AI-Powered Fraud Detection in Real-Time Financial Transactions

Article · March 2025

CITATIONS
72

READS
119

10 authors, including:

Venkata Reddy Pasam
Institute of Electrical and Electronics Engineers
6 PUBLICATIONS **72** CITATIONS
SEE PROFILE

Diganta Sen Gupta
Oracle Corporation
1 PUBLICATION **72** CITATIONS
SEE PROFILE

AI-Powered Fraud Detection in Real-Time Financial Transactions

Vishnu Ravi[1], Vineet Kumar Srivastava[2], Maninder Pal Singh[3], Ravi Kumar Burila[4], Srinivas Chippagiri[5], Venkata Reddy Pasam[6], Diganta Sengupta[7], Nuzhat Noor Islam Prova[8*], Indrajit De[9]

[1]Lead Software Engineer, Bayonne, New Jersey, 07002, USA.
[2]Senior Software Engineer, Peoria, Arizona, 85382, USA.
[3]Lead Software Engineer, Princeton, New Jersey, 08540, USA.
[4]VP, Cloud Services, JP Morgan Chase, Columbus, Ohio, 43240, USA.
[5]Sr. Member of Technical Staff, Salesforce Inc, Bellevue, WA, 98004.
[6]Sr. Data Engineer, Cornerstone Consulting, Irving, Texas, 75063, USA.
[7]Principal Enterprise Architect, Oracle Corp, Austin, Texas, 78759, USA.
[8*]Seidenberg School of CSIS, Pace University, New York, 10038, USA.
[9]Advisor, IIFON, Kolkata, 700091, India.

*Corresponding author(s). E-mail(s): nuzhatnsu@gmail.com;
Contributing authors: vishnu3186@gmail.com; icyvineet@gmail.com; mpsgotra85@gmail.com; reachburila@gmail.com; cvas22@gmail.com; venkatdbapp@gmail.com; seng78in@gmail.com; indrajit.de@iem.edu.in;

Abstract

The rising amount of digital financial transactions has brought about in a growth in fraudulent activity, consequently compromising users and financial institutions significantly. More complicated methods are needed since conventional rule-based fraud detection systems find it difficult to keep up with changing fraud schemes. In addition XGBoost, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Artificial Neural Networks (ANN), and Random Forest (RF) machine learning (ML) models in real-time fraud detection is looked at in this paper. These models are trained and tested on a vast dataset comprising over six billion transactions. With 100% accuracy, precision, and recall, the RF model surpasses all others, based to the results, hence it is the most dependable method for identifying fraudulent transactions with least false positives. Important difficulties in ML-based fraud detection such as handling imbalanced datasets, modifying to

new fraud techniques, and increasing computational efficiency for real-time processing also are covered in this paper. Furthermore suggested is an ensemble-based method using many ML methods to improve fraud detection accuracy. The findings show that artificial intelligence-driven fraud detection offers a scalable and quick fix for recent banking systems, consequently considerably improving financial security. Deep learning algorithms, real-time adaptive learning, and improved data integration techniques should all be investigated in future work to maximize fraud detection capacity.

Keywords: Fraud Detection, Machine Learning (ML), Real-Time Transactions, Financial Security, Random Forest (RF), Anomaly Detection.

1 Introduction

We are almost certain to live in today's cashless world. Various studies and opinions show that online transaction activity has grown substantially in the past few years, and this field should see much more progress. Although this is encouraging news, it usually comes with the negative side of more false transactions [1]. The banking and finance industry deals with uncommon challenges as fraud and risk management become more complicated in a digital environment.

Usually rule-based systems, typical fraud detection techniques sometimes give up because they cannot change to fit transforming deceptive schemes [2]. Often utilizing static models and historical data, specific fraud detection and risk management methods are substituted with AI-driven systems with real-time analysis and predictive responsibilities. Rising as game-changing technologies with great tools for fraud detection and risk assessment were artificial intelligence (AI) and ML. Bringing predictive analytics with AI into financial services is a significant step toward better risk management and fraud prevention. AI-powered systems can quickly process enormous volumes of data, spot trends, and highly precisely find anomalies[3] [4].

Trained on historical transaction data, several ML models including XGBoost, Random Forest (RF), Support Vector Machines (SVM), K-Nearest Neighbors (KNN) and Artificial Neural Network (ANN) identify fraud developments. Several models are provided on labeled historical transaction data to detect patterns dividing real and fake transactions. XGBoost effectively combines several models to enhance predictions; it is consequently capable of identifying intricate fraud patterns [5]. Although it can be slow for real-time use, SVM sets apart legal and fraud by establishing a boundary. While it can be slow with massive quantities of data, KNN classifies transactions relying on a connection to others [6]. ANN provides great accuracy but requires more excellent computer capability by learning patterns in data using several layers [7].

Beyond all of these, the RF is the indicated model due to its Interpretation, significant accuracy, and resiliency. Getting transactions sent through the trained models in real time, where each model uses acquired information to identify whether the transaction is genuine or a fraud. Considered an ensemble of decision trees, RF builds up several forecasts to provide substantial detection accuracy.

This research consequently has several limitations. XGBoost can be computationally expensive for massive data sets and requires careful tuning even though it is pretty accurate. High-dimensional data presents difficulties for SVM, which causes it to lose efficiency for real-time fraud detection. KNN compares every transaction with all past ones, so it is slow for massive data sets. Although ANN offers excellent accuracy, real-time processing is complex since it demands significant volumes of training data and great computational resources.

This research contributes the following key contributions:

- This article offers a real-time fraud detection system that produces few false positives that accurately recognizes fraudulent transactions, consequently guaranteeing improved financial system security.
- Comparatively analyzing several ML models (RF, XGBoost, SVM, KNN, and ANN), the work suggests that RF is best suited for real-time fraud prevention and possesses the highest detection accuracy (100%).
- Leveraging a dataset of more than 6 million transactions, the work confirms the scalability and resilience of ML models in managing biased data and identifying advanced fraud decisions.

This paper's remaining sections are organized as follows: A thorough analysis of previous research on AI-based financial market prediction is given in section 2. The methodology is fully described in section 3, covering model design, optimization techniques, and data pretreatment. The experimental results are shown in section 4, which compares the model's performance to the most advanced techniques. Finally, section 5 concludes by summarizing the main conclusions and reviewing possible improvements for further studies, such as incorporating more advanced deep learning (DL) methods and real-time financial data processing.

2 Literature Review

Research on multiple computing methodologies, such as rule-based systems, ML models, and DL approaches, has been implemented to improve automated financial transaction systems.

Johora et al.[8] demonstrated the limits of conventional fraud detection techniques and rising cybersecurity risks in banking. It underlines how well logistic regression and DT are two examples of AI and ML algorithms that improve fraud detection accuracy. Experiments demonstrate that AI models have great accuracy; logistic regression performs best at 98%, so providing enhanced security in financial transactions. Additionally, primary concerns, though, were data imbalance and algorithm transparency.

Likewise, Ismaeil et al.[9] examining the development of fraud detection from conventional rule-based approaches to AI-driven solutions. It emphasized how well ML and DL approaches identify intricate fraud patterns, lower false positives, and provide real-time fraud protection. With models achieving accuracy rates as high as 98%, the study highlighted the part AI plays in adjusting to new fraud tendencies.

Kotha et al.[10] emphasized ML, DL, and NLP, observing the shift from conventional rule-based fraud detection to AI-driven methodologies. AI algorithms show better accuracy in recognizing fraud by finding irregularities in financial transactions sometimes overlooked by more traditional techniques. Studies demonstrate that AI-driven fraud detection systems reach great accuracy up to 98% while also being very sensitive. Despite data privacy and system integration issues, AI keeps transforming fraud prevention to guarantee improved security and efficiency in the financial sector.

Rani et al.[11] demonstrated how rule-based fraud detection systems have transitioned to AI-driven solutions in modern banking, an exploration of AI-Driven approaches for enhanced fraud prevention, risk management, and regulatory compliance deals with appears. As the main AI tools for increasing risk management and fraud prevention, it demonstrates graph analytics, NLP, and deep learning. With AI models reaching up to 98% accuracy, the study noted great fraud detection capacity. Despite this, issues such as data security, AI bias, and regulatory compliance will be their top concern for financial institutions.

Furthermore, Khare et al.[1] investigated several ML methods for digital transaction fraud detection, showing how well supervised learning models—including logistic regression, RF, and decision trees—work in spotting fraudulent behavior. With a 99% accuracy level, the DT model turned out to be the most successful approach among the other models according to the research. Still, significant areas of additional investigation are class imbalance, data preparation, and the necessity of enhanced DL methods.

Additionally, Marripudugala et al.[12] examined how ML, DL, and NLP AI-powered technologies—improve banking transaction efficiency and fraud detection. According to the findings, AI-driven fraud detection systems can more accurately recognize complicated fraudulent developments than conventional rule-based approaches. Results showed that up to 98% of ML models detect fraud, lowering false positives and improving real-time risk management. Still, financial institutions' top concerns remain data reliance, regulatory compliance, and AI model biases.

Bello et al.[13] looked at changing from conventional fraud detection strategies to AI-based solutions. They demonstrate how well anomaly detection, DL, and ML identify criminal activity more efficiently and accurately. The study notes that Gradient Boosting Machines (GBM) significantly lower false positives by attaining the best accuracy of 95%. Nevertheless, implementing AI-driven fraud detection in real-time financial transactions depends on data quality, security concerns, and regulatory compliance.

Limitations and Difficulties with Stock Price Prediction Models Found:

- Unlike genuine transactions, fraudulent ones are rarely generated and generate imbalanced datasets that could skew models toward non-fraudulent scenarios.
- The constant evolution of their approaches enables fraudsters to maybe avoid detection systems based on their fraudulent movements.
- Real-time handling of enormous transaction data demands efficient systems and can present challenges with execution.
- Complex models would overfit the training data, thus decreasing their ability to generalize to unresolved fraud patterns.

3 Methodology

In this section, we briefly discuss our overall research methodology on fraud detection in real-time financial transactions, along with the dataset, and prepare the data and relevant techniques. We employ several AI models, including XGBoost, RF, SVM, KNN and ANN. Figure 1 displays the overall research methodology.

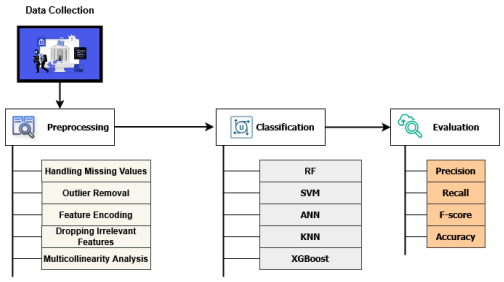

Fig. 1: Graphical Representation of the Overall Research Methodology

3.1 Dataset Description

In this study, we fetched the fraud dataset from Kaggle comprising 6,353,307 transactions, each with 11 features, designed to simulate and detect fraudulent activities in financial transactions. The data spans 744 time steps, each representing one hour over 30 days. The key columns include 'step,' which means the time unit, and 'type,' which categorizes the transaction as CASH-IN, CASH-OUT, DEBIT, PAYMENT, or TRANSFER. The 'amount' column records the transaction value in local currency, while 'nameOrig' and 'nameDest' denote the sender and recipient, respectively. The 'oldbalanceOrg' and 'newbalanceOrig' capture the account balance before and after the transaction for the sender, while 'oldbalanceDest' and 'newbalanceDest' track the same for the recipient, with some records lacking recipient information for transactions involving merchants. The 'isFraud' column indicates whether a transaction is fraudulent, while 'isFlaggedFraud' flags an attempt to transfer more than 200,000 in a single transaction as potential fraud. Where, 0 indicates a legitimate transaction, while 1 represents fraud. Figure 2 exhibits the correlation matrix of the dataset's features.

3.2 Data Preprocessing

In this study, we employ various data pre-processing techniques to improve the quality of the data set and provide a robust fraud detection model. The missing values were initially addressed, as the data set contained explicit null values. The average imputation method was applied to impute these missing values:

$$X_{\text{imputed}} = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{1}$$

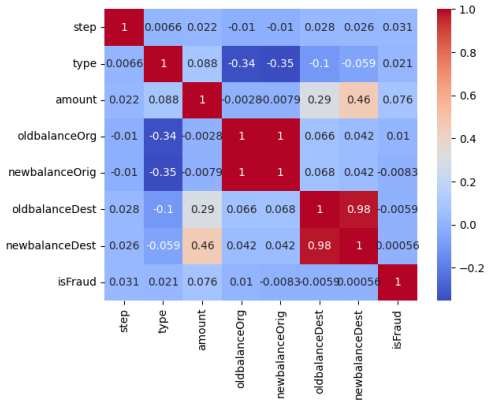

Fig. 2: The correlation matrix of the dataset's features

where X_{imputed} represents the imputed value, and X_i denotes the observed values.

Outlier removal was performed using the Z-score and interquartile range (IQR) methods. Transactions with Z-scores beyond ± 3 were considered outliers and removed. The Z-score was computed as:

$$Z = \frac{x - \mu}{\sigma} \tag{2}$$

where Z is the Z-score, x is the determined value, μ is the mean, and σ is the standard deviation.

For the IQR-based outlier removal, the following equations were used:

$$IQR = Q_3 - Q_1 \tag{3}$$

$$\text{Lower Bound} = Q_1 - 1.5 \times IQR, \quad \text{Upper Bound} = Q_3 + 1.5 \times IQR \tag{4}$$

In this case, the first and third quartiles are denoted by Q_1 and Q_3.

To enable numerical processing, categorical attributes were transformed. The *type* feature was encoded using label encoding:

$$X_{\text{encoded}} = f(X_{\text{categorical}}) \tag{5}$$

where $X_{\text{categorical}}$ represents the categorical feature, and f denotes the encoding function.

Additionally, irrelevant features such as *nameOrig*, *nameDest*, and *isFlaggedFraud* were dropped, as they provided minimal contribution to fraud detection.

Multicollinearity was assessed using Pearson correlation analysis and Variance Inflation Factor (VIF) calculations. The VIF for each feature was computed as follows:

$$VIF_i = \frac{1}{1 - R_i^2} \tag{6}$$

where R_i^2 is the coefficient of determination from a regression model predicting the ith feature using all other independent features.

Table 1: Random Forest Model's Hyperparameters for Real-Time Fraud Detection

Hyperparameter	Description
`n_estimators`	100
`max_features`	sqrt
`min_samples_leaf`	5
`n_jobs`	-1
`random_state`	42
`oob_score`	false

3.3 Classification

Our fraud detection model employs several AI techniques, including XGboost, SVM, RF, KNN, and ANN, which examine transactional data to identify fraudulent patterns in real-time transactions. By combining these several methods, the model reduces false positives and improves fraud detection effectiveness.

3.3.1 Random Forest Classifier (RF)

We employed the RF Classifier for fraud detection due to its ensemble learning capability, which builds numerous DTs and aggregates their predictions to enhance accuracy while reducing overfitting [14]. Mathematically, each decision tree $h_t(x)$ makes a prediction based on its learned decision function $f_t(x)$. The final classification output is determined through majority voting, where the predicted class $\hat{y}$ is given by:

$$\hat{y} = \text{mode}\{h_1(x), h_2(x), ..., h_T(x)\} \tag{7}$$

where, T denotes the total number of trees. This method reduces the risk of overfitting, effectively handles numerical and categorical data, and ensures resilience against noise, making it a practical choice for fraud detection.

3.3.2 XGboost

By reducing the loss function and avoiding overfitting, the potent gradient-boosting method XGBoost improves prediction accuracy [15]. The objective of this function is:

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{t=1}^{T} \Omega(f_t) \tag{8}$$

where $l(y_i, \hat{y}_i)$ defines the loss function, and $\Omega(f_t)$ is the regularization term. XGBoost's efficiency, scalability, and handling of imbalanced data make it suitable for fraud detection.

3.3.3 Support Vector Machines (SVMs)

The SVM is a supervised learning model employed for classification tasks [16],[17]. It finds the optimal hyperplane $\mathbf{w} \cdot \mathbf{x} + b = 0$ that separates data points of different

classes with the largest margin. The optimization problem is:

$$\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|^2 \tag{9}$$

subject to the constraint:

$$y_i\left(\mathbf{w}\cdot\mathbf{x}_i + b\right) \geq 1, \quad \forall i \tag{10}$$

where y_i is the label of the data point $\mathbf{x}_i$. For non-linear data, SVM uses a kernel function $\phi(\mathbf{x})$ to map the data into higher dimensions, allowing for linear separation. The decision function is given by:

$$f(\mathbf{x}) = \text{sign}\left(\sum_{i=1}^{N} \alpha_i y_i \left(\mathbf{x}_i \cdot \mathbf{x}\right) + b\right) \tag{11}$$

3.3.4 K-Nearest Neighbors (KNN)

A straightforward non-parametric classifier, the KNN algorithm provides a class label to a data point based on the majority class of its k nearest neighbors in the feature space [18]. Given a test point $\mathbf{x}$, the algorithm calculates the distance to all training points and selects the k closest points. Euclidean, Manhattan, or Minkowski distance are the most common distance metrics. The Euclidean distance between two points $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{in})$ and $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is given by:

$$d(\mathbf{x}_i, \mathbf{x}) = \sqrt{\sum_{j=1}^{n} (x_{ij} - x_j)^2} \tag{12}$$

After calculating the distances, the algorithm selects the k nearest neighbors and classifies $\mathbf{x}$ based on the majority vote of those neighbors:

$$\hat{y} = \text{mode}\{y_1, y_2, \ldots, y_k\} \tag{13}$$

where $y_1, y_2, \ldots, y_k$ are the class labels of the nearest neighbors.

3.3.5 Artificial Neural Network (ANN)

An ANN computes the output a_i of each neuron in the hidden layers as a weighted sum of inputs, followed by an activation function f:

$$a_i = f\left(\sum_{j=1}^{n} w_{ij} x_j + b_i\right) \tag{14}$$

To reduce the loss, the weights are changed during network training.

4 Result and Analysis

This section presents the model performance evaluation metrics, real-time adaptation, and experimental setup. The models' results are examined, and a comparison with current methods is presented. The outcomes demonstrate our model's exceptional accuracy and handling of big datasets and real-world situations.

4.1 Experimental Setup

In our research, we use several ML models for real-time fraud detection; among them, RF is the best model configured with 100 trees, a maximum depth of 10, and the Gini index as the splitting criterion. The model is trained on 80% of the data and tested on 20%, using a batch size of 64 transactions. We perform 5-fold cross-validation to ensure generalizability and train for 3000 iterations. The hyperparameters of the RF model, such as the number of estimators, maximum features for splitting, minimum samples per leaf, and parallelization options, are carefully tuned to enhance predictive accuracy, reduce overfitting, and optimize computational efficiency for real-time fraud detection, which is presented in Table 1. Hardware resources include a Tesla K80 GPU with 2496 CUDA cores and 12 GB of VRAM, a 2.3 GHz Xeon CPU, 12 GB of RAM, and 2 GB of disk space. Key software tools include numpy, scikit-learn, TensorFlow (with Keras), and pandas for efficient model implementation and data handling.

4.2 Real-Time Adaptation

In our fraud detection framework, real-time adaptation is achieved through a dynamic ML approach that continuously updates its model parameters based on incoming transaction data. Each transaction is characterized by a feature set X_t at time t, which includes transaction history, user behavior, and external threat intelligence. The fraud detection model processes these features to classify transactions as fraudulent or legitimate.

To improve detection accuracy and reduce false positives, we employ an adaptive learning mechanism that refines the classification model over time. Given an initial decision function $f(X_t)$, model updates are driven by error minimization:

$$\theta_{t+1} = \theta_t - \eta \nabla L(f(X_t), y_t), \tag{15}$$

where θ_t denotes the model parameters, η is the learning rate, and $L(f(X_t), y_t)$ represents the loss function according to classification errors. The model constantly modifies its decision boundaries to adapt to growing fraud patterns.

We incorporate an ensemble-based technique, in which several models influence the final decision, to improve real-time adaptation further:

$$F(X) = \sum_{i=1}^{N} w_i f_i(X), \tag{16}$$

where w_i means the weight allocated to each model $f_i(X)$. This approach leverages model diversity, decreasing sensitivity to distribution shifts in transaction data.

4.3 Evaluation Metrics

This experiment used several well-known evaluation metrics to assess the performance of the suggested ML models. While accuracy offered a general indicator of accurate predictions, precision and recall concentrated on the model's capacity to detect fraudulent transactions and accurately reduce false positives. The confusion matrix provided more in-depth information about the categorization performance by displaying true positives, false positives, true negatives, and false negatives. The F1 Score was then used to evaluate the balance between recall and precision.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{17}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{18}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{19}$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{20}$$

Where,

- TP = True Positive
- TN = True Negative
- FP = False Positive, and
- FN = False Negative

4.4 Outcomes of the Models

Table 2: Overall Performance on Fraud Detection in Real-Time Financial Transactions

Model	Accuracy	Precision	Recall	F1-Score
XGBoost	95%	94%	96%	95%
RF	100%	100%	100%	100%
SVM	92%	91%	93%	92%
ANN	88%	85%	90%	87%
KNN	90%	89%	91%	90%

In Table 2, some primary performance indicators like accuracy, precision, recall, and F1-score that are suitable for evaluating the proficiency of models in detecting frauds involved in real-time financial transactions are presented. These measures reduce false positives and negatives by providing a model to identify highly trustworthy fraudulent transactions.

The RF performs more effectively than any other model when compared to the others, achieving perfect scores on all metrics: 100% accuracy, precision, recall, and F1-score. This implies that with its high sensitivity and specificity, RF is the best method

for identifying fraud. On the other hand, the XGBoost model performs impressively with 95% accuracy and 94% precision, but its recall is slightly lower at 96%, showing a good balance but not as perfect as RF. The SVM performs reasonably well with 92% accuracy and 91% precision but falls short in recall compared to RF and XGBoost, indicating that it misses some fraudulent transactions. The ANN achieves the lowest scores among the models, with 88% accuracy and an F1 Score of 87%, indicating that it is less reliable for this particular task. The KNN model, with 90% accuracy and an F1 Score of 90%, is comparable to SVM but still not as effective as XGBoost or RF. The RF model is the most dependable for identifying fraud in real-time financial

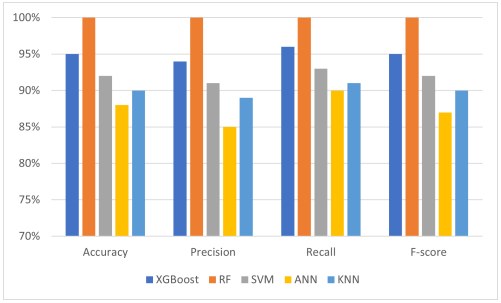

Fig. 3: The correlation matrix of the dataset's features

transactions because of its flawless precision, recall, and F1 score shown in Figure 3. Although alternative models such as XGBoost and SVM exhibit potential, RF is the best option for this task due to its ability to address both true and false positives.

The confusion matrix illustrates in Figure 4 the performance of our fraud detection model in distinguishing between fraudulent and legitimate transactions. The TN (1,906,334) indicates the number of legitimate transactions correctly classified. The FP (28) represent legitimate transactions incorrectly flagged as fraudulent. An important area for development is the FN (511), which shows fraudulent transactions that were mistakenly classed as lawful. Concurrently, the TP (1,909) indicates fraud cases that were appropriately identified. The FN show that more work is needed to improve fraud detection accuracy, while the low FP rate suggests there will be minor inconvenience to real users. The model accuracy and loss curves over several training epochs are shown in Figure 5. Accuracy trends are displayed in the top figure, demonstrating how training and testing accuracy quickly converge to a high value, signifying robust model performance with little overfitting. The bottom plot displays model loss, which significantly drops within the first epoch and stabilizes at a low value for training and testing. This suggests effective learning, with the model achieving convergence early in training. The minimal difference between training and testing curves in both plots indicates that the model generalizes well to unseen data.

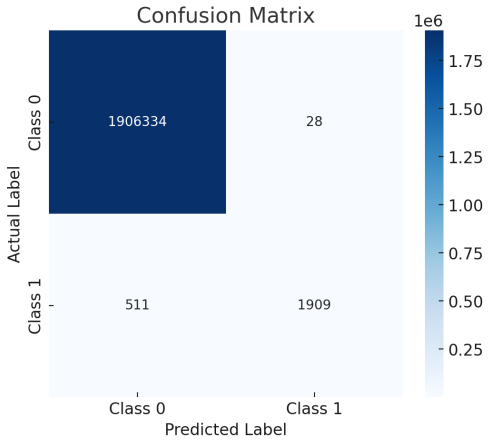

Fig. 4: Confusion Matrix of the RF Model

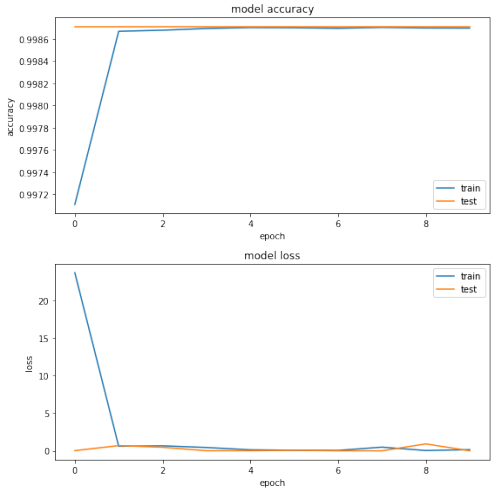

Fig. 5: The Model Accuracy and Model Loss

4.5 Comparative Analysis

This paper evaluates many ML models for fraud detection where RF has the highest accuracy (100%), followed by XGBoost (95%), SVM (92%), KNN (90%), and ANN (88%). Our analysis of 6,353,307 transactions shows RF's excellence when compared to Khare et al.[1], which used DT with 99% accuracy on 284,807 transactions. Although LR helps Johora et al.[8] attain 98% accuracy, LR suffers with complicated fraud patterns. Though GBM lacks efficiency in real-time fraud detection, Oguntibeju et al.[19] and Shabir et al.[3] employed Gradient Boosting Machine (GBM), obtaining 95% and 96% accuracy, respectively. Unlike past models, RF guarantees real-time fraud

Table 3: Compare our model with existing works

Reference	Dataset	Data Size	Model	Accuracy
Johora et al.[8]	Interviews and surveys	Not specified	LR	98%
Oguntibeju et al.[19]	industry	197,471 transactions	GBM	95%
Shabir et al.[3]	Banking institutions	Not specified	GBM	96%
Khare et al. [1]	Kaggle	284,807 transactions	DT	99%
Our	Kaggle	6,353,307 transactions	RF	100%

avoidance and reduces false positives. This comparison underscores the significance of dataset size and model selection. While previous studies achieved commendable accuracy, our approach demonstrates a distinct advantage by effectively handling an extensive dataset, ensuring reliability and real-world applicability.

5 Conclusion

This study emphasizes the efficiency of ML models in spotting fraudulent financial transactions, where the RF model attained the best accuracy, precision, and recall. This makes the most dependable method. Significant difficulties such as managing imbalanced datasets, adjusting to changing fraud strategies, and optimizing computational efficiency for real-time detection are highlighted in the paper. Using a dataset of almost six million transactions, the results validate the scalability and dependability of AI-driven fraud detection in financial security. The study also highlights the significance of minimizing false positives to maintain high detection accuracy and mitigate changes in legitimate transactions. Future studies should investigate expert DL models like LSTMs and Transformers to improve fraud detection capacities even more. Adaptive learning methods can also enable models to change to dynamically fit newly emerging fraud patterns of behavior. Additionally, federated learning methods that facilitate safe group fraud detection while maintaining user privacy should be considered. Finally, explainable artificial intelligence (XAI) methods will help to increase trust in automated fraud detection systems, regulatory compliance, and honesty.

References

[1] Khare, P., Srivastava, S.: Ai-powered fraud prevention: A comprehensive analysis of machine learning applications in online transactions. J. Emerg. Technol. Innov. Res **10**(9), 2349–5162 (2023)

[2] Olushola, A., Mart, J.: Fraud detection using machine learning. ScienceOpen Preprints (2024)

[3] Shabir, G., Khalid, N.: AI-Powered Fraud Detection and Risk Assessment: The Future of Financial Services

[4] Mujtaba, N., Yuille, A.: Ai-powered financial services: Enhancing fraud detection and risk assessment with predictive analytics

[5] Ali, A.A., Khedr, A.M., El-Bannany, M., Kanakkayil, S.: A powerful predicting model for financial statement fraud based on optimized xgboost ensemble learning technique. Applied Sciences **13**(4), 2272 (2023)

[6] Vaquero, P.R.: Literature review of credit card fraud detection with machine learning (2023)

[7] Abiodun, O.I., Jantan, A., Omolara, A.E., Dada, K.V., Mohamed, N.A., Arshad, H.: State-of-the-art in artificial neural network applications: A survey. Heliyon **4**(11) (2018)

[8] Johora, F.T., Hasan, R., Farabi, S.F., Akter, J., Al Mahmud, M.A.: Ai-powered fraud detection in banking: Safeguarding financial transactions. The American journal of management and economics innovations **6**(06), 8–22 (2024)

[9] Ismaeil, M.K.A.: Harnessing ai for next-generation financial fraud detection: A datadriven revolution. Journal of Ecohumanism **3**(7), 811–821 (2024)

[10] Kotha, R.: Ai-powered fraud detection in financial services. J Artif Intell Mach Learn & Data Sci 2022 **1**(1), 1337–1341

[11] Aziz, L.A.-R., Andriansyah, Y.: The role artificial intelligence in modern banking: an exploration of ai-driven approaches for enhanced fraud prevention, risk management, and regulatory compliance. Reviews of Contemporary Business Analytics **6**(1), 110–132 (2023)

[12] Narsina, D., Gummadi, J.C.S., Venkata, S., Manikyala, A., Kothapalli, S., Devarapu, K., Rodriguez, M., Talla, R.: Ai-driven database systems in fintech: Enhancing fraud detection and transaction efficiency. Asian Accounting and Auditing Advancement **10**(1), 81–92 (2019)

[13] Bello, O.A., Ogundipe, A., Mohammed, D., Adebola, F., Alonge, O.A.: Ai-driven approaches for real-time fraud detection in us financial transactions: challenges and opportunities. European Journal of Computer Science and Information Technology **11**(6), 84–102 (2023)

[14] Eteng, I.E., Chinedu, U.L., Ibor, A.E.: A stacked ensemble approach with resampling techniques for highly effective fraud detection in imbalanced datasets. Journal of the Nigerian Society of Physical Sciences, 2066–2066 (2025)

[15] Khurshid, M.R., Manzoor, S., Sadiq, T., Hussain, L., Khan, M.S., Dutta, A.K.: Unveiling diabetes onset: Optimized xgboost with bayesian optimization for enhanced prediction. PloS one **20**(1), 0310218 (2025)

[16] Kousar, F., Sultana, A., Albahar, M.A., Shamkuwar, M., Heyat, M.B.B., Hayat, M.A.B., Parveen, S., Lira, J.I.G., Rahman, K., Alammari, A., *et al.*: A cross-sectional study of parental perspectives on children about covid-19 and classification using machine learning models. Frontiers in Public Health **12**, 1373883 (2025)

[17] Prova, N.N.I.: Healthcare fraud detection using machine learning. In: 2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI), pp. 1119–1123 (2024). IEEE

[18] Gøttcke, J.M., Zimek, A., Campello, R.J.: Bayesian label distribution propagation: A semi-supervised probabilistic k nearest neighbor classifier. Information Systems **129**, 102507 (2025)

[19] Oguntibeju, O., Adonis, M., Alade, J.: Systematic review of real-time analytics and artificial intelligence frameworks for financial fraud detection. International Journal of Advanced Research in Computer and Communication Engineering **13**(9) (2024)