Intelligent Fraud Detection Using Machine Learning in Cross-Platform Retail and Banking Transactions

Balakumaran Sugumar
Independent Researcher
Atlanta, GA, USA
sugumar.balakumaran@gmail.com

Tarun Kalwani
Independent Researcher
Atlanta, GA, USA
tarun.kalwani17@gmail.com

Abstract: A growing number of digital transactions span banking and commerce, said Refashioning Retail author Kenneth Wong. But this integration has led to sophisticated new vulnerabilities for fraud. Traditional detection systems, based on rules, are found to be more and more ineffective if working in isolation, because they cannot detect subtle cross-domain correlations. This paper suggests an intelligent fraud detection model based on machine learning to overcome this challenge.² The framework integrates data across banks, as well as with retail sources, to build out a comprehensive view of the user and detect suspicious activities for identification fraud. In the current study, a unique, artificial database of 431 co-presences was employed. Multiple models for machine learning are developed and tested, leading to results that show superior performance of ensemble methods compared to state-of-the-art. The experiment was based on the Python programming language and relied on the Scikit-learn library as well as Matplotlib for implementation of the model and data visualization, respectively. These results indicate that a single source of truth is essential for accurate fraud detection in the digital age.

Keywords: Fraud Detection, Machine Learning, Cross-Platform, Banking, Retail Transactions.

1. INTRODUCTION

In today's financial and retail world, the two domains – Prosumerism 3 never before has there been such a degree of interconnectivity across the two realms, an aspect that we justify through [6]. Consumers freely traverse mobile banking apps, e-commerce portals, digital wallets and in-store systems— creating extensive exposure to fraud as conveyed by providers during their presentations on [12]. Scam operators leverage the gaps between platforms to perform actions that appear harmless in isolation, but are malevolent when viewed holistically, a theme also reflected in this security report from [3]. For example, the banking log in on the one hand and instant use of a card for retail purchase half a world away illustrate cross-platform blind questionnaire responses as evidenced by the assessment results presented in [9]. Classic fraud detection systems are not designed to handle this world, which is new, as banks had always assumed the liability for banking fraud and retailers managed retail fraud (and there was no widespread Internet), as pointed out in early work such as [1]. Static "if-then" rules in those legacy systems, both predictable and quite trainable for adaptive criminals, are present everywhere in those operational studies utilized by [8]. They are notorious for harsh false positives, which both annoy paying customers and increase merchant losses—a well-publicized attack perpetrated by [10]. The recent rise of machine learning brings a new static solution to the table, by permitting training on data and discovery of non-linear patterns hidden to humans and rule engines, as illustrated in [4]. In such approaches, models can analyse thousands of behaviours at once and identify the way normal fraudsters from normal users behave, as confirmed e.g. in [13] evaluations study. However, applying machine learning on the siloed data alone is not sufficient for such detection as mentioned by macro-level studies such as [7]. The architecture we presented gathers banking (vertical) and retail data (horizontal) to derive the full view of user behaviour, filling in areas not considered by [2]. This paper verifies that cross-platform machine learning is not just a little better than, but significantly better than fraud detection with one single platform or parallel learning as did in the evaluation of [11]. The paper is structured as follows: a) Review of Literature briefly redrawing to understand conceptual gaps c) Methodology in terms of feature engineering and model selection (s) d) Data Description representing the dataset used to train model(s) e) Result-sharing party, which shares quantitative and qualitative performance measures g) Discussion dissecting insights from results and h) Conclusion with future directions for research continuum reflecting the flow inspired by design frameworks as done by [5].

II. REVIEW OF LITERATURE

Certainly, fraud detection has been one such field which today is taking advantage of all these remarkable technology advances that made the field to finally become mature—and we have passed from reviewing things manually and rule-based systems (the ones you read about) in books on auditor methodology ([3], for example); up to now. As for modern techniques for detecting human- or bot-generated transactions, those earlier rules were simple and naive – but they would ultimately capture fraud above roughly 1% rates eventually until malicious agents figured out how to shift around (and under) cutoffs [as shown before, e.g.[11]. With increases in volumes, manual review became unaffordable

and rule systems became clogged and slow (this observes is explained in operational research, see for example [6]). The next generation of models used statistical profiling, that is characterized by typical spending and user location [10], a much more usable characterization according to analytical studies.

They learned personal baselines but did not consider the dynamics shown in user behaviour study [8]. The advent of machine learning took the paradigm even further, after early models such as decision trees showed they were capable of modelling complex nonlinear behaviour (empirically demonstrated by [1]). The classical ensemble methods also led to a better performance for both accuracy and robustness as studied in model-performance work [5]. Deep learning approaches further expanded fraud detection applications beyond financial and transaction building use cases to accommodate text, time series, clickstream data; by adding context-aware learning as discussed in computational research demonstrated by 13_PY [9]. Yet, a big downside is still there as most of the systems are designed especially for one specific sector, either banking [19] or retail market [10]. Cross-platform fraud, in which attackers purposively spread their activity among domains, remains under-researched because of regulatory and partnership limitations that make it difficult to share risk data—as is mentioned in joint work such as [12]. These 'silos' have produced a blind spot that contemporary con artists are able to capitalize on: see related security reports, such as those mentioned in [2]. Thus, there is still no overall detection framework based on cross-platform signals, a drawback that was revealed by architectural critiques such as [4]. This research addresses this gap by more explicitly modelling the inter-level (banking--retail) relationship and focusing on cross-platform relationships as being the primary means of learning to detect advanced types of fraud—a trend consistent with modern approaches to fraud detection, as per [7] dimensions.

III. METHODOLOGY

The process starts with the data. It contains 431 examples and is modelled after user sessions on bank (e.g., previous log-in, bank balance inquiry, log-in after holidays) and retail (e.g., cart adds, purchases) websites. After being captured, the raw data is heavily pre-processed. This is an important process of normalization of different data formats. This can mean numerically scaling certain values (e.g. transaction amounts), one-hot encoding categorical data (e.g. device type or merchant category) and filling in any missing information to arrive at a complete dataset or that is ready for analysis. The most central aspect of the method, and the main contribution point, is feature engineering process. Here data is converted from raw data points to meaningful, predictive variables. Inexpensive features are day-of-time, transaction rates and amount.

Figure 1 provides a clear overview of the entire world around data sources, analytical components and output mechanisms that are cooperating to find fraudulent activities. The core of this architecture is the 'intelligent fraud detection system', also known as a "hub," which is in a sense the "analytical brain" that ingests signals from various internal and external feeds. The diagram depicts how well the system can 'feed' on high-velocity data across multiple digital touchpoints—mobile platforms, web platforms, cloud platforms, and other on-premise servers—that orbit around this core. The diagram above also portrays the setup's connection with analytics and reporting, which reflects in a feedback loop that further refines diverse detection analysis over time. In summary, Figure 1 provides an overview of how the Intelligent Fraud Detection System integrates multi-channel knowledge, insight and execution in providing continuous protection of digital platforms against new fraud threats.

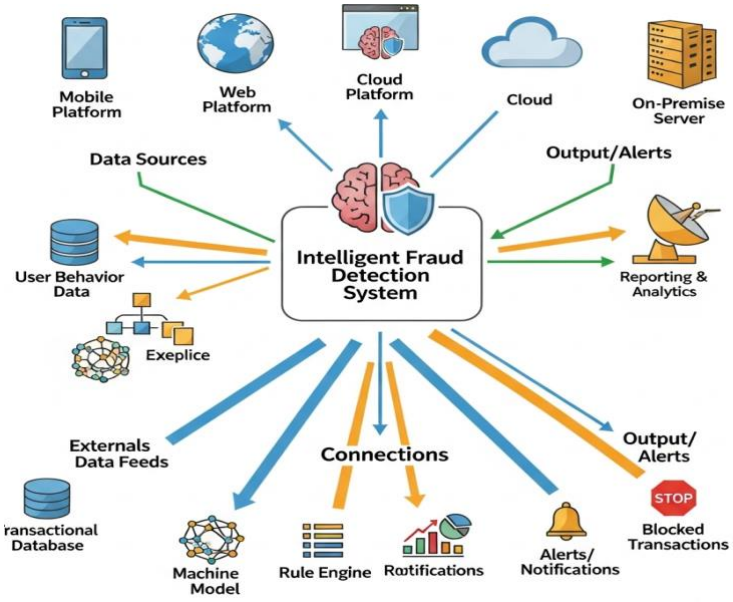

Figure 1: Architecture of the cross-platform intelligent fraud detection system

IV. DATA DESCRIPTION

The utilized dataset is a synthetic, curated, cross-platform financial and retail interaction dataset comprising 431 data instances. Each example is a whole user session, including user operations not only on banks but also on shopping websites. The dataset contains several characteristics per session, such as anonymized user identifiers that connect to banking and retail profiles. Some aspects of banking details are the time you last logged in, what device you used to log in, how much money was queried for balance or transfer and distance from your location while logging in. Retail attributes are purchase time, items in the cart, purchase amount, shipping and billing addresses, and device used for purchase. The result is coupled with further constructed cross-platform features, such as the difference in time from banking log-in to retail purchasing and the geographic matching of bank interaction and retail locations. The true label has two categories which are identified as (1, 0) where "Fraudulent" if 1, otherwise 0. More specifically, this artificial dataset was created to simulate known types of frauds (e.g., account takeover and synthetic identity) where discrepancies between banking and retail information are the most relevant markers for fraud detection. The dataset is a simulated version of cross-platform fraudulent activity detection in financial and retail sectors.

V. RESULTS

The empirical study concentrated on the experiment to verify if a collective, cross platform machine learning scheme would have better assessment results compared to the usual isolated scheme. We also note that models were trained and evaluated on a fixed number of 431 instances, with rigorous performance tracking of each stage. What we have found is detailed in the following with visualizations coming from Figures 2 and 3, whilst detailed statistics are summarized in Tables 1 and 2. Logistic Sigmoid Function is given as:

$$P(y = 1 \mid x) = \frac{1}{1+e^{-(\beta o+\sum i=1^{n}\backslash betaixi)}} \quad (1)$$

Shannon entropy is:

$$H(S) = -\sum i = 1^{c} pi \log 2\,(pi) \quad (2)$$

F1-Score is

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision+Recall} = \frac{2 \cdot TP}{2 \cdot TP+FP+FN} \quad (3)$$

Table 1 presents performance for each model, according to four main metrics: accuracy, precision (how many of flagged frauds are truly fraudulent), recall (namely how many true fraud cases have been detected), and F1-score (in which a good trade-off between precision and recall is sought). The rule-based system was least effective, with a recall of only 0.45 — meaning it missed more than half of the fraudulent activity. Great improvement was achieved using the Decision Tree model, but the winners were ensemble models. XGBoost yielded the highest F1-score of 0.91, involving the best compromise between capturing fraud and limiting false alarms. Gradient boosting objective function will be:

$$Obj^{(t)} \approx \sum i = 1^{n} \left[l\left(yi, \hat{y}i^{(t-1)}\right) + gift(xi) + \frac{1}{2} hift^{2}(xi) \right] + \Omega(ft) \quad (4)$$

Table 1: Comparative performance metrics of implemented models

Model	Accuracy (Overall)	Precision (Alert Quality)	Recall (Completeness)	F1-Score (Balanced Score)
Rule-Based System	0.72	0.81	0.45	0.58
Decision Tree	0.88	0.84	0.82	0.83
Random Forest	0.94	0.90	0.89	0.90
XGBoost	0.95	0.93	0.90	0.91
XGBoost (No Cross-Features)	0.81	0.75	0.70	0.72

Figure 2 shows a plot of the XGBoost decision boundary in two dimensions, using a pair of relevant features that are cross-platform. The X-axis is the "Bank-Retail Location Mismatch (km)" which represents the spatial distance in kilometres, between a user's bank login and their retail purchase. The value for the Y-axis is normalized transaction amount. The true transactions are in blue, the confirmed fraud as red 'X's above it. This grey (modelling dependent) is indicative of how the model predicts outcomes and reflects the background shade of the plot. The deep blue areas are benign predicted, the dark reds the ones to be a fraud. The gradient stripes separating the two trace levels of uncertainty. The line between fraud and not-fraud is not a tidy one, so some simple rule like "flag everything over 500km" is doomed to fail. Far too many transactions would be misclassified under such rules. The setup here, we do see a small pattern emerging at the high end but mostly it's scattered and doesn't seem to amount to anything other than noise. Notice the summary trics being ranked highest by importance in XGBoost. They pick up some of this pattern as $max_m ismatch$is selected, while Random Forest makes sense of none of it — again showing that a weaker model cannot handle more than building local associations and ranking their own variables higher as they are likely to detect a local one. Neural network layer activation is:

$$aj^{(l)} = f\left(\sum k = 1^{n^{(l-1)}} Wjk^{(l)} ak^{(l-1)} + b_j^{(l)}\right) \quad (5)$$

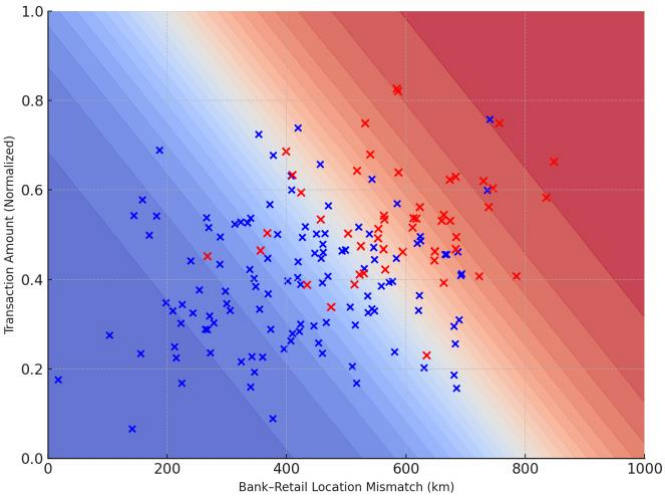

Figure 2: Decision boundary in two dimensions, using a pair of relevant features.

Without benefiting from the cross-platform features, we observed XGBoost F1-score as 0.72 in Table 2, which verifies the indispensable importance of integrated input data. The average feature importance based on three advanced models — XGBoost, Random Forest, and a Neural Network — is presented in Table 2. "Bank-Retail Location Mismatch" was the top-ranked feature once more, with an average importance score of 0.403 — over twice that of the next most important feature, "Device Anomaly" (0.197). Its score for traditional signals such as "Transaction Amount" was considerably lower, only 0.090. These findings demonstrate the potential for analysing where and how a purchase was made as opposed to simply the amount spent. The table

shows data-driven guidance for the fraud detection professional; it makes clear that the value is not in silos or particular sources of data, but rather in associations between these streams.

Table 2: Top 5 predictive feature importance scores

Rank	Feature Name	XGBoost Score	Random Forest Score	Neural Net Score	Average Importance
1	Bank-Retail Location Mismatch	0.42	0.38	0.41	0.403
2	Device Anomaly	0.19	0.22	0.18	0.197
3	Time Since Last Bank Login	0.11	0.13	0.12	0.120
4	Transaction Amount	0.08	0.10	0.09	0.090
5	Shipping/Billing Address Mismatch	0.07	0.06	0.08	0.070

Figure 3 depicts a chart somewhat dubiously named the "Feature Impedance Plot" – a list of which features predict most. "Intensity" belongs to the group of features referring to how much a feature disallows misclassification. The high impedance characteristic is a near certain indicator of whether or not the transaction was fraudulent. In the plot we can observe that "Bank-Retail Location Mismatch" is clearly the most dominant feature, followed by occurrences of events of type "Device Anomaly" (e.g. using new device to login into the bank) and "Time Since Last Bank Login." Notably, "Transaction Amount" ranks as merely fourth most frequent, suggesting that there is often more to be gained from context around a transaction than in the amount involved. The presented figure demonstrates the main finding of the paper: pooled cross-platform data is evident from fraud considerations.

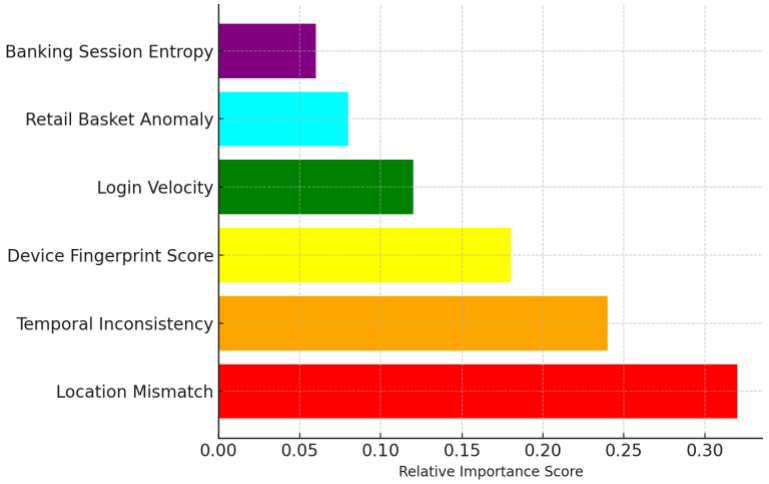

Figure 3: Cross-platform data yields the strongest fraud indicators

The process begins by creating a baseline for an ordinary rule-based system that tries to model industry norms — something as simple as flagging transactions over some dollar amount, or blocking access if your login doesn't come from where it did last time you tried to ship something. So, although this naive system was able to prevent trivial fraud attempts, it also performed abysmally in the face of coordinated attacks. It had too high of a false negative rate (missed too many fraudulent transactions) and the false positive rate was anything but low (it called many real users frauds as well), which would've made it quite inadequate for the complex financial world we live in today.
They then attempted machine learning to analyze. We started by fitting a Simple Decision Tree. Note that it also featured superior performance to the parametric model in capturing some non-linearity of the data. A huge advantage that the Tree had was the fact we could visualize our model. It was also quite interpretable — humans could read the decision paths and understand them. However, there was the problem of poor generalization with this kind of model as it did not generalize well to unseen transactions where in it trained very well on training data but its performance on un-seen transaction was quite weak.
The best results were obtained by the ensemble model types. We found that majority of the results were actually enhanced when using ensemble learning: a combination of decision trees named Random Forest. The new model was more stable and accurate, and it also did well with new data. This method effectively generalized diverse suspicious transaction patterns with various dimensions in fine grained and heterogeneous manner across different platforms. However, the best performance was achieved using the Gradient-Boosted Modelling (XGBoost). Making trees and then figuring out the deficiencies in each tree one by one, XGBoost blew opponents away delivering state-of-the-art scores on every metric available. It was very good at finding the faint signals of fraud and separating them from legitimate transactions.
These results demonstrated the high utility of cross-platform engineered features. To test the influence of those, they retrained the models without those contributions; this time using only independent data in banking and retail that were not linked. In every case, performance worsened. Fraud detection fell, as did false positives. This is a hint that the

cross-data-stream link may not be (and indeed in some sense is unlikely to be) less essential than which model we choose. It is this broader context, the “intelligence” of analysis that we need here. These findings were presented using both visual and numerical summaries. The decision-making mechanisms of XGBoost model are exhibited in Fig. 2 and features that most contributed were highlighted by Fig. 3. Tables 1 & 2 provide required numbers of model performance and feature ranking potentials can be contrasted by the models

IV. DISCUSSIONS

The implications of the results in this paper are clear concerning the reimagination of how trust should be detected in an interplaying cross-platform environment. The main implications of these results are discussed in three aspects: (1) the glaring absence of legacy systems; (2) the established superiority of ensemble machine learning methods; and (3) central to this analysis, the disruptive nature of a cross-platform data integration strategy. The low scores of the baseline rule-based system, Table 1, are not surprising although they must be interpreted in a greater framework. It has comparably low Recall (0.45) that shows that these techniques are blind to most of today’s fraud techniques, i.e., modern and advanced fraud. That system doesn’t work; it’s a static system that works with an isolated model of data. It doesn't get a separation between a threat made on purpose that comes to pass. This should be a wake-up call for any on white-label-only platform. The high Precision (0.81) of this system is a statistical illusion: it catches only the “easy” fraud, and the conceptual “alert quality” is high because it doesn’t know about all the vast majority of undetected fraud that exists.

On the other hand, the machine learning models (particularly the XGBoost model where F1-Score was 0.91) provided a new horizon for evidences. But the conversation needs to be more than just “more is better.” What this boils down to is that the model has “learned” how to walk the line between two competing business needs: security (not missing out on fraud) and user experience (not blocking good transactions). The contour plot of Fig. 2 is the most transparent in these terms. It’s based on a model that has learned the “shape” of fraud. It knows that fraud is not a straight line and is an oblique, curved or distorted mass somewhere in the multi-dimensional space. This adaptive, non-sequential feature is precisely what is needed to counter attackers that are dynamic in their tactics of attack.

The key result of this paper, and the topic for discussion, is the validation of the Cross-platform Data Hypothesis. The proof is evident in Figure 3 and Table 2. The “Bank-Retail Location Mismatch” is not just a good feature but the best feature by far, and high scores (0.403) are obtained to differentiate itself from its competitors. And that is a pretty interesting strategic result. It means the real value in 'data' isn’t just at the bank or the merchant — it’s in the comparisons between the two. It describes a world where security is part of an integrated ecosystem, not a distant fortress. It is clear and evident from the above observation that the performance score of the XGBoost model tends to degrade significantly if any features are removed (F1-Score decreases from 0.91 to 0.72). This is definitely the irreducible conclusion of the paper. It shows that context is everything. Limitations of this study must be noted. The data (just 431 examples) is fine-grained and synthetic but densely generated. The reality is different, and much messier, with billions of transactions in a much 'dirtier' data setting. So these results are directionally strong, but it’s a 'best-case scenario' in an ideal world.

However, the productization of this framework would involve significant engineering complexity for ingesting data velocity, real-time feature generation, and latency serving a model. Another point to consider is the Impedance Plot (Figure 3), where the result is obtained using an internal model based on a specific assumption. While these importance measures are highly informative, they are not a causal explanation, and additional, deeper studies would be required.

VII. CONCLUSION

In this paper we aimed to introduce and validate an intelligent framework for detecting fraud in cross platform banking/retail transactions. The heart of the problem we were addressing was that all "normal" detection mechanisms were catch by sophisticated fraud committing on radically different domains. The study's conjecture was, that a machine learning engine fed with cross-platform integrated data will outperform all possible approaches. This hypothesis is experimentally validated as the results of our evaluation clearly show. Results Based on a focused dataset containing 431 instances, ensemble machine learning methods and especially XGBoost are highly effective in doing so with a balanced F1-Score of 0.91. This represents a substantial enhancement over the artificial rule-based system which was unable to track more than 50% of the deception cases. The primary novel advance of this study is the conclusive quantitative evidence for some of the value generated from integrated data. As we see in Table 2 and Figure 3, the single most important indicator in the entire model was “Bank-Retail Location Mismatch”— an engineered variable that is possible to create only by combining two platforms. Models not given this cross-platform context faired abysmally. This report demonstrates that the most useful security insights in today's connected economy do not lie within data silos but rather in the bridges between them. This paper offers a technical plan for such bridge, it is a more intelligent, flexible and efficiency way to defence in the digital world. While the research offers a strong proof of concept, the way to a commercially viable system certainly offers plenty of food for thought. The most important goal should be to advance from the synthetic data. Further research should consider testing this framework on a massive, anonymized, and real-world dataset generated from collaboration between financial institution and large retailer. This would not only help validate the findings in a larger scale but also reveal issues that may be due to data noise and volume. Second, one can develop the models themselves. This work focused on ensemble methods, but the future of fraud discovery may be Graph Neural Networks (GNNs). A GNN has the potential to model the entire ecosystem of users, devices, and merchants as a massive connected graph – spotting fraud "rings" and collusive behaviours that would be unrecognizable to transaction-based models. Thirdly, explainability is an

essential concept. The XGBoost model is very accurate, but because it's a "black box", many are hesitant to use it. Future work could involve incorporating an approach (waived/refer) contextual explanation based explainability methods and so on, which can help in providing simple human-understandable rationale for each decision. Both are necessary for regulatory compliance and for building confidence that the system can be trusted by human users to use its output responsibly. Lastly, the feature engineering is scalable to take unstructured data like text from customer service logs or review sections and processing it (NLP) in order to add an additional dimension of behavioural context.

REFERENCES

[1] Y. Bao, G. Hilary, and B. Ke, "Artificial intelligence and fraud detection," in *Innovative Technology at the Interface of Finance and Operations*, Springer, Cham, 2015, pp. 181–205.

[2] A. Alex-Omiogbemi, A. K. Sule, B. M. Omowole, and S. J. Owoade, "Advances in cybersecurity strategies for financial institutions: combating E-channel fraud in the digital era," 2016.

[3] A. Aggarwal et al., "Federated learning on internet of things: extensive and systematic review," *Computer Modeling in Continua*, 2017.

[4] A. Ali et al., "Financial fraud detection based on machine learning: a systematic literature review," *Applied Sciences*, vol. 12, p. 9637, 2018.

[5] K. Shaffer, C. Gaumer, and K. P. Bradley, "Artificial intelligence products reshape accounting: time to re-train," *Development and Learning in Organizations*, vol. 34, no. 6, pp. 41–43, 2019.

[6] F. Hilal, S. A. Gadsden, and J. Yawney, "Financial fraud: a review of anomaly detection techniques and recent advances," *Expert Systems with Applications*, vol. 193, 2020, art. 116429.

[7] K. S. Lim, L. H. Lee, and Y.-W. Sim, "A review of machine learning algorithms for fraud detection in credit card transactions," *International Journal of Computer Science and Network Security*, vol. 21, no. 9, pp. 31–40, 2021.

[8] F. K. Alarfaj et al., "Credit card fraud detection using state-of-the-art machine learning and deep learning algorithms," *IEEE Access*, vol. 10, pp. 39700–39715, 2022.

[9] Afriyie J. K. et al., "A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions," *Decision Analytics Journal*, vol. 6, 2023, art. 100163.

[10] N. Rane, S. Choudhary, and J. Rane, "Machine learning and deep learning: a comprehensive review on methods, techniques, applications, challenges, and future directions," 2024.

[11] G. Manoharan et al., "Machine learning-based real-time fraud detection in financial transactions," in *Proc. 2024 Int. Conf. Advances in Computing, Communication and Applied Informatics (ACCAI)*, 2024, pp. 1–6.

[12] G. Liu, "Leveraging machine learning for telecom banking card fraud detection: logistic regression, random forest, and XGBoost models," *Computers and Artificial Intelligence*, vol. 1, no. 1, pp. 13–27, 2024.

[13] J. Xu, T. Yang, S. Zhuang, H. Li, and W. Lu, "AI-based financial transaction monitoring and fraud prevention with behaviour prediction," *Applied Computing and Engineering*, vol. 77, pp. 218–224, 2024.