortcomings of traditional systems since it provides better predictive opportunities [10].

Significant developments in AI-based financial analytics have revealed tremendous possibilities in terms of improving fraud detection, credit risk analysis, and regulatory compliance. The high potential of Hybrid and ensemble learning methods in dealing with complex and imbalanced financial data sets as it is encountered in different studies has been revealed [11]. In addition, research on the systematic reviews of AI-related applications in financial compliance has indicated that machine learning-based models could have more potential in identifying fraudulent acts in different financial contexts, including cryptocurrency and digital banking systems[12].

Despite the abovementioned developments, there are still certain issues when it comes to the implementation of AI-based fraud detection system. The first obstacle is the unbalanced nature of financial transaction datasets as the percentage of bad transactions is much lower than that of the good ones. This ratio tends to impact the work of the system because the percentage of fraudulent operations is not high in many cases compared with the normal ones. Moreover, the problem of time validity, information leakage, and use of the right performance measures are also worth considering in the introduction of the effective fraud detection systems.

To this end, the present research study aims to assess the effectiveness of artificial intelligence-based models in the detection of fraud in online financial transactions on a large scale dataset of transactions. In this respect, this research study will frame this problem as a binary classification problem and will compare the effectiveness of a simple Logistic Regression model with a more sophisticated XGBoost model, and in this respect, will apply methods like feature engineering, temporal splitting, and threshold optimization to make this a more realistic fraud detection system.

This research has three-fold contributions. To begin with, this study presents empirical findings of traditional and advanced machine learning models in an environment with unbalanced financial data. Second, the study highlighted the significance of threshold optimization and precision-recall-based assessment measures in fraud detection activities. Lastly, the study provides information on the ways of developing scalable and effective AI-based fraud detection systems in a contemporary financial environment.

Objective of the study

1. To develop robust artificial intelligence-based models for accurate detection of fraudulent transactions in digital financial systems.
2. To compare the performance of different machine learning algorithms in order to identify the most effective model for fraud detection.
3. To evaluate model performance using key metrics such as precision, recall, F1-score, and their effectiveness under imbalanced data conditions.

2. Literature Review

Odofin et al. [13] studied on artificial intelligence in telecommunications data infrastructure in anomaly detection and revenue recovery. The authors were able to conclude that artificial intelligence systems can effectively

detect anomalies in a complex data setting and can contribute to revenue recovery by identifying irregular patterns. The research shows that artificial intelligence has great value in managing large-scale systems related to transactions.

Randhawa et al. [14] focused on credit card fraud detection using AdaBoost and majority voting techniques. The researchers were able to show how ensemble learning can improve the accuracy of fraud detection and classification. The research is significant in showing the power of machine learning in detecting fraudulent activities in finance.

Jean-Philippe [15] studied the development of computer network defense technology with the application of machine learning and artificial intelligence. The research highlighted the significance of the application of AI-based technology in enhancing security by efficiently identifying threats. It is relevant to the field because digital financial fraud is related to cybersecurity.

Byambaa et al. [16] empirically examined investigation into the adoption of artificial intelligence in banking services in Mongolia has been carried out. It is evident that the adoption of AI enhances the efficiency and innovation of banking services. It is clear that the adoption of AI is a significant aspect of the current transformation in banking.

Abulsaoud Ahmed Younis and Adel [17] studied the strategy of artificial intelligence, creativity-oriented human resource management, and knowledge-sharing quality in AI-powered organizations are discussed in this paper. The researchers found that a good strategy in artificial intelligence improves individual and organizational effectiveness with the support of creative human resource management and quality knowledge sharing.

Goswami [18] studied the advancement in threat detection through artificial intelligence and machine learning-based cybersecurity audits has been observed. The study has shown that artificial intelligence can greatly enhance cybersecurity audits in an efficient manner to identify potential risks and threats. This has highlighted the role of artificial intelligence in protecting digital systems from malicious activities.

Singh and Sinha [19] studied the way in which artificial intelligence is expected to impact non-banking financial companies. The authors found that artificial intelligence can increase efficiency, provide automated services, increase customer engagement, and monitor fraud. The authors' research revealed that artificial intelligence has significant transformative potential in the non-banking finance sector.

Akter and Kudapa [20] A comparative analysis was conducted on AI-integrated business intelligence dashboards for real-time decision support systems in operations. The research proved that AI-based business intelligence dashboards improve the speed of analytics, quality of decisions, and responsiveness of operations. This emphasizes the value of AI in real-time decision-making.

Kuswara et al. [21] in this study, the role of Artificial Intelligence in the financial reporting process and its impact on the effectiveness and efficiency of the process were investigated. It was concluded that the quality, speed, and accuracy of financial reporting processes can be improved using Artificial Intelligence.

Edozie et al. [22] studied Improvements in the application of artificial intelligence for anomaly detection in telecommunications systems. The authors conclude that anomaly detection systems based on artificial intelligence are very effective in identifying abnormal activities in large-scale and dynamic telecommunications systems. The authors' research shows the general applicability of artificial intelligence for anomaly detection purposes.

Ali et al. [23] The research aimed to examine the role of artificial intelligence in the automation of the banking services industry in Jordan. The authors established that artificial intelligence helps improve the automation of the banking industry. The research highlights the significance of artificial intelligence in the banking industry.

Bou Reslan and Jabbour Al Maalouf [24] assessed, evident in the transformative effect that the adoption of AI can have on efficiency, fraud detection, and the skills of individuals working in accounting practices. It is evident from the study that the adoption of AI increases the efficiency and fraud detection capacity, while at the same time transforming the skills required by individuals working in the field.

Islam et al. [25] study examined the practice of artificial intelligence to enhance business performance in Bangladesh. The researchers found that the adoption of artificial intelligence helps businesses innovate, enhance productivity, and strategically grow. The study suggests that artificial intelligence improves various aspects of businesses.

Dubey et al. [26] studied the unified transformer-BDI architecture was examined for its applicability in detecting financial fraud, using distributed knowledge transfer on different datasets. The authors found that advanced

AI architectures can improve the performance of detecting financial fraud in heterogeneous financial datasets. This further emphasizes the significance of using advanced AI models in detecting financial fraud.

Raihan [27] research focused on the optimization of financial technology using artificial intelligence with the aim of achieving sustainable development goals. The researcher concluded that artificial intelligence in financial technology can improve financial efficiency, innovation, and sustainability. The search shows that AI in finance can help with both economic and developmental goals.

3. Methodology

3.1 Data Sources and Integration

The current study utilised three primary datasets: accounts, transactions, and alerts, which were systematically integrated to create a cohesive transaction-level analytical dataset [28]. The integration process entailed the process of connecting transaction records and account related information and the alert labels on the use of unique identifiers on the accounts and transactions. The resulting dataset (after this merging process) consisted of 1,323,234 transactions that occurred on 10,000 accounts with 1,719 transactions classified as fraudulent according to the alert linkages. This built in architecture allowed extensive modelling of financial behavior at the transaction level and consistency between transactional and alert based fraud indicators.

3.2 Data Preprocessing and Feature Engineering

3.2.1 Feature Construction and Selection

Following the integration, 31 variables were initially developed and this included the transaction-level variables, the account-level financial variables, behavioral variables, and the derived variables. These were the transaction amount, receivers and senders balance, time based variables and behavior identifiers. Variables obtained directly based on the alert labels or containing information related to the target were not entered into the modelling process in order to guarantee methodological rigour and to prevent data leakage. Consequently, a sub-set of 22 leakage-free predictive features was selected to be used in the development of models. These engineered variables were critical among others such as financial ratio variables as AMOUNT_TO_SENDER_BALANCE_RATIO and AMOUNT_TO_RECEIVER_BALance Ratio, balance-based variables as BALANCE-DIFF, log-transformed transaction amount (LOG_TX-AMOUNT) and temporal segmentation variables as TIME-BIN-24 and TIME-BIN-7. These characteristics were developed to represent transactional scale as well as behavioral dynamics of fraud detection.

3.2.2 Handling Class Imbalance

The dataset was highly skewed in terms of the distribution of the classes as fraudulent transactions comprised only 0.1299 percent of the entire observations. With this gross imbalance, no artificial resampling of any kind was done, to retain the natural distribution of fraud in natural financial systems. As an alternative, the modelling approach focused on the application of class imbalance-resistant evaluation metrics, including precision, recall, F1-score, and Precision-Recall Area Under the Curve (PR-AUC). Further, the optimization of the decision threshold was also added so as to enhance the sensitivity of classification without overinflating the false positives.

3.3 Temporal Data Partitioning

In order to guarantee methodological validity and capture real-life deployment conditions, the dataset was divided on a chronological splitting policy. Namely, the training set was assigned with earlier transactions, whereas the more recent transactions were assigned to validation and testing sets. The resulting partitions were 926,263 training, 198,485 validation and 198,486 tests. Notably, there were no changes in the levels of fraud in these subsets, which implies that the time-based division had no distributional drift due to the fixed class distribution. This strategy avoided any form of temporal leakage and made sure that future-like unobserved data was evaluated in the models.

3.4 Exploratory Data Analysis

3.4.1 Transaction Amount Analysis

Exploratory analysis has been done to investigate the distributional properties of transaction amounts in fraudulent and non-fraudulent classes. To reduce skew and allow comparison of classes, log-transformed histograms, and boxplots were used. This study identified specific differences in distributions, and it implies that the magnitude of transactions is an effective discriminative characteristic to detect fraud.

3.4.2 Temporal Pattern Analysis

Besides the characteristics at the transaction level, the temporal patterns of fraud were examined on both the fine-grained (24-bin) and coarse-grained (7-bin) time segmentation. Fraud rates have been calculated within these temporal bins in order to establish a pattern of fraud activity over time. It was shown that the occurrence of fraud was not evenly distributed, which showed that there were temporal dependencies. These were added to the modelling structure via engineered time characteristics.

3.5 Machine Learning Models

Three monitored machine learning models were adopted to deal with the task of fraud detection. A linear baseline model of performance under simple decision boundaries was performed using Logistic Regression. An ensemble learning algorithm, which was a bagging based algorithm called Random Forest, was used to identify non-linear relationships and interactions amongst features. Also, XGBoost, a gradient boosting model, was employed because it has a high capability of processing structured data and imbalanced classification issues. The combination of these models allowed making a comparative assessment of the linear and non-linear methods.

3.6 Model Training and Validation

All the models were trained on the chronologically defined training data and the hyperparameter tuning and threshold calibration operated on the validation set. The independent test set was used in the final performance evaluation to make sure that it was done without any bias. This training validation testing pipeline guaranteed that information leakage of the test data was not caused by model optimization.

3.7 Evaluation Metrics

Since the data was highly imbalanced, numerous measures of evaluation were also used to evaluate the performance of the models in an inclusive manner. These were accuracy, precision, recall, F1-score, ROC-AUC and PR-AUC. Although accuracy had been used as a general performance measure, more focus was given to precision and recall as they are relevant in detecting fraud. Particularly, the PR-AUC was regarded as an important metric, since it gives more informative assessment in those situations when the positive class is infrequent.

3.8 Threshold Optimization

To further improve the model performance, decision thresholds were optimized using validation data as opposed to using default threshold of 0.5. The aim was to obtain an ideal tradeoff between precision and recall, namely, maximising the F1-score in case of class imbalance. In the case of the best-performing model, which was XGBoost, a threshold value of 0.05 was determined to be the optimal threshold, indicating that in a situation with rare events, lower classification cutoffs are required.

3.9 Model Performance Analysis

3.9.1 ROC and Precision–Recall Analysis

ROC curves were used to measure model discrimination ability whereas Precision Recall curves were used to give a more trustworthy measure in the case of class imbalance. This two-fold review structure made sure that performance and classification effectiveness got the due evaluation.

3.9.2 Confusion Matrix Analysis

The confusion matrices helped in understanding the distribution of true positives, false positives, true negatives and false negatives of every model. This discussion helped to understand the error format and the working appropriateness of the models especially in reducing false alarms and maximizing the detection of fraud.

3.9.3 Probability Distribution Analysis

They tested the probability distributions that were predicted in order to determine the level of separation between fraudulent and non-fraudulent transactions. This discussion was useful in justifying the capacity of the model to generate discriminative and well-calibrated probability outputs.

3.10 Feature Importance and Model Interpretability

The best-performing XGBoost model was analyzed using the built-in importance metrics of the model to determine the most important features. This step revealed the strongest predictors that lead to decisions made by a model, such as variables of financial ratios, balance-based variables, and behavioral variables. Temporal variables were also added but they were found to have a lower importance. This interpretability examination was a good revelation of the background motivation of fraudulent actions.

3.11 Alert-Level Analysis

Besides transaction-level modelling, there was a descriptive analysis of the alert data. The alerts were classified as cycle and fan in and all the alerts related to verified fraudulent transactions. This validation was to make sure that there is consistency between machine learning labels and rule-based alert mechanisms, which makes the dataset used in the analysis valid.

4. Results

4.1 Dataset composition and class imbalance

The combined experimental data consisted of 1,323,234 computerized financial transactions, created between 10,000 accounts and connected with 1,719 alert records once the accounts, transactions and alerts files were merged. The combined transaction-level analytical data had 31 engineered variables, 22 of which were leakage-free predictive variables preserved to model after the rejection of target-related and alert-derived variables. In the transaction stream, there was a high imbalance in classes. Out of the 1,323,234 transactions, 1, 321,515 were not fraudulent and only 1,719 were fraudulent, with the total prevalence of fraud being 0.1299% (Table 1; Figure 1). This proves that the fraud detection issue in the current study is a very unbalanced binary classification problem.

A chronological segregation of the dataset was performed in order to maintain the integrity of time through creating training, validation and testing segments. The training set consisted of 926,263 transactions, validation set consisted of 198,485 and test set consisted of 198,486. The rates of fraud were very stable between the three partitions, as the values were 0.1261% in training, 0.1406% in validation, and 0.1370% in testing (Table 1; Figure 1). This allocation shows that the chronological division did not cause significant driftage between subsets because of the structure of rarity of the fraud class.

Table 1. Dataset distribution across chronological splits

Dataset	Non-Fraud	Fraud	Total	Fraud Rate
Full	1,321,515	1,719	1,323,234	0.1299%
Train	925,095	1,168	926,263	0.1261%
Validation	198,206	279	198,485	0.1406%
Test	198,214	272	198,486	0.1370%

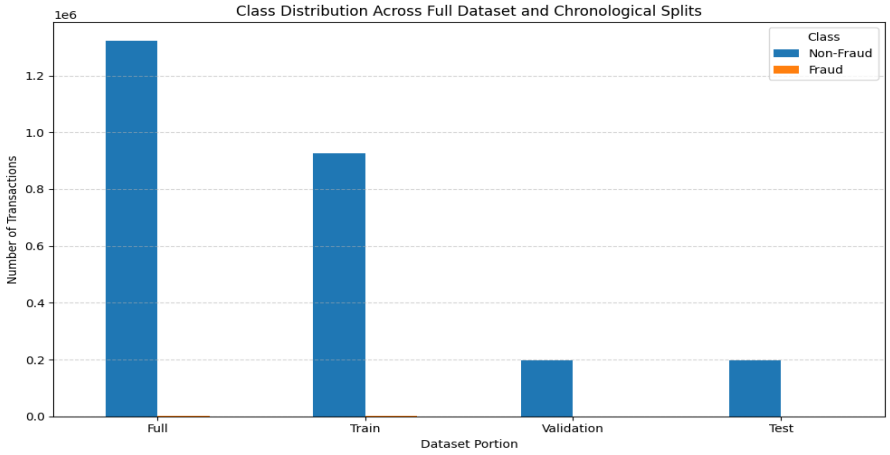

Figure 1. Class distribution across full dataset and chronological splits

4.2 Exploratory transaction and temporal fraud patterns

The mode of transaction values varied with the fraud and non-fraud classes as shown by log-transformed histogram and boxplot of the transaction values (Figure 2). The visual distinction indicates that the amount of transaction was a significant discriminative attribute, which was subsequently justified by the analysis of feature-importance. Time analysis also indicated that the presence of fraud was not entirely time-bin uniform (Figure 3). Across the 24 time bins, the fraud rate ranged from 0.1070% (Time Bin 2) to 0.1603% (Time Bin 23), with relatively elevated fraud rates also observed at Time Bin 14 (0.1542%) and Time Bin 6 (0.1522%) (Table 2; Figure 3a). The range of fraud rate over the 7 coarser time bins was 0.1235 percent to 0.1385 percent with the largest concentration of fraud in Time Bin 6 0.1385 percent (Table 2; Figure 3b). Such results suggest that there is a slight yet significant temporal change in fraudulent transaction behaviour.

Table 2. Fraud rate across temporal bins

Time Bin	Fraud Rate	Total Transactions	Fraud Count
24-bin minimum	0.1070%	58,854	63
24-bin maximum	0.1603%	53,636	86
7-bin minimum	0.1235%	185,358	229
7-bin maximum	0.1385%	185,524	257

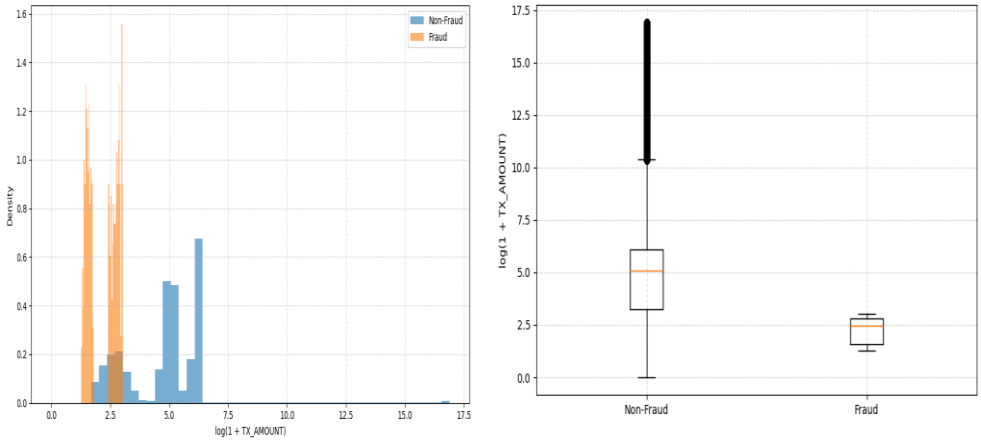

Figure 2. Log-transformed transactn amount distribution and boxplot by class

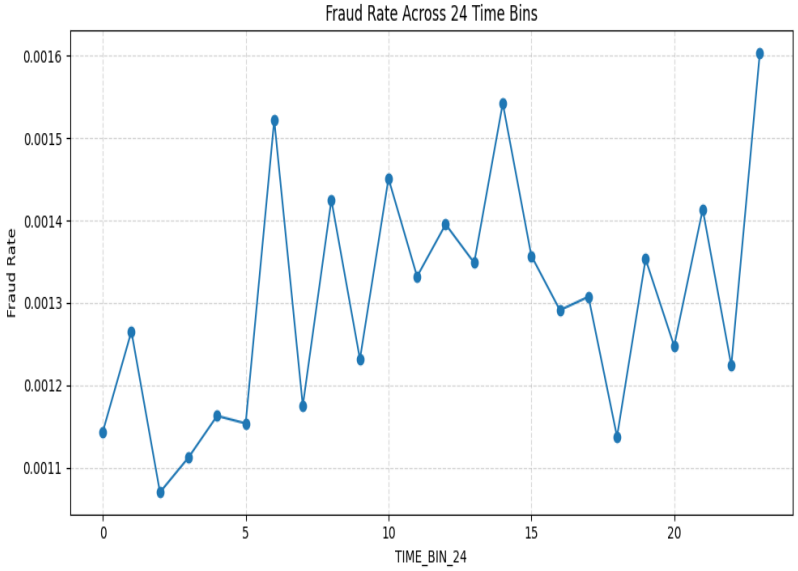

Figure 3a. Fraud rate across 24 time bins

Figure 3b. Fraud rate across 7 time bins

4.3 Comparative model performance

The three machine learning models considered in the study included Logistic Regression, Random Forest, and XGBoost, which were assessed in fraud detection. The results of their comparison in the temporarily held-out test set are summarised in Table 3 and shown graphically in Figures 4 through 6. XGBoost was the highest performing overall classifier with a decision threshold of 0.05, the highest accuracy of 99.9950, precision of 100.0, recall of 96.3235 and F1-score of 0.9813 (Table 3). It has found the highest ranking-based performance, with ROC-AUC of 0.9989 and PR-AUC of 0.9708 (Table 3; Figures 4 and 5). These findings imply extremely high discrimination in fraudulent and legitimate transactions in extreme classes imbalance.

The Random Forest model performed comparably, achieving 99.9929% accuracy, 99.2366% precision, 95.5882% recall, and an F1-score of 0.9738, with ROC-AUC = 0.9985 and PR-AUC = 0.9689 (Table 3). Though it was a little worse than XGBoost, its performance was also very competitive which implies that tree-based ensemble techniques were particularly favorable in the current fraud detection task. Logistic Regression, in turn, demonstrated significantly worse results. It has a high recall (88.6029%), but with a low precision (6.8641), its F1-score was significantly lower (0.1274) and PR-AUC was also low (0.0438) (Table 3). Such a difference suggests that the linear

model was far less applicable to the operational fraud detection in this dataset despite the significant percentage of fraudulent cases that it captures, as it generated too many false positives.

Table 3. Test-set performance comparison of the evaluated models

Model	Threshold	Accuracy	Precision	Recall	F1-Score	ROC-AUC	PR-AUC
XGBoost	0.05	0.999950	0.9999	0.963235	0.981273	0.998933	0.970752
Random Forest	0.41	0.999929	0.992366	0.955882	0.973783	0.998458	0.968906
Logistic Regression	0.87	0.983369	0.068641	0.886029	0.127412	0.974832	0.043792

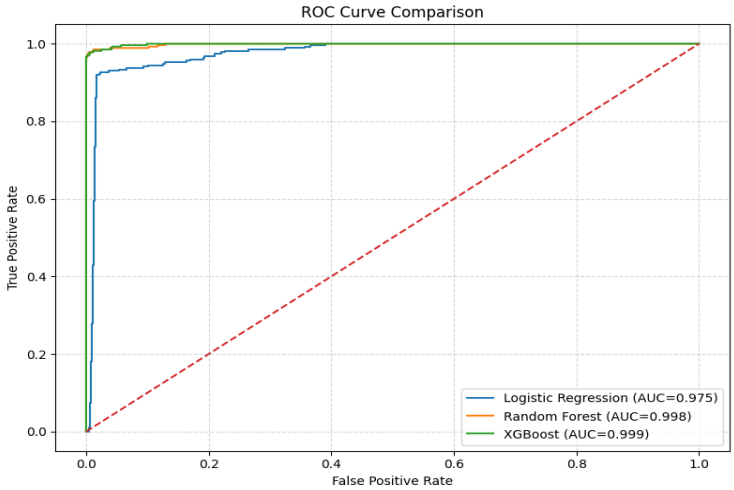

Figure 4. ROC curve comparison of evaluated models

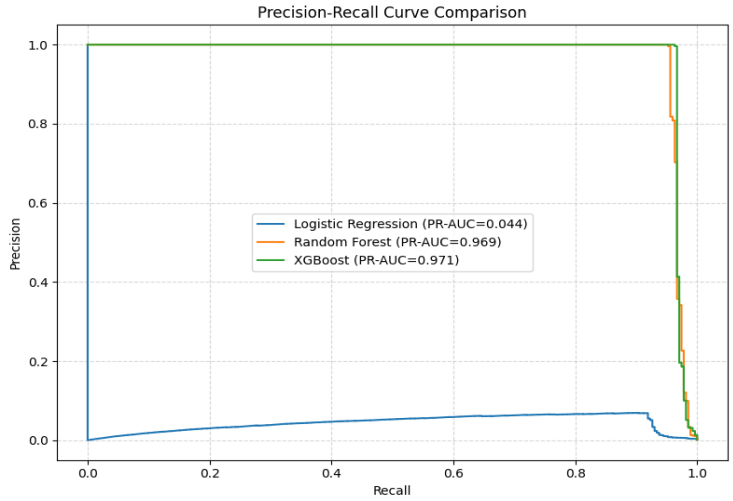

Figure 5. Precision–Recall curve comparison of evaluated models

4.4 Discrimination and ranking performance

The ROC curves revealed that all the three models gave high values of area-under-the-curve, however the separation of the classifiers was more evident with the use of PR-based evaluation (Figures 4 and 5). Since the class of fraud was very low, the Precision Recall curve provided a more informative method than accuracy itself. XGBoost

and Random Forest were very precise in recall levels, but logistic regression declined much, which is in line with its low PR-AUC (Figure 5).

The performance heatmap also revealed that the two ensemble models were leading on the linear baseline in all the key metrics that were sensitive to fraud, particularly the precision, F1-score, and PR-AUC (Figure 6). This supports the finding that non-linear ensemble learning offered better decision boundaries to the non-linear transactional patterns in the dataset.

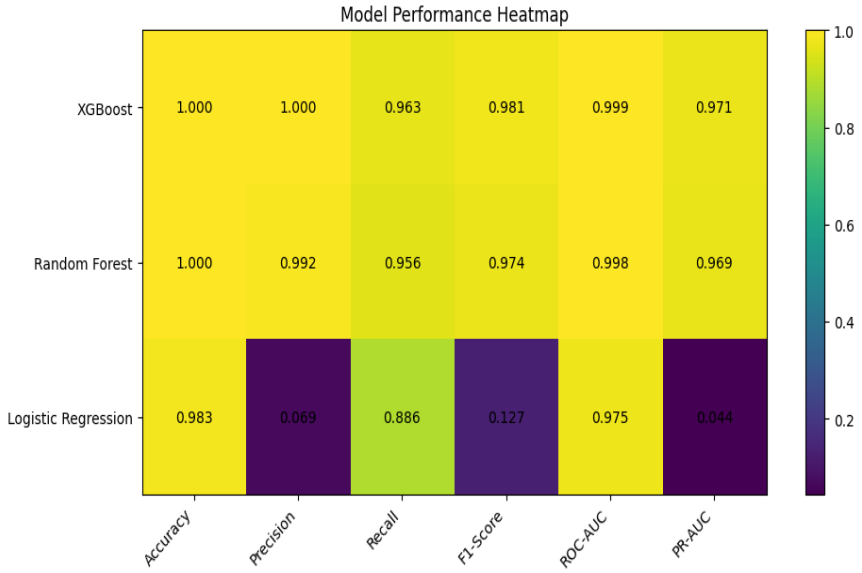

Figure 6. Model performance heatmap

4.5 Error structure and confusion matrix analysis

The heatmaps in the confusion matrices further elaborated the classification pattern of the three models (Figure 7). XGBoost had the best error structure with almost all fraudulent transactions detected and almost no false alarms. With a test fraud rate of 272, the XGBoost findings of 262 true positives, 10 false negatives, and 0 false positives indicated 100 percent accuracy of the selected threshold (Figure 7a).

The confusion structure was also highly effective with random forest that had few false positives and slightly more false negative than XGBoost (Figure 7b). Logistic Regression, in its turn, produced a significantly higher false positive and still managed to retrieve the majority of fraud cases, which is why its precision is low and why it is not much applicable in practice in terms of its use in monitoring fraud (Figure 7c).

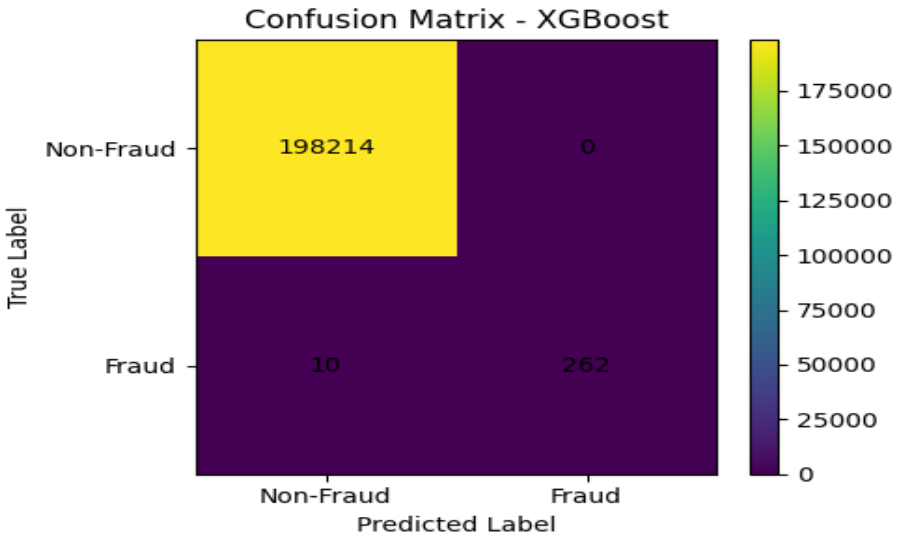

Figure 7a. Confusion matrix heatmap for XGBoost

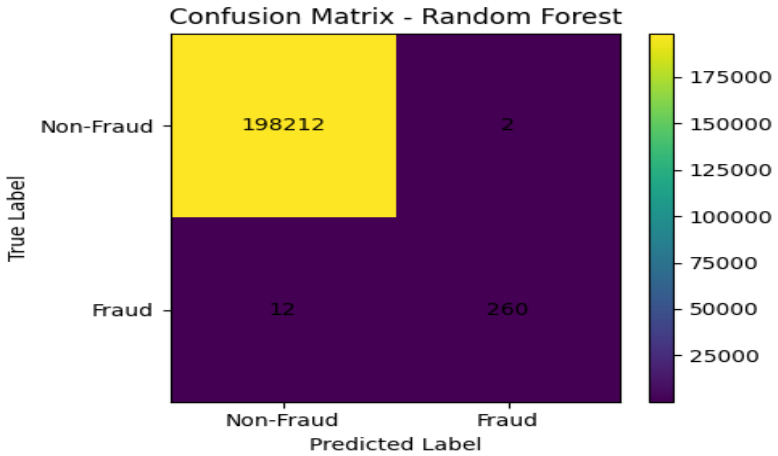

Figure 7b. Confusion matrix heatmap for Random Forest

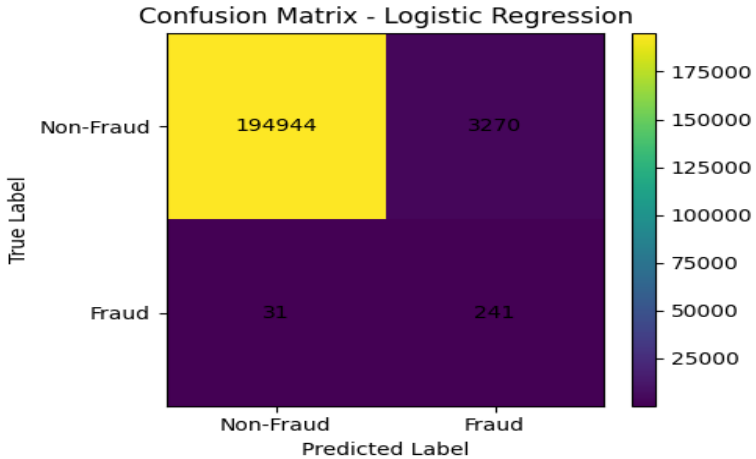

Figure 7c. Confusion matrix heatmap for Logistic Regression

4.6 Threshold optimization and probability separation

The threshold analysis of the best-performing model revealed that the decision cutoff was sensitive to the performance, and the chosen threshold of 0.05 gave the most desirable balance between the precision, recall, and F1-score on the validation set (Figure 8). This validates the fact that default thresholds are not always optimal in imbalanced financial fraud detection environments.

XGBoost predicted probability distributions were well separated in fraud and non-fraud transactions (Figure 9). The legitimate transactions were clustering towards the low fraud-probability values, and the fraudulent transactions were clustering towards a significantly higher predicted probability. This distributional separation provides a graphic support of the robust ROC-AUC and PR-AUC values that the model acquired.

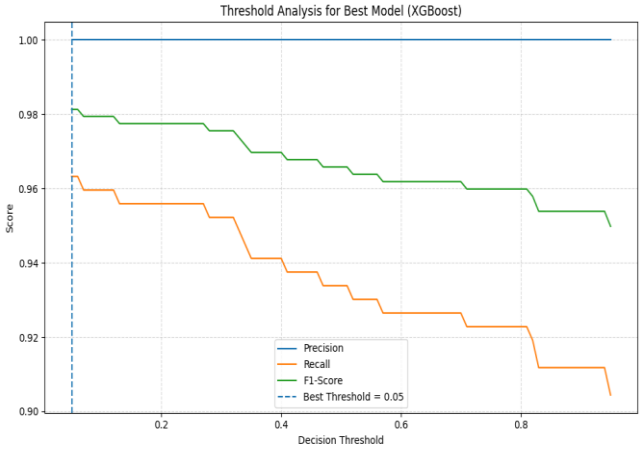

Figure 8. Threshold analysis for the best-performing model

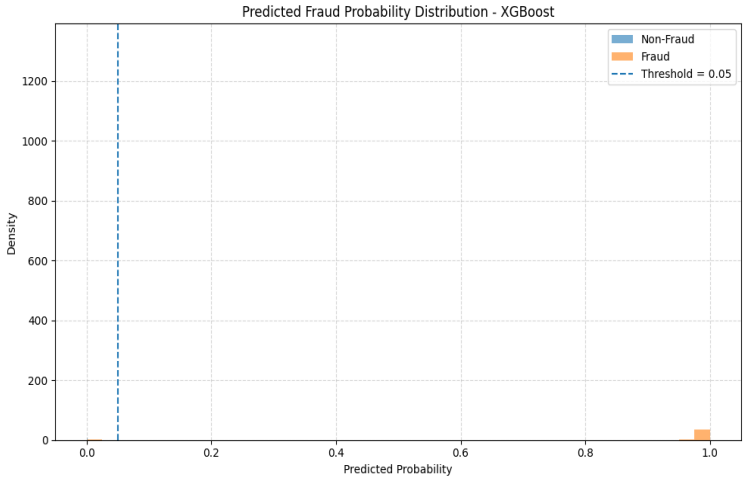

Figure 9. Predicted fraud probability distribution for XGBoost

4.7 Feature contribution analysis

Feature-importance analysis of the best-performing XGBoost classifier identified a small set of dominant predictors (Figure 10; Table 4). The most influential variable was AMOUNT_TO_SENDER_BALANCE_RATIO with an importance score of 0.2694, followed by SENDER_INIT_BALANCE (0.2528), SENDER_TX_BEHAVIOR_ID (0.1654), and TX_AMOUNT (0.1529). Additional meaningful features included LOG_TX_AMOUNT (0.0623) and BALANCE_DIFF (0.0296).

These findings suggest that the financial size of a transaction compared to account balance, the initial balance picture of the senders, and the behavioral pattern of the senders were the main factors (fraud detection) in this dataset. The temporal indicators like TIME_BIN_24 and TIME_BIN_7 were relatively minor, whereas a number of categorical country and account-type indicators had insignificant impact on the final model (Table 4).

Table 4. Top feature-importance scores of the XGBoost model

Rank	Feature	Importance
1	AMOUNT_TO_SENDER_BALANCE_RATIO	0.269358
2	SENDER_INIT_BALANCE	0.252771
3	SENDER_TX_BEHAVIOR_ID	0.165421
4	TX_AMOUNT	0.152916
5	LOG_TX_AMOUNT	0.062313
6	BALANCE_DIFF	0.029588
7	SENDER_ACCOUNT_ID	0.025260
8	TIMESTAMP	0.014779
9	TIME_BIN_24	0.005796
10	AMOUNT_TO_RECEIVER_BALANCE_RATIO	0.005781

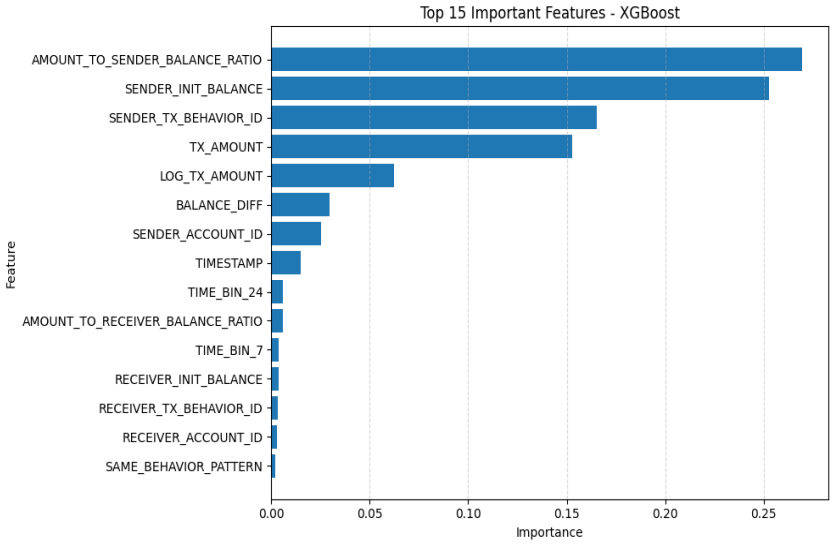

Figure 10. Feature-importance plot for the XGBoost model

4.8 Alert-level descriptive findings

Descriptive summary of alert information indicated that AML alert system produced 1,719 alerts, with all being fraudulent transactions in the labelled alert file (Table 5; Figure 11). There were two types of alerts: cycle alerts (936 cases) and fan in alerts (783 cases). The alert types showed an identical rate of 1.0 in fraud, which means that the alert data set was a set of confirmed suspicious patterns but not a mixed suspicious/non-suspicious set.

Table 5. Alert type distribution

Alert Type	Total Alerts	Fraud Rate	Fraud Count
cycle	936	1.0	936
fan_in	783	1.0	783

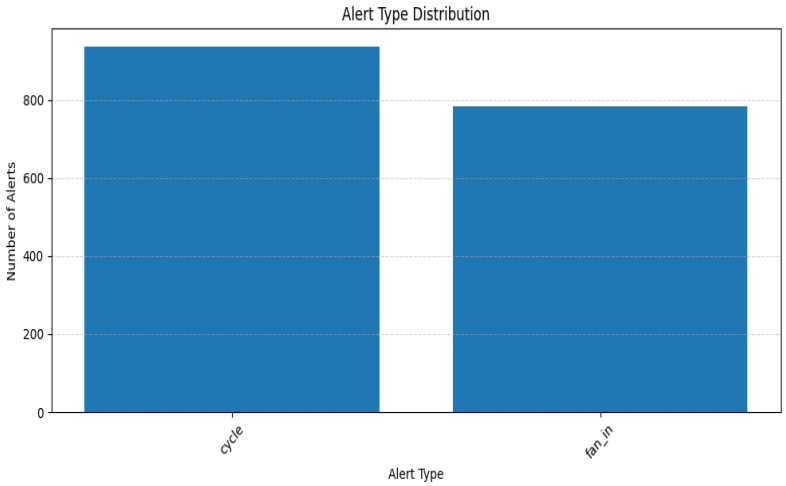

Figure 11. Alert type distribution

5. Discussion

The results of the current paper support the idea that artificial intelligence-driven models are quite efficient in detecting fraud in online financial operations, especially in the situation of extreme class imbalance when the percentage of fraudsters out of the total transactions is extremely low. The high performance of XGBoost and the Random Forest compared to the Logistic Regression suggests that the ensemble learning methods can better reflect the multi-layered behavioral, complex, and non-linear nature of transactional fraud. This finding aligns with the previous literature that has indicated that advanced AI systems are becoming more useful in the banking setting to detect and manage fraud, and comply with regulations [29] and that ensemble techniques are specifically useful in capturing dynamic and anomalous financial frauds [30]. The methodological decision to employ chronological splitting, leakage-sensitive feature selection and validation-based threshold optimization used in the current research seems to have enhanced the realism and reliability of the results by more closely approximating the future transaction monitoring as opposed to maximizing the performance with random sampling.

The difference between recall and precision between models is one of the results of specific significance. Although the Logistic Regression was able to identify a substantial proportion of frauds, it has a comparatively low accuracy, and, thus, it generates an unacceptable high ratio of false alarms, which would be inefficient in the real financial systems. Comparatively, XGBoost had much more favorable balance between the precision and recall, and the highest F1-score and PR-AUC of the tested models. This shows that not only did this model detect fraud it did so at significantly lower false-positive load. This balance is very essential in financial settings since missed fraud and too much generation of alerts are both costly to the economy and administration. This fact that XGBoost is better than the other can be logically justified by the fact that this model can learn complex relations between transaction amount ratios, sender balance conditions, behavioral indicators, and time-derived variables, which is consistent with the prior studies demonstrating that sophisticated AI methods are significant in improving the efficiency and safety of the financial system [31].

The feature-importance findings also add to the view that it is not one variable that can be said to be driving the fraud detection, but rather a combination of the financial magnitude, the account context and the behavior of transactions. The other variables, such as transaction-balance ratios, balance changes, and log-transformed transaction values, were significant predictors, which demonstrated that anomalous financial relationship are central to the discovery of suspicious activity. The threshold optimization process was also found to have methodological importance. The trade-off between sensitivity and precision in the study was reasonable because the cutoff of the decisions was selected on the validation set rather than default cutoff. It is particularly applicable to fraud analytics, where the best classification lies in the relative cost of a false positive and false negative. Similar studies also highlight that AI-aided optimization enhances the quality of classification and decisions in financial systems [32], whereas predictive analytics helps to make more data-driven institutional decisions [33]. The overall implications can also be applied to other related financial applications beyond fraud surveillance like customer relationship management and operational intelligence [34].

The other interesting result is the very high PR-AUC of the best model which is especially relevant in the context of rare event detection since it represents the capacity of the model to detect fraud on extreme imbalance more precisely compared to the accuracy. It is directly applicable to compliance-based financial structures where fraudulent events are not common, but when they occur, they have a significant impact that needs to be identified with high levels of reliability. Other past researchers have also noted the applications of AI systems in identifying tax scams and other financial fraud in bulk data [35]. The obtained findings, at the same time, validate the long-term relevance of anomaly-driven machine learning in organized financial databases, where deviant behaviour of transactions may serve as the preliminary signs of a criminal activity [36]. Despite these advantages, the study also possesses certain limitations including the synthetic nature of the data and the need to possess more effective interpretability processes. To make it even stronger and more applicable, the future research should examine the real-life data, clarify the AI models, privacy-preserving learning and hybrid or deep learning models.

6. Conclusion

In this paper, ensemble algorithms such as XGBoost, which rely on artificial intelligence, are efficient in fraud detection in an online financial transaction in a scenario involving extreme class imbalance. The empirical findings indicate that XGBoost performed significantly better than Logistic Regression and Random Forest based on the main evaluation parameters, such as precision, F1-score, ROC-AUC, and PR-AUC, which means that it has the better capacity to find non-linear, more complex patterns in the transactional data. The synergy of the strong feature engineering, temporal data separation, and threshold optimization using validation also enhanced the stability and the relevance of the proposed framework to the real-life situation. The findings reinforce the growing importance of AI-based systems in risk management in the financial industry, fraud prevention, and regulation compliance. Nevertheless, these contributions have certain limitations to the study. A synthetic data may not be as rich and dynamic as the actual world of financial environments. Besides, despite high predictive performance of the models, issues of interpretability and transparency remain a major issue, especially in regulated financial systems. The proposed framework should be applied to real-world data in future studies, explainable AI methods should be considered to enhance the interpretability of the results, and innovative architectures in the form of deep learning and graph-based models can be considered. In addition to this, privacy protective systems and hybrid anomaly detection methods can be integrated to enhance the robustness and scalability of the fraud detection systems in evolving financial ecosystems.

References

1. S. P. Bhavnani, K. Parakh, A. Atreja, R. Druz, G. N. Graham, S. S. Hayek, et al., “2017 Roadmap for innovation—ACC health policy statement on healthcare transformation in the era of digital health, big data, and precision health: A report of the American College of Cardiology Task Force on Health Policy Statements and Systems of Care,” J. Amer. Coll. Cardiol., vol. 70, no. 21, pp. 2696–2718, 2017.
2. A. Q. Li and P. Found, “Towards sustainability: PSS, digital technology and value co-creation,” Procedia CIRP, vol. 64, pp. 79–84, 2017.
3. T. Yu and W. Shen, “Funds sharing regulation in the context of the sharing economy: Understanding the logic of China’s P2P lending regulation,” Comput. Law Secur. Rev., vol. 35, no. 1, pp. 42–58, 2019.
4. N. Tariq, M. Asim, and F. A. Khan, “Securing SCADA-based critical infrastructures: Challenges and open issues,” Procedia Comput. Sci., vol. 155, pp. 612–617, 2019.
5. O. Acar and Y. E. Çıtak, “Fintech integration process suggestion for banks,” Procedia Comput. Sci., vol. 158, pp. 971–978, 2019.
6. J. Schot and L. Kanger, “Deep transitions: Emergence, acceleration, stabilization and directionality,” Res. Policy, vol. 47, no. 6, pp. 1045–1059, 2018.
7. S. P. Jun, H. S. Yoo, and S. Choi, “Ten years of research change using Google Trends: From the perspective of big data utilizations and applications,” Technol. Forecast. Soc. Change, vol. 130, pp. 69–87, 2018.
8. N. Passas, “Report on the debate regarding EU cash payment limitations,” J. Financial Crime, vol. 25, no. 1, pp. 5–27, 2018.
9. M. M. Singh and A. A. Bakar, “A systemic cybercrime stakeholders architectural model,” Procedia Comput. Sci., vol. 161, pp. 1147–1155, 2019.
10. S. Elyassami, H. N. Humaid, A. A. Alhosani, and H. T. Alawadhi, “Artificial intelligence-based digital financial fraud detection,” in Proc. Int. Conf. Intelligent and Fuzzy Systems, Cham, Switzerland: Springer, 2021, pp. 214–221.
11. K. Goyal, M. Garg, and S. Malik, “Adoption of artificial intelligence-based credit risk assessment and fraud detection in the banking services: A hybrid approach (SEM-ANN),” Future Bus. J., vol. 11, no. 1, p. 44, 2025.
12. L. Rodríguez Valencia, M. J. Ochoa Arellano, S. A. Gutiérrez Figueroa, C. Mur Nuño, B. Monsalve Piqueras, A. D. V. Corrales Paredes, et al., “A systematic review of artificial intelligence applied to compliance: Fraud detection in cryptocurrency transactions,” J. Risk Financial Manag., vol. 18, no. 11, p. 612, 2025.

13. O. T. Odofin, A. A. Abayomi, A. C. Uzoka, B. I. Adekunle, O. A. Agboola, and S. Owoade, “Integrating artificial intelligence into telecom data infrastructure for anomaly detection and revenue recovery,” Iconic Res. Eng. J., vol. 5, no. 2, pp. 222–234, 2021.
14. K. Randhawa, C. K. Loo, M. Seera, C. P. Lim, and A. K. Nandi, “Credit card fraud detection using AdaBoost and majority voting,” IEEE Access, vol. 6, pp. 14277–14284, 2018.
15. R. Jean-Philippe, “Enhancing computer network defense technologies with machine learning and artificial intelligence,” Master’s thesis, Utica College, 2018.
16. O. Byambaa, C. Yondon, E. Rentsen, B. Darkhijav, and M. Rahman, “An empirical examination of the adoption of artificial intelligence in banking services: The case of Mongolia,” Future Bus. J., vol. 11, no. 1, p. 76, 2025.
17. R. Abulsaoud Ahmed Younis and H. M. Adel, “Artificial intelligence strategy, creativity-oriented HRM and knowledge-sharing quality: Empirical analysis of individual and organisational performance of AI-powered businesses,” in Proc. Annu. Int. Conf. Brit. Acad. Manage. (BAM), Sep. 2020.
18. D. Goswami, “Advancing threat detection through artificial intelligence and machine learning enhanced cybersecurity audits,” Amer. J. Scholarly Res. Innov., vol. 4, no. 01, pp. 428–457, 2025.
19. V. K. Singh and K. Sinha, “How artificial intelligence is going to revolutionize the non-banking financial companies,” Strad Res., vol. 11, no. 10, pp. 81–95, 2024.
20. M. Akter and S. P. Kudapa, “A comparative analysis of artificial intelligence-integrated BI dashboards for real-time decision support in operations,” Int. J. Sci. Interdisciplinary Res., vol. 5, no. 2, pp. 158–191, 2024.
21. Z. Kuswara, M. Pasaribu, F. Fitriana, and R. A. Santoso, “Artificial intelligence in financial reports: How it affects the process's effectiveness and efficiency,” J. Ilmu Keuangan dan Perbankan (JIKA), vol. 13, no. 2, pp. 257–272, 2024.
22. E. Edozie, A. N. Shuaibu, B. O. Sadiq, and U. K. John, “Artificial intelligence advances in anomaly detection for telecom networks,” Artif. Intell. Rev., vol. 58, no. 4, p. 100, 2025.
23. M. S. Ali, I. A. Swiety, and M. H. Mansour, “Evaluating the role of artificial intelligence in the automation of the banking services industry: Evidence from Jordan,” Humanities Soc. Sci. Lett., vol. 10, no. 3, pp. 383–393, 2022.
24. F. Bou Reslan and N. Jabbour Al Maalouf, “Assessing the transformative impact of AI adoption on efficiency, fraud detection, and skill dynamics in accounting practices,” J. Risk Financial Manag., vol. 17, no. 12, p. 577, 2024.
25. M. T. Islam, M. M. Hasan, M. Redwanuzzaman, and M. K. Hossain, “Practices of artificial intelligence to improve the business in Bangladesh,” Soc. Sci. Humanities Open, vol. 9, p. 100766, 2024.
26. P. Dubey, P. Dubey, and P. N. Bokoro, “A unified transformer–BDI architecture for financial fraud detection: Distributed knowledge transfer across diverse datasets,” Forecasting, vol. 7, no. 2, p. 31, 2025.
27. A. Raihan, “Financial technology optimization using artificial intelligence (AI) to accomplish sustainable development goals (SDGs),” in Proc. Int. Conf. Digital Finance, Green Innov., Sustain. Develop., 2024.
28. MLG-ULB, “Credit Card Fraud Detection,” Kaggle, 2018. [Online]. Available: https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud.
29. L. A. R. Aziz and Y. Andriansyah, “The role of artificial intelligence in modern banking: An exploration of AI-driven approaches for enhanced fraud prevention, risk management, and regulatory compliance,” Rev. Contemporary Bus. Analytics, vol. 6, no. 1, pp. 110–132, 2023.
30. M. K. Alam, M. A. Mahmud, and M. A. Alam, “Adversarial machine learning for robust fraud detection in high-frequency financial transactions,” J. Comput. Sci. Technol. Stud., vol. 7, no. 8, pp. 314–335, 2025.
31. A. Ahmed, A. Shah, T. Ahmed, S. Yasin, F. E. A. Longa, W. Hussaini, and M. Zubair, “AI-driven innovations in modern banking: From secure digital transactions to risk management, compliance frameworks, and AI-based ATM forecasting systems,” J. Manage. Sci. Res. Rev., vol. 4, no. 3, pp. 1145–1183, 2025.
32. M. A. Jaafar, “The impact of artificial intelligence in detecting manipulations in financial statements,” Futurity Econ. Law, vol. 5, no. 2, pp. 51–72, 2025.
33. S. Malpani, “To study the role of artificial intelligence in financial decision making.”
34. D. Jindoliya, D. Y. Choudhary, D. S. Kaur, and S. Asthana, “Role of artificial intelligence (AI) based customer relationship management (CRM) in customer retention: An empirical investigation for fintech organisations,” SSRN Electron. J., Art. no. 5466507, 2025.
35. J. Alm and R. Belahouaoui, “Emerging trends in tax fraud detection using artificial intelligence-based technologies,” Tulane Univ., Dept. Econ., Working Papers, no. 2511, 2025.
36. R. Maddali, “Machine learning for SQL-based anomaly detection & fraud analytics in financial data,” Int. J. Leading Res. Publication, vol. 2, no. 10, pp. 1–11, 2021.