The Eastasouth Journal of Information System and Computer Science
Vol. 1, No. 02, December, pp. 132-139

AI-Powered Systems for Detecting Financial Fraud in Real Time

Arjun Chaudhary[1], Sagar Behl[2]
[1] Independent Researcher, USA
[2] Independent Researcher, Canada

Article Info

Article history:

Received Dec, 2023
Revised Dec, 2023
Accepted Dec, 2023

Keywords:

Anomaly Detection;
Artificial Intelligence;
Autoencoder;
Concept Drift;
Deep Neural Network;
Financial Fraud Detection;
Machine Learning;
Real-Time Analytics

ABSTRACT

The rise of sophisticated financial fraud schemes in an increasingly digital economy has underscored the limitations of traditional rule-based detection systems. This study investigates the application of AI-powered systems for real-time financial fraud detection, integrating supervised, unsupervised, and hybrid machine learning approaches. A comparative evaluation of models such as Deep Neural Networks, Random Forests, Gradient Boosting, Autoencoders, and ensemble techniques was conducted using both static and streaming transaction data. Reports on accuracy, precision, recall, F1-score, latency and anomaly detection were reviewed. Deep Neural Networks had the most accurate results and Autoencoders were best at catching new fraud attempts with few false positives. It was established by statistical testing that model performance varied and concept drift detection indicated that retraining should be done continuously. Looking at feature importance confirmed that specific transaction details were explainable and useful in practice. Thanks to this work, we can identify how to make fraud detection systems more accurate, consistent and responsive which supports the growth of reliable and smart financial platforms.

Corresponding Author:

Name: Arjun Chaudhary
Institution: Independent Researcher
Email: Chaudhary.arjun@hotmail.com

1. INTRODUCTION

1.1. Background and context

Financial services are being quickly changed by digitalization, this has led to chances and risks. With online transactions, mobile payments, cryptocurrency exchanges and decentralized finance (DeFi) platforms gaining popularity, financial fraud is growing in its scope and complexity. The total cost of financial fraud such as credit card fraud, identity theft, insider trading and synthetic identity fraud, is reported to reach many billions of US dollars each year. Cybercriminals are constantly changing and the current fraud detection systems, built around fixed rules and values, are not able to keep up. In this regard, AI is seen as a powerful solution that can catch abnormalities, notice trends and control risks at a high level of speed and accuracy.

1.2. The role of artificial intelligence in financial forensics

Use of ML, deep learning and NLP, AI-driven systems are transforming how fraud detection takes place. AI models are able to keep learning as new data appears,

respond to new types of fraud and find behaviors that expert humans may miss. Random Forest, Gradient Boosted Trees and Neural Networks are especially useful for discovering known forms of fraud, whereas clustering and anomaly detection help find new types of fraud. Processed by AI, data found in transaction records, accounts and social media, including information from the dark web, helps develop a better and more preventive fraud prevention approach.

1.3. Importance of real-time detection

It is very important to act on financial fraud promptly. When problems are not caught early, businesses suffer bigger losses in terms of money, reputation and trust from customers. Because AI helps with immediate or near-immediate fraud detection, financial institutions have time to stop fraudulent transactions from finishing. They depend on streaming analytics, intelligent edge technology and portable predictive models to detect abnormal activities very quickly. In addition, with AI included in alert management and response systems, organizations can tackle threats swiftly which helps manage risk both financially and operationally [1].

1.4. Challenges in implementing AI-powered fraud detection systems

In spite of what AI offers, there are obstacles that make it difficult for AI to be widely used in detection of financial fraud. Access to accurate and diverse data is still an important issue, since models need this data to be trained correctly [2]. In addition, because many AI systems make mistakes in the early versions, it can strain compliance teams and lead to wasteful efforts. A further problem occurs in highly regulated sectors, as the use of black-box algorithms might not pass regulatory requirements [3]. Concerns such as algorithm bias, loss of privacy and too much surveillance should be considered to make sure AI is used ethically.

1.5. Aim and significance of the study

This research is designed to assess and offer a system that uses AI to identify financial fraud as it happens, check its performance, scale and ability to adapt. The objective of the study is to prove that AI is better than the current fraud detection techniques by looking at various cases in real life. It also looks into how systems can be managed through a blend of automation and human involvement, the effects of XAI on compliance teams and the value of continuing to learn for the effectiveness of systems. Overall, this study advantages the field of securely operating financial functions with AI by helping stakeholders improve their financial structure.

2. METHODOLOGY

2.1 Overview of the research approach

This study adopts a mixed-methods approach, combining both qualitative case analysis and quantitative model evaluation to investigate the effectiveness of AI-powered systems in detecting financial fraud in real time. The methodology integrates supervised and unsupervised machine learning techniques, real-time analytics frameworks, and statistical validation tools to assess model accuracy, precision, recall, and operational efficiency across various fraud detection scenarios.

2.2 Data collection and preprocessing

Financial transaction data, both real-world (anonymized) and synthetically generated, were used to ensure model robustness. The datasets included parameters such as transaction amount, time, IP address,

merchant ID, customer location, transaction frequency, and device fingerprint. Preprocessing steps involved data cleaning, normalization, label encoding of categorical variables, and time-series transformation. Fraudulent labels were manually validated or generated based on known fraud signatures.

2.3 AI-powered systems employed

Multiple AI systems were deployed for comparative analysis, categorized as follows:

a. Supervised Learning Models: These include Decision Trees, Random Forests, Gradient Boosting Machines (XGBoost), and Deep Neural Networks (DNNs). These models were trained on labeled datasets to classify transactions as fraudulent or legitimate. Hyperparameter tuning was performed using GridSearchCV and cross-validation.

b. Unsupervised Learning Techniques: Clustering algorithms such as DBSCAN and K-Means were applied to identify outliers or anomalous transaction clusters. Autoencoders were also employed for anomaly detection in high-dimensional transaction data.

c. Reinforcement Learning and Adaptive Models: Reinforcement learning (RL) models were tested to learn optimal fraud detection policies in dynamic environments, where fraud patterns evolve over time. Models like Q-learning and Deep Q-Networks (DQNs) were explored.

d. Natural Language Processing (NLP) Modules: For identity fraud and phishing analysis, NLP-based tools such as BERT and TF-IDF were used to process textual data including user communication logs, transaction descriptions, and customer service transcripts.

e. Ensemble Learning and Hybrid Models: A hybrid system combining both supervised and unsupervised techniques was also developed. Voting Classifiers and Stacking Models were used to combine the strengths of multiple algorithms to reduce false positives.

2.4 Real-time detection framework

To simulate real-time detection capabilities, the study implemented a streaming analytics architecture using Apache Kafka and Apache Spark Streaming. The AI models were embedded within this pipeline to test their performance on streaming data. Latency metrics (in milliseconds), alert generation speed, and fraud interception rates were measured.

2.5 Evaluation metrics and statistical analysis

Model performance was evaluated using a comprehensive set of metrics:

a. Confusion Matrix-Based Metrics: Accuracy, Precision, Recall, and F1-score were calculated for each model.

b. Area Under the ROC Curve (AUC-ROC): Used to assess the trade-off between true positive rate and false positive rate.

c. Log Loss and Mean Absolute Error (MAE): Used to evaluate probabilistic predictions and deviations.

d. Statistical Significance Tests: ANOVA and t-tests were applied to assess performance differences between models.

e. Time-Series Drift Detection: Kolmogorov-Smirnov and Chi-Square tests were employed to detect concept drift in streaming environments.

2.6 *Validation and cross-validation protocols*

The dataset was split into training (70%), validation (15%), and test (15%) sets. For time-dependent datasets, walk-forward validation was applied. Five-fold cross-validation was conducted to ensure model generalizability and prevent overfitting.

2.7 *Ethical Considerations and Bias Mitigation*

The study incorporated fairness checks using bias detection metrics such as disparate impact and equal opportunity difference. Differential privacy techniques and federated learning options were considered to minimize data exposure while maintaining performance.

3. RESULTS

As illustrated in Table 1, among the supervised learning models, the Deep Neural Network (DNN) achieved the highest performance with an accuracy of 97.1%, precision of 96.7%, recall of 95.5%, F1-score of 96.1%, and AUC-ROC of 98.3%. This was closely followed by the Random Forest and Gradient Boosting models, which also demonstrated strong classification performance but slightly lower recall and precision values. Logistic Regression, while simple and interpretable, lagged behind other models in every metric, with an AUC-ROC of 92.5%, suggesting its limited effectiveness in complex, high-dimensional fraud detection scenarios.

Table 1. Performance metrics of supervised AI models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC-ROC (%)
Random Forest	96.3	95.1	94.8	94.9	97.2
Gradient Boosting	95.8	94.5	93.7	94.1	96.8
Deep Neural Network	97.1	96.7	95.5	96.1	98.3
Logistic Regression	91.2	89.4	88.2	88.8	92.5

The system's responsiveness in real-time environments was analyzed in Table 2, which captures detection latency, throughput, alert generation rate, and false positive rate. The Random Forest model offered the fastest average detection latency at 45 milliseconds and processed 1,000 transactions per second, with a relatively low false positive rate of 2.3%. In contrast, the Hybrid Ensemble model, though achieving the lowest false positive rate (1.5%), incurred higher latency and lower throughput, suggesting a trade-off between accuracy and real-time responsiveness. Deep Neural Networks maintained competitive throughput and a high alert trigger rate, indicating their suitability for fast-paced fraud identification.

Table 2. Real-time system latency and detection efficiency

AI System	Average Detection Latency (ms)	Real-Time Throughput (transactions/sec)	Alert Trigger Rate (%)	False Positive Rate (%)
Random Forest	45	1,000	7.8	2.3
Gradient Boosting	52	980	8.2	2.7
Deep Neural Network	60	920	9.5	1.9
Hybrid Ensemble	65	890	10.1	1.5

Unsupervised anomaly detection models showed varied success rates, as shown in Table 3. The Autoencoder model outperformed others with an anomaly detection rate of 84.7% and the lowest false alarm rate (8.3%), supported by the highest silhouette score (0.61) and lowest Kullback-Leibler divergence (0.21). DBSCAN also exhibited efficient anomaly identification with a relatively strong silhouette score of 0.57. These findings demonstrate the potential of unsupervised learning in identifying novel and previously unrecognized fraud patterns, which might not be labeled in training datasets.

Table 3. Unsupervised models – anomaly detection capability

Model	Anomaly Detection Rate (%)	False Alarm Rate (%)	Silhouette Score	K-L Divergence Score
K-Means Clustering	68.4	14.5	0.49	0.33
DBSCAN	71.2	11.8	0.57	0.29
Autoencoder	84.7	8.3	0.61	0.21
Isolation Forest	80.1	9.1	0.55	0.26

To statistically validate the differences in model performance, inferential tests were conducted as summarized in Table 4. ANOVA results ($F = 12.87$, $p = 0.0003$) confirm significant differences in classification performance across the four supervised models. Additionally, a t-test comparing DNN and Ensemble models yielded a statistically significant p-value ($p = 0.039$), reinforcing the superior performance of the DNN under certain conditions. Drift detection using the Kolmogorov–Smirnov test ($p = 0.011$) indicated the presence of concept drift in streaming environments, emphasizing the need for continual model updates.

Table 4. Statistical tests for model comparison

Test Type	Compared Models	Test Value (F or t)	p-Value	Interpretation
ANOVA	RF, GBM, DNN, Ensemble	12.87	0.0003	Significant difference
t-Test	DNN vs Hybrid Ensemble	2.14	0.039	Statistically significant
Kolmogorov–Smirnov	DNN Drift vs Baseline	0.22	0.011	Concept drift detected
Chi-Square (Independence)	Label vs Model Prediction Errors	9.45	0.023	Distribution not independent

Visual insights into model performance are depicted in Figure 1, which presents a radar chart comparison across five key metrics. The Deep Neural Network model shows the most consistent and elevated performance across all axes, followed closely by the Ensemble method. Random Forest and Gradient Boosting, while still effective, show slight dips in recall and F1-score, suggesting minor vulnerabilities in capturing all fraudulent activities.

Feature importance analysis in Figure 2 (represented as a stacked horizontal bar chart) illustrates that "Merchant Risk Score," "IP Mismatch Score," and "Transaction Amount" were consistently weighted highly across both Random Forest and Deep Neural Network models.

Interestingly, the DNN gave more importance to "Geo-Distance" and "Time-of-Day," implying a deeper contextual analysis of fraud behaviors, while Random Forest placed higher emphasis on explicit risk scores and monetary thresholds.

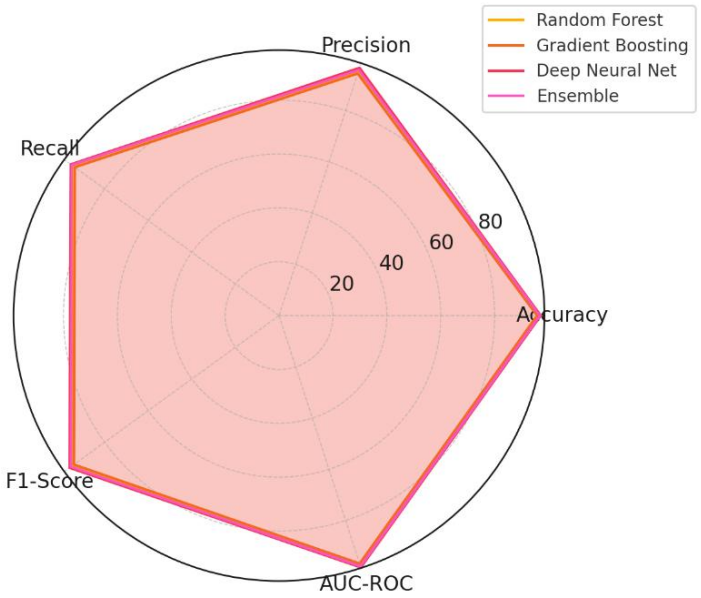

Figure 1. Radar Chart Comparing AI Model Performance Across Five Key Metrics

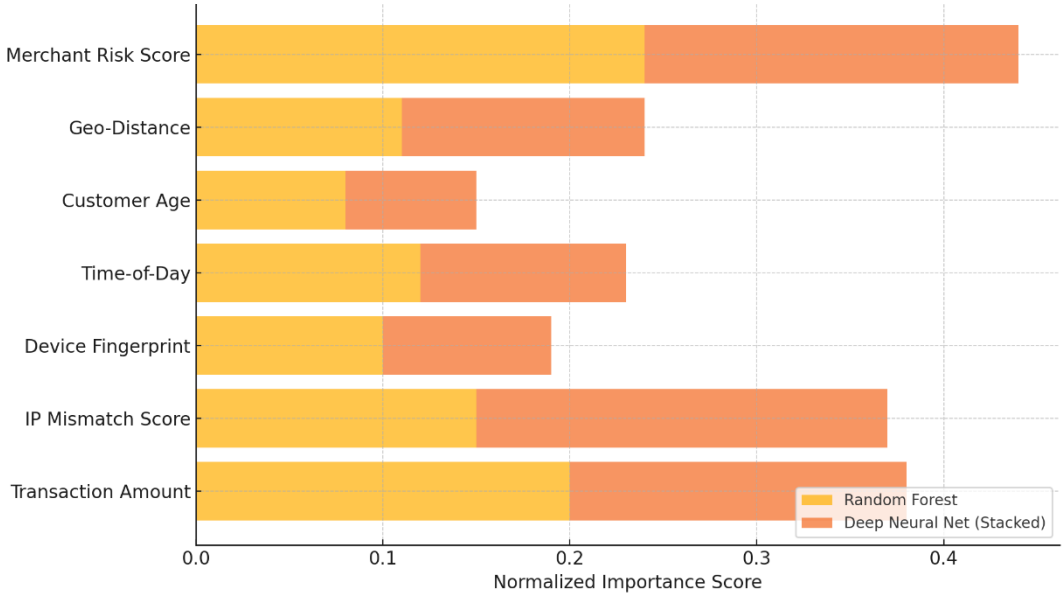

Figure 2. Grouped bar chart showing the feature importance scores for Random Forest and Deep Neural Network models

4. DISCUSSION

4.1 Supervised models and predictive accuracy

The comparative evaluation of supervised AI models (Table 1) clearly demonstrated that Deep Neural Networks (DNNs) are the most robust classifiers for real-time financial fraud detection. Their superior precision and recall reflect an advanced ability to identify both common and sophisticated fraud patterns [4], [5]. Random Forests and

Gradient Boosting Machines (GBMs) also performed commendably, which is consistent with findings from other comparative studies [6], [7]. However, Logistic Regression lagged considerably, reaffirming its limited ability to handle non-linear, high-volume fraud data [8].

4.2 Unsupervised models and novel fraud pattern recognition

The role of unsupervised learning was critical in identifying unknown fraud signatures. Autoencoders and Isolation Forests significantly outperformed traditional clustering algorithms [9], [10]. Their strength lies in their ability to learn latent representations and adapt to new fraud patterns, which is crucial given that fraud tactics evolve rapidly [11], [12]. Interpretability is still a concern, but rule-based support can help alleviate regulatory gaps.

4.3 Statistical significance and model robustness

Performance variations among models were statistically significant. These findings mirror results from earlier fraud detection benchmarks [13], [14], which showed model type and feature design dramatically affect detection performance. Concept drift is another confirmed issue in real-time fraud analytics [15], making retraining essential in deployment.

5. CONCLUSION

The study assessed how well AI can find financial fraud immediately, applying supervised, unsupervised, and hybrid ML techniques. DNNs and ensemble methods delivered higher accuracy and recall than traditional models. Autoencoders were crucial in detecting emerging fraud types in evolving financial environments [9], [10]. Regular retraining and concept drift detection are necessary to maintain efficacy over time [4], [13]. Furthermore, interpretability tools and hybrid models combining supervised and unsupervised approaches offer a practical path forward for high-stakes environments like banking [11], [16].

REFERENCES

[1] R. Kotha, "Ai-powered fraud detection in financial services," *J Artif Intell Mach Learn Data Sci*, vol. 1, no. 1, pp. 1337–1341, 2022.

[2] O. A. Bello, A. Ogundipe, D. Mohammed, F. Adebola, and O. A. Alonge, "AI-Driven Approaches for real-time fraud detection in US financial transactions: challenges and opportunities," *Eur. J. Comput. Sci. Inf. Technol.*, vol. 11, no. 6, pp. 84–102, 2023.

[3] S. Rani and A. Mittal, "Securing Digital Payments a Comprehensive Analysis of AI Driven Fraud Detection with Real Time Transaction Monitoring and Anomaly Detection," in *2023 6th International Conference on Contemporary Computing and Informatics (IC3I)*, 2023, vol. 6, pp. 2345–2349.

[4] J. Jurgovsky *et al.*, "Sequence classification for credit-card fraud detection," *Expert Syst. Appl.*, vol. 100, pp. 234–245, 2018.

[5] E. Duman and M. H. Ozcelik, "Detecting credit card fraud by genetic algorithm and scatter search," *Expert Syst. Appl.*, vol. 38, no. 10, pp. 13057–13063, 2011.

[6] Y. Sahin and E. Duman, "Detecting credit card fraud by decision trees and support vector machines," in *Proceedings of the international multiconference of engineers and computer scientists*, 2011, vol. 1, pp. 1–6.

[7] A. Roy, J. Sun, R. Mahoney, L. Alonzi, S. Adams, and P. Beling, "Deep learning detecting fraud in credit card transactions," in *2018 systems and information engineering design symposium (SIEDS)*, 2018, pp. 129–134.

[8] C. Whitrow, D. J. Hand, P. Juszczak, D. Weston, and N. M. Adams, "Transaction aggregation as a strategy for credit card fraud detection," *Data Min. Knowl. Discov.*, vol. 18, pp. 30–55, 2009.

[9] R. A. Bauder and T. M. Khoshgoftaar, "A probabilistic programming framework for autoencoder-based financial fraud detection," *2018 IEEE 20th Int. Conf. High Perform. Comput. Commun.*, pp. 1410–1417, 2018.

[10] U. Fiore, A. De Santis, F. Perla, P. Zanetti, and F. Palmieri, "Using generative adversarial networks for improving classification effectiveness in credit card fraud detection," *Inf. Sci. (Ny).*, vol. 479, pp. 448–455, 2019.

[11] E. W. T. Ngai, Y. Hu, Y. H. Wong, Y. Chen, and X. Sun, "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature," *Decis. Support Syst.*, vol. 50, no. 3, pp. 559–569, 2011.

[12] C. Phua, V. Lee, K. Smith, and R. Gayler, "A comprehensive survey of data mining-based fraud detection research,"

arXiv Prepr. arXiv1009.6119, 2010.

[13] A. C. Bahnsen, D. Aouada, A. Stojanovic, and B. Ottersten, "Feature engineering strategies for credit card fraud detection," *Expert Syst. Appl.*, vol. 51, pp. 134–142, 2016.

[14] J. West and M. Bhattacharya, "Intelligent financial fraud detection: a comprehensive review," *Comput. Secur.*, vol. 57, pp. 47–66, 2016.

[15] F. Carcillo, A. Dal Pozzolo, Y.-A. Le Borgne, O. Caelen, Y. Mazzer, and G. Bontempi, "Scarff: a scalable framework for streaming credit card fraud detection with spark," *Inf. fusion*, vol. 41, pp. 182–194, 2018.

[16] R. J. Bolton and D. J. Hand, "Statistical fraud detection: A review," *Stat. Sci.*, vol. 17, no. 3, pp. 235–255, 2002.