

Real-Time Anomaly Detection Using ML, DL, and XAI for Financial Fraud: A Systematic Review

Padilla-Lopez Jimmy Andres, Lazo-Ordoñez Pedro Antonio, Juan Jesús Soria-Quijate
Escuela de Ingeniería de Sistemas, Universidad Tecnológica del Perú, Lima, Perú

Abstract—Digital payments, fintech, and online banking are now part of our daily lives. Making a transfer, paying for a service, or shopping online is faster, but this progress has also given rise to new forms of fraud. Rule-based systems remain useful when the pattern is already known; the problem arises when attackers change their tactics. For this reason, this study reviews the use of machine learning, deep learning, and Explainable Artificial Intelligence to detect anomalies and prevent fraud in digital payment systems. The research was conducted as a systematic literature review. PICOC helped define the scope and search terms, while PRISMA guided the selection of articles. Scopus, Web of Science, and ScienceDirect were consulted. After applying the inclusion, exclusion, and methodological evaluation criteria, 33 studies were retained. The articles addressed cases of card fraud, digital banking, blockchain, money laundering, and other suspicious transactions. The most frequently mentioned models were Random Forest, XGBoost, SVM, Isolation Forest, LSTM, CNN, autoencoders, and graph networks. Hybrid approaches and tools such as SHAP and LIME also appeared, used to show which variables influenced an alert. Several studies reported good results in terms of accuracy, recall, and F1 score. However, challenges remain regarding synthetic data, class imbalance, limited information on latency, and insufficient validation with financial specialists. Their performance still needs to be tested with real-world data and under conditions similar to those of a financial institution.

Keywords—Financial fraud; machine learning; deep learning; XAI; digital payments

I. INTRODUCTION

Electronic payments, fintech, and online banking are part of everyday financial services. Their growth has sped up many transactions, but it has also opened up new opportunities for fraud. This affects businesses through financial losses and also affects users, who may lose trust in these platforms.

Aparcana-Tasayco et al.[1] note that the rise of connected digital services necessitates improvements in anomaly detection methods.

The reason is that forms of fraud do not remain static: they evolve, test new strategies, and seek to circumvent existing controls. Consequently, methods based on static rules prove insufficient for detecting complex fraudulent behavior. In light of this situation, anomaly detection has emerged as a key approach, enabling the identification of unusual patterns in large volumes of data. In this context, Zhang et al.[2] note that machine learning models based on probabilistic approaches improve the early detection of suspicious activities in financial systems.

Likewise, the use of machine learning techniques has enabled the development of predictive models capable of classifying financial transactions by learning from historical patterns. These models include supervised and unsupervised algorithms, which have proven to be more efficient than traditional methods in identifying fraud. In this regard, Kumar et al. [3] note that the use of deep learning techniques has significantly increased detection capabilities by enabling the modeling of complex and nonlinear relationships in financial data.

The digitization of payment systems has led to more sophisticated forms of financial fraud, which have outpaced traditional rule-based methods. In this context, Wang et al. (2024) propose the use of dual-channel memory networks to extract features from credit review texts, demonstrating that it is possible to improve the accuracy of fraud risk detection by leveraging unstructured data [4].

Digital payment systems require accurate models capable of processing large volumes of transactions in real time. Raja et al. (2025) propose a hybrid model that combines GraphSAGE with LSTM networks to detect suspicious patterns, achieving 95.4% accuracy in money laundering detection. [5]

Finally, hybrid models that combine machine learning and deep learning techniques have proven to be a promising alternative for improving accuracy and reducing false positives in fraud detection. [6]

A. Problem Definition

1) *Current situation*: Digital payments and fintech platforms have become a common part of financial activities due to the speed and convenience they offer. However, the increase in electronic transactions has also created more opportunities for fraud. In the United States alone, more than 2.7 billion electronic transactions are recorded each year, while organizations lose nearly 5% of their annual revenue due to this problem [7].

Detecting these crimes is not easy. Attackers combine cyber techniques with social engineering and constantly change their methods. At the same time, a system that is too sensitive may block legitimate transactions. Added to this is the need to protect user privacy and explain how automated decisions are made [8].

Traditional systems operate using manually defined rules and thresholds. The problem is that these rules must be updated every time a new form of fraud emerges, so they often react too

late. Černevičienė and Kabašinskas [9] explain that this reliance on fixed rules reduces adaptability. Furthermore, these systems can generate between 85% and 99% false positives, which affects customers and increases the workload for the teams responsible for reviewing alerts [10].

Another limitation lies in response time. Some systems still process transactions in batches, meaning that fraud may be detected after the transaction has already been completed [11]. This delay, coupled with the rise in attacks, continues to erode user confidence in digital financial services [12].

Deep learning models offer better capabilities for recognizing complex patterns, but they also present challenges. Their real-time application requires significant processing power, especially when they must analyze thousands of transactions in just a few seconds. Furthermore, fraud accounts for a very small portion of total transactions—in some cases, less than 0.1%. For this reason, a model that performs well in laboratory tests does not always maintain the same performance when applied on a large scale [13].

B. Rationale

Fraud in digital payments has a direct impact on businesses and the people who use these services. It is estimated that organizations lose nearly 5% of their annual revenue due to this issue [14]. Furthermore, more than 2,700 million electronic transactions are conducted annually in the United States. This growth facilitates financial transactions, but it also increases the number of situations that can be exploited to commit fraud. Ketenci et al. (2021) combined time-frequency analysis with transactional information and CRM data. With this approach, they reduced false positives to 11.85% and were able to better distinguish suspicious entities from those showing no signs of fraud [15].

The rise of digital financial services has not only expanded access but has also changed the rules of the game when it comes to attacks. Systems that operate with fixed rules and predefined limits can still respond effectively to known threats, but they fall short when new or unexpected behaviors emerge. This is where many traditional approaches fall short: they fail to detect in a timely manner those unusual relationships between transactions or the subtle changes in attackers' behavior [16].

Faced with this gap, machine learning began to be seen as a more promising alternative, precisely because it learns from past events. By processing vast amounts of transactions, these models can flag movements that deviate from the norm. And when the information is fluid, or the relationships do not follow a clear logic, probabilistic methods are often the best fit [17]. For his part, Sørensen (2025) analyzes how artificial intelligence and machine learning help uncover signs of fraud that would go unnoticed in a manual review. But he also mentions the problems that remain unresolved, such as false positives and the opacity of some models—two fundamental issues if we are to move toward real-time detection systems that are reliable and, above all, can be clearly explained in the financial sector [18].

The article by Kulkarni et al. (2025) aligns well with what we're seeing here, because it delves deeply into the use of generative AI—VAEs and GANs, to be specific—applied to fraud detection. What's most striking is that these models not

only detect very subtle anomalies but can also generate synthetic data without leaking private information. The idea is to build systems that instantly detect fraud in digital payments, without compromising either reliability or the ability to provide clear explanations of how they work [19]. And when we turn to deep learning, the landscape changes because it allows us to analyze transaction sequences at a level of detail that was previously impossible. Networks such as CNNs, LSTMs, autoencoders, and transformers have demonstrated their ability to uncover connections that go completely unnoticed by simpler methods. In fact, in the studies we've reviewed, these architectures have yielded excellent results, both in terms of accuracy and the ability to anticipate future events [20].

But if there's one thing that's proving particularly interesting, it's hybrid models—those that combine the best of traditional machine learning with deep learning. For example, integrating the One-Class SVM with a probabilistic approach has yielded remarkable results: it achieves extremely high accuracy and recall in fraud detection, and best of all, it reduces those false positives. Furthermore, the models become more versatile and adapt better to different scenarios [21].

Finally, speed is key. There's no point in having a system that detects fraud if, by the time it does, it's already too late. That's why the ability to identify suspicious transactions in real time has become essential in digital payments. What researchers have found is that combining models such as transformers, LOF, and Random Forest yields results that excel in precision, recall, and F1-score, demonstrating their enormous potential when applied to complex and fast-paced financial environments [22]. In the words of experts in the field, integrating these approaches is not just a technical improvement: it is a real contribution to the security and efficiency of the digital financial ecosystem as a whole [23].

II. METHODOLOGY

This research is being conducted in accordance with the Systematic Literature Review (SLR) format.

A. PICOC Formulation

The PICOC method was used to determine the scope of the research (Table I).

TABLE I. PICOC COMPONENTS

Component	Description
P (Problem / Population)	Financial transactions suspected of fraud in digital payment systems.
I (Intervention)	Deep learning algorithms with explainability layers.
C (Comparison)	"Black box" deep learning models (without XAI)
O (Results)	Improved interpretability of alerts without significantly compromising accuracy or speed
C (Context)	Fraud prevention departments that require explanations to make operational or regulatory decisions

B. Formulating Questions

After identifying the components of the PICOC method, the general research question (RQ) for this systematic review was formulated:

“What are the deep learning algorithms with Explainable Artificial Intelligence (XAI) applied to the detection of financial fraud in digital payment systems, compared to traditional black-box models, in terms of interpretability, accuracy, and speed?”

Based on this general question, specific questions were formulated (Table II).

TABLE II. PICOC QUESTIONS

RQ1	What machine learning algorithms have been used to detect financial fraud and anomalies in digital payment systems?
RQ2	What deep learning algorithms have been used to detect financial fraud and anomalies in digital payment systems?
RQ3	What machine learning and deep learning models have been used to detect financial fraud in digital payment systems, and to what extent do they outperform or are they outperformed by traditional methods (rule-based and statistical) in terms of effectiveness?
RQ4	What improvements in the interpretability of alerts—while maintaining high levels of precision, recall, F1-score, and a low false positive rate—have been used for fraud detection in digital payment systems?
RQ5	What evidence does the literature provide regarding the effectiveness of real-time models using XAI techniques to improve interpretability without significantly compromising fraud detection performance?

C. Identification of Keywords

The purpose of identifying keywords corresponding to each component of the PICO-C method is to create effective search queries for databases (Table III).

TABLE III. KEYWORDS WITH BOOLEAN OPERATORS

P	"financial transactions" OR "digital payments" OR "payment systems" OR fintech OR banking OR "financial data"
I	"machine learning" OR "deep learning" OR "artificial intelligence" OR "neural networks" OR autoencoder
C	"rule-based" OR "statistical methods" OR "traditional methods" OR "baseline models"
O	"fraud detection" OR "financial fraud" OR "anomaly detection" OR "outlier detection"
C	"real-time" OR "online detection" OR "streaming data" OR "transaction monitoring" OR "explainable AI" OR XAI OR "interpretable machine learning"

D. PICOC Search Formula

Once we had identified the components of the PICOC question and the keywords, we proceeded to develop the search formula that would allow us to find the most relevant literature for this systematic review. To do so, we combined the terms from each component using the Boolean operators “OR” to combine the keywords and “AND” to link the components (Table IV).

TABLE IV. SEARCH EQUATION

Search equation	
Scopus	(TITLE-ABS-KEY ("financial transactions" OR "digital payments" OR "payment systems" OR fintech OR banking OR "financial data") AND TITLE-ABS-KEY ("machine learning" OR "deep learning" OR "artificial intelligence" OR "neural networks" OR autoencoder) AND TITLE-ABS-KEY ("rule-based" OR "statistical methods" OR "traditional methods" OR "baseline models") AND TITLE-ABS-KEY ("fraud detection" OR "financial fraud" OR "anomaly detection" OR "outlier detection") AND TITLE-ABS-KEY ("real-time" OR "online

	detection" OR "streaming data" OR "transaction monitoring" OR "explainable AI" OR XAI OR "interpretable machine learning")) AND (LIMIT-TO (OA, "all")) AND (LIMIT-TO (LANGUAGE, "English")) AND PUBYEAR > 2020 AND PUBYEAR < 2025
Web of Science	(((((TS=("financial transactions" OR "digital payments" OR "payment systems" OR fintech OR banking OR "financial data")) AND TS=("machine learning" OR "deep learning" OR "artificial intelligence" OR "neural networks" OR autoencoder)) AND TS=("rule-based" OR "statistical methods" OR "traditional methods" OR "baseline models")) AND TS=("fraud detection" OR "financial fraud" OR "anomaly detection" OR "outlier detection")) AND TS=("real-time" OR "online detection" OR "streaming data" OR "transaction monitoring" OR "explainable AI" OR XAI OR "interpretable machine learning")) AND DT=(Article) AND LA=(English)) AND PY=(2021-2026)
Science Direct	("digital payments" OR fintech) AND ("machine learning" OR "deep learning") AND ("fraud detection" OR "anomaly detection") AND ("real-time" OR XAI OR "interpretable machine learning")

E. Inclusion and Exclusion Criteria

The selection of the studies was not arbitrary: specific inclusion and exclusion criteria were applied to ensure their relevance and methodological soundness (Table V).

TABLE V. INCLUSION AND EXCLUSION CRITERIA (PICOC)

Component	Inclusion criteria	Exclusion criteria
Population / Problem	We included research on fraudulent financial transactions in digital payment systems—banks, fintech companies, payment processors, card issuers, digital wallets, and payment gateways—as well as real or synthetic transaction datasets that replicated the financial environment and its characteristic class imbalance.	Studies on non-financial fraud (healthcare, academia, telecommunications, network intrusions) were excluded, unless they justified their transfer to the financial domain. Research on financial transactions not linked to digital payment systems was also excluded.
Scope	Articles that propose, implement, or evaluate deep learning algorithms with explicit Explainable Artificial Intelligence (XAI) mechanisms, such as attention layers, SHAP, LIME, Grad-CAM, gradient-based models, or interpretable architectures.	Studies that used only classical machine learning without deep learning, deep learning models without explainability (black-box models), and studies where XAI was mentioned only as future work without actual development were excluded.
Comparison	Studies that compare explainable models with baseline or reference models—such as neural networks without XAI, Random Forest, XGBoost, SVM, rule-based methods, or traditional statistical approaches—using equivalent experimental conditions.	Research that evaluates only the proposed model without a baseline, or comparisons conducted using different datasets, different preprocessing methods, or non-reproducible metrics.
Results	Preference was given to studies that provided specific performance and	Qualitative studies without accuracy or speed metrics were excluded, as

	interpretability metrics—such as precision, recall, F1-score, ROC-AUC, PR-AUC, latency, throughput, accuracy, or stability—as well as those that included validation by users in the financial sector.	were those using indicators unsuitable for imbalanced data and those with severe performance degradation not justified by their methodology.
Context	The review prioritized studies on fraud detection systems that need to explain operational and regulatory decisions, and included approaches involving online detection, continuous monitoring, and auditing.	Academic studies without real-world validation were excluded, as were batch-processing approaches that ignore latency or fail to explain decisions to regulators.
General Criteria	Articles from databases such as Scopus, Web of Science, and ScienceDirect, preferably in English, published between 2021 and 2026, with full text available and aligned with the review's PICO question.	We excluded theses, unreviewed technical reports, editorials, unverified documents, duplicates, articles without a DOI or full text, and publications outside the established time frame, language, or scope.

F. PRISMA Statement

PRISMA helped organize the review from the outset. Using these guidelines, we documented how articles were searched for, which ones were excluded, and which ones were included in the final analysis. It also allowed us to present the selection process in a more structured way, so that readers can understand the source of the included studies. According to Page et al.[24], the 2020 version of PRISMA updated the 2009 guidelines to incorporate technological and methodological advances in the search for and analysis of evidence, resulting in a more comprehensive and organized report. Its ability to improve the quality and credibility of publications has led many journals and scientific organizations to adopt it. [24]

According to this study, the PRISMA statement was essential for systematically organizing the process of selecting and analyzing research related to wearable sensors, the Internet of Things (IoT), and artificial intelligence used in mental health monitoring. This guideline was used to establish clear criteria during the selection, exclusion, and final review stages of the articles, ensuring the study's methodological consistency. According to Hutton et al. [25], PRISMA promotes the understanding and comparison of results by encouraging a clear presentation of the techniques used. Thus, it supports transparency, strengthens confidence in the results, and enables future research to replicate and expand upon the findings with greater reliability.

G. PRISMA Process

The PRISMA process was used to review this research. A total of 373 articles were identified: 51 from Scopus, 15 from Web of Science, and 287 from ScienceDirect. After removing duplicates and applying the inclusion and exclusion criteria, the remaining studies were evaluated by reviewing their titles, abstracts, and full texts (Fig. 1).

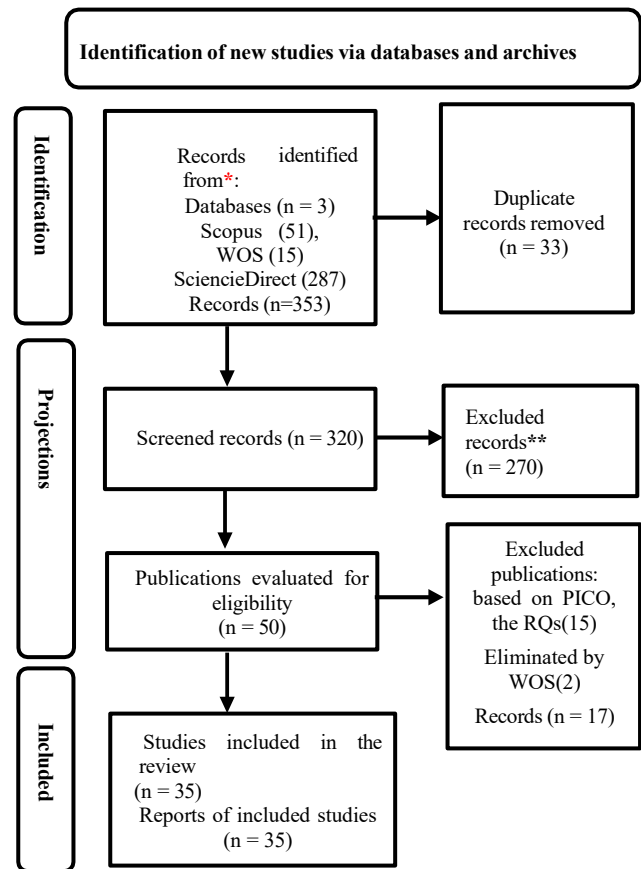


Fig. 1. PRISMA process diagram.

H. Assessment of Methodological Quality

To assess the methodological quality of the 35 studies included in this review, we followed the adapted guidelines of the Joanna Briggs Institute (JBI).[25] Seven criteria were considered: 1) adequate characterization of the dataset (source, size, imbalance, delayed feedback); 2) appropriateness of the data and preprocessing; 3) clear explanation of the algorithm's methodology and the XAI mechanism; 4) rigorous validation (time-series analysis, confidence intervals, statistical tests); 5) use of appropriate metrics for imbalanced data (F1, PR-AUC, latency); 6) explicit discussion of limitations; and 7) quantitative assessment of interpretability. Each of the defined criteria was evaluated on a binary scale—that is, using 0 and 1—which allowed the studies to be classified according to their quality.

Weighted synthesis: Findings from high-quality studies will be given greater weight in the conclusions, while those from moderate-quality studies will be used as a supplement.

I. Bibliometrix

Bibliometrix is an open-source R package, developed by Aria and Cuccurullo (2017), that enables comprehensive bibliometric analyses, including the import of data from databases such as Scopus and Web of Science, as well as the construction of co-citation, collaboration, and keyword co-occurrence networks [26].

One of its most highly valued features is Biblioshiny, its graphical interface based on the Shiny web application, which

allows researchers to use all of its functions without needing prior knowledge of R programming. This combination of analytical power, flexibility, and accessibility has made Bibliometrix one of the most comprehensive and popular tools for the quantitative analysis of scientific output.[27]

III. RESULTS

Following the PRISMA process, 33 articles remained for review. These articles presented different approaches to addressing fraud in digital payments. Several articles utilized machine learning and deep learning. Other studies tested graph networks or incorporated XAI techniques. The applications were varied, although most were related to credit cards, digital banking, blockchain, and money laundering.

The comparison was based on the type of fraud studied, the algorithms used, and the results reported in each article. We also examined how the studies handled the imbalance between normal and fraudulent transactions. In the studies that included explainability, we considered the use of tools such as SHAP and LIME.

The methodological assessment revealed differences among the articles. Of the 33 selected articles, 24 were considered high quality, 5 were of moderate quality, and 2 were of low quality. In the highest-rated studies, it was common to find real-world or publicly available data, comparisons between various models, and results expressed in terms of precision, recall, and F1-score.

When reviewing the limitations, we found that several studies still rely on synthetic data or tests conducted under controlled conditions. In other cases, the time it takes for the model to respond is not reported in sufficient detail. We also found proposals with good results but with little explanation as to why a transaction was considered suspicious.

A. Comparative Overview of Reviewed Studies

1) *ML/DL Models and metrics*: The results of the reviewed articles are shown in Table VI, which lists the reference, the types of financial fraud in digital payment systems, and the applied ML and DL algorithm models.

TABLE VI. FRAUD TYPES, ML/DL MODELS, AND METRICS

Reference	Types of Financial Fraud in Digital Payment Systems	ML and DL Models	Metrics
Ampumuza, D. et al.[28]	Mobile payment fraud, phishing, card fraud, cryptocurrency scams, identity theft, account takeover	Random Forest, XGBoost, Neural Networks, clustering (K-means), autoencoders, hybrid models (reglas + ML)	No reporta métricas cuantitativas (accuracy, F1, recall, AUC, FPR)
Vazallo-Bautista, J. et al.[29]	CNP, skimming, unauthorized digital payments, account takeover, fake accounts, money laundering using money mule accounts, card fraud (cloning and CNP), phishing, identity theft, suspicious transactions, fraudulent digital payments	Random Forest, XGBoost, Isolation Forest, Autoencoders, LSTM/GRU, CNN, GNN, LightGBM, CatBoost, Hoeffding Tree, Adaptive Random Forest, RNN, GRU, GAN, One-Class SVM, KNN, Logistic Regression, DT, AdaBoost, Stacking Ensemble, CNN-LSTM with Attention	Accuracy (94-99%), Precision (>95%), Recall (90-98%), F1-Score (0.90-0.999), AUC-ROC (>0.98), MCC, False Positive Rate (reducción >40%), Latency (<100 ms), throughput (tx/s)
Alatawi, M. N. et al.[30]	Credit card fraud (CNP, skimming, unauthorized digital payments, online fraud, IoT-based transactions)	Logistic Regression, Random Forest, Gradient Boosting Machine (GBM), Multilayer Perceptron (Neural Networks)	Accuracy (RF: 99.999%, GBM: 99.9865%, MLP: 99.876%, LR: 95.87%), Precision (RF: 1.0, GBM: 0.9995, MLP: 0.9970, LR: 0.8911), Recall (RF: 0.9999, GBM: 0.9990, MLP: 0.9888, LR: 0.6000), F1-Score (RF: 0.9999, GBM: 0.9992, MLP: 0.9929, LR: 0.7171)
Abd-Ellatif, L. et al.[31]	Credit card fraud, identity theft, phishing, unauthorized fund transfers	ATAD-Net (CNN + LSTM), CNN, RNN	Accuracy: ATAD-Net 98.65%, Precision: 97.12%, Recall: 96.74%, F1-Score: 96.93%, Average latency: 8.2ms por transacción, Throughput: reportado en Figura 8 (TPS), True Positives, False Positives, True Negatives, False Negatives (confusion matrix Figura 2)
Al-Mustaqbal University et al. [32]	Credit card fraud, unauthorized online transactions, electronic financial crimes	Decision Tree, Logistic Regression, Naive Bayes, Random Forest, ANN, SVM, KNN, CNN, LSTM, GAN, Ensemble methods (Bagging, Boosting)	Accuracy (SMOTE: 98.1%, undersampling: 94.9%, oversampling: 95.3%, GAN: 99.9%), Precision (SMOTE: 99.7%, undersampling: 98.8%, oversampling: 99.7%, GAN: 99.9%), Recall (SMOTE: 98.1%, undersampling: 94.9%, oversampling: 95.2%, GAN: 99.9%), F1-Score (SMOTE: 98.1%, undersampling: 94.7%, oversampling: 95.1%, GAN: 99.9%), ROC-AUC (SMOTE: 96%, undersampling: 95%, oversampling: 95.5%, GAN: 100%)
Leong, W. Y. et al.[33]	Identity theft, payment fraud, account takeover, credit card fraud, phishing	Logistic Regression, Decision Trees, Random Forest, Neural Networks, CNN, RNN, Isolation Forest, Autoencoders	Accuracy (LR: ~91%, RF: ~95%, SVM: ~90%, LSTM: ~97%, Autoencoder: ~82%), Precision, Recall, F1-Score,

Reference	Types of Financial Fraud in Digital Payment Systems	ML and DL Models	Metrics
			AUC (LSTM: 0.98, RF: 0.96), curvas precision-recall, matrices de confusión
El Youbi, R. et al.[34]	Electronic payment fraud, credit card fraud, unauthorized transactions, phishing	DNN, LSTM RNN, Logistic Regression, KNN, Random Forest, CNN, Deep Reinforcement Learning + ANN	Accuracy: 99.72%, AUC: 97.43%, Precision (fraud): 96.0%, Recall (fraud): 93.8%, F1-Score (fraud): 94.9%, Precision (no fraud): 99.8%, Recall (no fraud): 99.9%, F1 (no fraud): 99.8%, ROC-AUC: 0.999, PR-AUC: 0.913, Average Precision: 0.913, Confusion Matrix: TN=14,855, FP=17, FN=1, TP=127
Al Lawati, H. M. R. et al. [35]	Card fraud (skimming, CNP, counterfeit cards, loss/theft), payment fraud, phishing, collusion with merchants	CNN, ADASYN + Mutual Information + SelectKBest, EDDM, ADWIN, Random Forest, XGBoost, LightGBM, SVM, KNN, Decision Tree, J48, K-means, GRU	Accuracy (Credit Card: 100%, Real: 100%, Spam: 99.34%), Precision (Credit Card: 100%, Real: 100%, Spam: 100%), Recall (Credit Card: 100%, Real: 100%, Spam: 98.69%), F1-Score (Credit Card: 100%, Real: 100%, Spam: 99.34%), ROC-AUC (Credit Card: 100%, Real: 100%, Spam: 99.34%), Drift Detection Rate (Credit Card: 100%, Real: 100%, Spam: 99.34%)
Pavlova, O. et al.[36]	Money laundering, drop accounts, cryptocurrency fraud (Bitcoin)	GNN, GCN, DNN, Random Forest, Logistic Regression, MLP, CNN, KNN, LASSO, Gradient Boosting, Elastic Net	Accuracy: 93%, Recall: 88%, F1-Score: 90%, Precision-Recall AUC: valores altos según la curva (Figura 5), Confusion Matrix con TP, TN, FP, FN (Figura 4). La Precision-Recall curve muestra buen balance pero Precision más baja que Recall
Iqbal, S. et al.[37]	Credit card fraud, online transaction fraud	Random Forest, XGBoost, SHAP, LIME	RF: Accuracy 0.9996, Precision 0.8909, Recall 0.9423, F1 0.9159, ROC-AUC 0.9979, CV-F1 0.9998. XGB: Accuracy 0.9995, Precision 0.8776, Recall 0.9556, F1 0.9149, ROC-AUC 0.9997, CV-F1 0.9998
Renuga Devi, R. et al.[38]	Credit card fraud, real-time fraudulent transactions, clustered fraud patterns (coordinated attacks or organized fraud), detection of suspicious communities (cardholders, merchants, and compromised devices)	GNN (Graph Attention Networks), RL (Deep Q-Network), GAN, RNN (GRU, LSTM), SVM, Random Forest, XGBoost, Autoencoders, Restricted Boltzmann Machines, CART	AUROC 0.872, AP 0.683, Recall 0.85, Precision 0.83, F1 0.839, Balanced Accuracy 0.916, MCC 0.54, FPR 0.08, Latency 42ms/batch, Throughput 12,000 tx/s
Černevičienė, J. & Kabašinskas, A.[39]	General fraud detection (including credit card fraud, detection of suspicious transactions, electronic payment fraud, and P2P loan fraud)	ANN, XGBoost, Random Forest, SHAP, LIME, LightGBM, CNN, LSTM, Autoencoders, Gradient Boosting, Decision Trees, Logistic Regression, KNN, SVM, frameworks multi-enfoque	SLR cualitativa. 138 artículos, 397 autores, 80% publicados 2020-2022, China (33), Italia (17), Alemania (11), coautoría internacional 10%
Prashanth, B. S. et al.[40]	Banking transaction fraud, digital banking fraud, detection of suspicious transactions in electronic payment systems	TabNet, DNN, GRU, LSTM, CNN1D, SMOTE	Accuracy 97.39%, ROC-AUC 0.9739, Fold AUCs: 0.9736,0.9733,0.9747. LSTM AUC 0.5118, GRU AUC 0.517, CNN1D AUC 0.50, DNN AUC 0.5
Kungu, C. O. et al.[41]	Money laundering, cross-border payment fraud, suspicious transactions	VAE, GNN, One-Class SVM, Isolation Forest, LOF, SHAP	Precision@5%: OCSVM 99.63%, IF 84.92%, LOF 89.76%. AP: OCSVM 0.8478, IF 0.6742, LOF 0.8019. ROC-AUC: OCSVM 0.8388, IF 0.7826, LOF 0.8459. Throughput: 1000 TPS
Karnavou, E. et al.[42]	Credit card fraud (banking transaction fraud)	Isolation Forest, One-Class SVM, Autoencoder, SHAP	Isolation Forest (Merchants DE FS): TP 59%, FP 5%, Acc 93%, F1 15%, AUC 80%, Precision 12%. Persons: TP 48%, F1 20%, AUC 70%, Precision 17%. Autoencoder Persons: TP 85%, F1 20%, AUC 88%, Precision 11%
Gupta, R. K. et al.[43]	Credit Card Fraud	SVM, KNN, ELM, Stacking Ensemble (Gradient Boosting), Autoencoder.	Stacking Ensemble: Acc 99.95%, Prec 99.93%, Rec 99.97%, F1 99.95%, AUC 1.00, FPR 0.02%. SVM: Acc 97.33%, Prec 98.69%, Rec 96.03%, F1 97.34%, AUC 0.98. KNN: Acc 99.87%, Rec 100%, F1 99.87%, AUC 0.99

Reference	Types of Financial Fraud in Digital Payment Systems	ML and DL Models	Metrics
Hasan, M. et al.[44]	Fraud in blockchain transactions (Bitcoin), fraudulent cryptocurrency transactions	1D CNN, XGBoost, Random Forest, Decision Tree, Gradient Boosting, AdaBoost, SHAP	1D CNN+SMOTE (early stop): Acc 0.98, TPR 0.36, FPR 0.02. 1D CNN+ADASYN: Acc 0.99, TPR 0.27, FPR 0.005. 1D CNN+SMOTEENN: Acc 0.99, TPR 0.27, FPR 0.0008
Chen, Y. et al.[45]	Credit card fraud, bank fraud, insurance fraud, cryptocurrency/blockchain fraud, money laundering	CNN, LSTM, Transformer, GNN, GAN, VAE, Autoencoder-LSTM, CatBoost-DNN, Random Forest, XGBoost, LightGBM.	108 estudios. LSTM 25.7%, MLP 20%, CNN 15.7%, RNN 14.3%, Transformers 11.4%. Credit card sector más estudiado. 92/108 reportan desbalanceo. Métricas: accuracy, precision, recall, F1, AUC-ROC, AUC-PR, FPR, FNR, cost metrics
Murugaraj, G. et al.[46]	Credit card fraud, fraud detection, anti-money laundering (AML)	QSVM, QNN, QAOA, Quantum Clustering, Quantum GNN, Quantum Amplitude Estimation, Quantum Boltzmann Machines, Hybrid Quantum-Classical models	QSVC (ZFeatureMap): F1=0.98 (BankSim). QNN: 97.63% accuracy. QGNN: 94.5% accuracy. QAOA: speedups 10-1000x. QAE: quadratic speedup. QADQN: 75.37% returns
Mia, M. S. et al.[47]	Credit card fraud	Isolation Forest, CNN, Autoencoder, XGBoost, Decision Tree, KNN, Logistic Regression, LSTM, Random Forest, SVM, VQC (Quantum), SMOTE, AutoEncoder + Bagging (RF) + Boosting (XGBoost)	Proposed Model-1: Acc 100%, Prec 1.00, Rec 1.00, F1 1.00, AUC 0.9998. XGBoost: Acc 100%, Prec 0.81, Rec 0.93, F1 0.86, AUC 0.9841. RF: Acc 100%, Prec 0.94, Rec 0.91, F1 0.92, AUC 0.9684. VQC: Acc 86%, Prec 1.00, Rec 0.72, F1 0.84
Uddin, N. et al.[48]	Financial fraud in general (credit card fraud, money laundering, identity theft, digital payment fraud)	Decision Trees, Random Forest, XGBoost, Neural Networks, Isolation Forest, Hybrid Models (reglas + ML), Explainable AI (XAI)	Accuracy puede ser engañosa en datasets desbalanceados. Se recomienda Precision (>70%), Recall, F1-Score, AUC-ROC. Pérdidas EE.UU.: \$12.5B (2024)
Usman, A. et al.[49]	Money laundering (AML), fraudulent transactions in payment systems	Graph Convolutional Networks, Semi-supervised learning, PageRank, Word2Vec, Doc2Vec, TransE, ComplEX	Accuracy 79.36% promedio (79.96% max). RF F1 0.619, NN 0.479, DT 0.423. Dataset: 100K nodos, 5.4M aristas
Marazqah Btoush, E. et al.[50]	Credit card fraud, cyber fraud in banking transactions, online payment fraud	LR, RF, SVM, DT, NB, KNN, ANN, CNN, RNN, LSTM, GRU, XGBoost, LightGBM, CatBoost, AdaBoost, GAN, Autoencoder, Isolation Forest, LOF, HMM, K-means, SMOTE, ADASYN, RUS, Tomek	Revisión sistemática; no aplica una métrica única por modelo. Sintetiza resultados y tendencias de ML/DL en fraude con tarjetas.
Ahmad, H. et al.[51]	Credit card fraud	ANN, XGBoost, RF, LR, SVC, GB, Tomek Links (undersampling)	Accuracy ANN 99.36% (D1), 99.36% (D2); Recall ANN 99.41% (D1), 98.92% (D2); Precision ANN 99.73% (D1), 98.73% (D2); F1 ANN 99.36% (D1), 99.37% (D2); G-Mean ANN 99.36% (D1), 99.21% (D2). RF Accuracy 98.57% (D1), GB 98.57% (D1), SVC 97.61% (D1)
Miao, Z. et al.[52]	Financial fraud (credit card fraud, e-commerce fraud, electronic payment fraud, cryptocurrency fraud)	Transformer, LOF, Random Forest, XGBoost, LightGBM, LSTM, Isolation Forest	Accuracy 95.47%, Precision 94.32%, Recall 96.58%, F1 95.44%, AUC 97.12%, MCC 92.76%, Specificity 95.13%. Transformer: CCFD Acc 89.63%, EPF Acc 96.23%, FTTS Acc 91.63%, VCT Acc 92.89%. LOF: CCFD Acc 96.2%, EPF Acc 95.48%, FTTS Acc 96.22%, VCT Acc 92.39%
Khan A. et al.[53]	Credit card fraud	LSTM-RNN, Naive Bayes, SVM, ANN	LSTM-RNN: Accuracy 100%, Precision 1.00, Recall 1.00, F1-Score 1.00, Sensitivity 1.00. ANN: Accuracy 91.85%, Precision 92.6%, F1 71.6%. SVM: Accuracy 86.56%, Precision 68.8%. Naive Bayes: Accuracy 69.54%, Precision 71.6%
Venkata Lakshmi Narayana Gorle. et al.[54]	Debit and credit card fraud (unauthorized purchases), online banking fraud, document forgery, false financing, scams, fake accounts, financial statement fraud (alteration of accounting records), identity theft, phishing, mobile payment	Deep Learning Models: AM-1DCNN (Adaptive Multiscale 1D Convolutional Neural Network), A-CNet (Adaptive Capsule Network). Machine Learning Models (feature extraction): PCA (Principal Component Analysis), t-SNE (t-distributed Stochastic	IWPA-AM-1DCNN: Acc 94.18%, Prec 94.15%, Rec 94.23%, F1 93.86%, Spec 94.17%, MCC 88.36%, FPR 5.83%. IWPA-A-CNet: Acc 93.76%, Prec 93.62%, F1 93.69%, MCC 87.52%, FPR 6.25%.

Reference	Types of Financial Fraud in Digital Payment Systems	ML and DL Models	Metrics
	fraud, cryptocurrency scams, account takeover.	Neighbor Embedding), SVM (Support Vector Machine). Optimization algorithm: IWPA (Improved Waterwheel Plant Algorithm). Compared ensemble techniques: XGBoost, IMEML, HEML, RILHHO.	Validación externa: Dataset2 94.27%, Dataset3 93.91%, Dataset4 95.87%, Dataset5 94.96%.
Wang, Y., Zheng, Q.[54]	Blockchain/cryptocurrency fraud, money laundering, phishing, scams on dark markets (Silk Road, Hydra), illicit Bitcoin/Ethereum transactions, identity theft, account takeover.	DL: CoSemiGNN (dynamisches Graph-Neuralnetzwerk mit semi-überwachter Ko-Assoziation), EvolveGCN, LRGCN, GConvGRU, GC-LSTM, DCRNN, A3T-GCN, AGCRN, DyGrEncoder. ML: GCN, GAT, Random Forest, XGBoost, logistische Regression, Entscheidungsbaum, SVC, KNN, MLP, AdaBoost, Struc2Vec. Techniques: Halbüberwachtes Lernen (SelfTrain, sim-kernelKmeans), LSTM, GRU, Multi-Head-Attention.	CoSemiGNN-E (mejor): Acc 98.47%, Illegal-F1 80.16%, Legal-F1 99.17%, Micro-F1 89.66%, Illegal-Prec 96.43%, Illegal-Rec 71.11%. CoSemiGNN-V: Acc 98.14%, Illegal-F1 74.46%, Legal-F1 98.99%, Micro-F1 86.73%, Illegal-Prec 77.57%, Illegal-Rec 73.02%. Comparativa: Superior a GCN (Illegal-F1 38.45%), GAT (36.74%), EvolveGCN-O (43.83%), DCRNN (35.78%).
Lebichot, B. et al.[55]	Credit card fraud (unauthorized transactions), e-commerce fraud, digital payment scams.	DL: NN (dense multilayer network), EWC (Elastic Weight Consolidation), IMM (Incremental Moment Matching), GenR (Generative Replay with CGAN), Frz (Frozen Network). ML: Batch, Retrain, Incremental, Experience Replay (ExpR), Negative Correlation Ensemble (NCE). Techniques: Ensemble learning, rehearsal, regularization, pseudo-rehearsal, weight transfer.	FrzE (Frozen Network Ensemble - mejor): Pr@100 25.36% ± 6.49%, AUPRC 11.28% ± 2.90%. Comparativa: - NCE (Ens): Pr@100 26.24%, AUPRC 11.66% - Incr (Ens): Pr@100 24.28%, AUPRC 11.20% - EWC: Pr@100 24.19%, AUPRC 10.59% - Batch: Pr@100 20.54%, AUPRC 6.97% Tiempos: FrzE 89 min, NCE 126 min, EWC 657 min.
Chang, V. et al.[56]	Credit/debit card fraud (unauthorized transactions), digital payment fraud, e-commerce, online banking.	ML (supervised): Logistic Regression (LR), K-Nearest Neighbors (KNN), Decision Tree (DT), Random Forest (RF). DL (no supervisado): Autoencoder (AE). Techniques: PCA (feature extraction), SMOTE (oversampling), NearMiss (undersampling), GridSearchCV, 5-fold cross-validation.	RF + PCA + NearMiss (mejor): Acc 98%, AUROC 0.976, Avg Prec 0.927, Sens 0.83, Spec 1.00. LR + NearMiss (mejor precisión): AUROC 0.972, Avg Prec 0.928. Autoencoder (5b - mejor AE): Acc 96%, AUROC 0.947, Avg Prec 0.809, Sens 0.81, Spec 0.98. Comparativa supervisada (PCA): - LR: AUROC 0.971, Avg Prec 0.927 - KNN: AUROC 0.933, Avg Prec 0.865 - DT: AUROC 0.900, Avg Prec 0.858 - RF: AUROC 0.960, Avg Prec 0.921
Kolli, C.S. et al. [57]	UPI (Unified Payment Interface) transaction fraud, phishing, SIM swap fraud, social engineering, fake UPI apps, payment interface cloning, identity theft, unauthorized access to bank accounts.	Deep Learning Models: SNN (Spiking Neural Network) optimized with SCWO.Models de Machine Learning: RF (Random Forest), ML (Machine Learning), XFL (Explainable Federated Learning). Techniques: FL (Federated Learning), Borderline-SMOTE (oversampling), SCWO (Snow Carpet Weaver Optimization—a combination of SAO and CWO), Comprehensive Normalization.	SCWO-SNN (propuesto): Loss 0.119, NMSE 0.197, NRMSE 0.443, Prec 91.400%, Rec 91.275%, F1 91.448%. Comparativa (Training data=90%): - RF: Prec 77.626%, Rec 78.415%, F1 77.848% - XFL: Prec 79.966%, Rec 79.636%, F1 79.801% - Federated SDT: Prec 80.349%, Rec 80.389%, F1 80.663% - ML: Prec 82.114%, Rec 82.996%, F1 82.851%
Backiyalakshmi, V., & Umadevi, B.[58]	Fraud in the banking sector (bank accounts), credit/debit card fraud, identity theft, unauthorized transactions, insurance fraud, accounting fraud, and fraud involving bank card registrations.	Deep Learning Models: DCL-MDA (Dilated ConvLSTM + Multi-scale Dense Attention). Machine Learning Models: GAN, DRL, GBDT, GNN, ABC, HLBO, POA, MRSA. Techniques: PCA, t-SNE, Statistical Features, MRV-RSA (Modified Random Value Reptile Search Algorithm), weighted feature fusion.	DCL-MDA + MRV-RSA (mejor modelo): - Dataset 1: Acc 93.86%, Prec 2.57%, True fraud 3,200, False fraud 26,684,700, Error 6.34%. - Dataset 2: Acc 97.15%, Prec 18.91%, True fraud 300, False fraud 965,000, Error 4.03%.

The blue area includes machine learning, convolutional networks, graph models, financial fraud, and money laundering. The articles in this group compare different ways of identifying transactions that could be fraudulent.

The red group includes deep learning, recurrent neural networks, LSTM, and CNN. These models are used when fraudulent behavior depends on multiple transactions or relationships that are not easily recognized. Looking at all three groups, the map reveals that the review is primarily divided among explainability, fraud detection, and deep models.

E. Thematic Map

Fig. 5: In the figure, machine learning, artificial intelligence, finance, blockchain, fintech, and fraud detection appear among the most prevalent topics. This is because they recur in several of the analyzed articles. Deep learning and federated learning appear in another section of the map. Meanwhile, document examination, digital twin, metaverse, privacy, and security compliance are less prevalent within the analyzed dataset. A similar pattern is observed for ensemble learning, class imbalance, generative AI, and large language models, which do not yet have the same prominence as the central topics. The position of each group reveals that some research areas already have more studies, while others still appear less frequently.

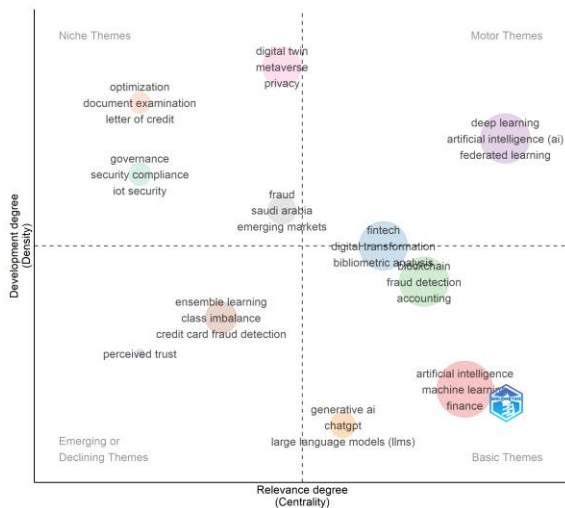


Fig. 5. Thematic map based on the centrality and density of the identified topics. (Source: Prepared by the author based on bibliometric analysis in Bibliometrix.)

V. DISCUSSION

The review shows that rule-based systems are still useful when fraud follows known patterns. The problem arises when those patterns change, or new forms of fraud emerge, because these systems are less adaptable. For this reason, the most recent studies increasingly rely on machine learning and deep learning models. Even so, the results do not allow us to identify a single model as the best in all cases.

For RQ1, Random Forest, XGBoost, Support Vector Machine, Logistic Regression, and Isolation Forest are among the most commonly used algorithms. In several studies, they achieve competitive results, although accuracy may not be very representative when fraudulent transactions are much less

frequent than normal ones. In these situations, metrics such as recall, F1-score, ROC-AUC, and the false positive rate allow for a better evaluation of the model's performance.

In RQ2, architectures such as LSTM, CNN, GNN, GRU, and autoencoders were identified. These techniques can analyze more complex relationships and transaction sequences, but they also require more resources and tend to be more difficult to interpret. This can become a limitation when the model is used in a financial environment where an alert must be processed and understood quickly.

RQ3 revealed that traditional methods retain the advantage of being easy to interpret, although they are more difficult to adapt to changes in fraudulent behavior. Mill et al. point out that the increase in the volume of information and the emergence of new forms of fraud make it necessary to supplement these systems with machine learning techniques. However, a more complex model does not necessarily guarantee a better result.

In RQ4, SHAP and LIME make it possible to show which variables carried the most weight in an alert. However, this does not mean that the explanation is necessarily useful to the analyst. Diaz et al. indicate that there is still little empirical evidence on how AML specialists use these explanations in real-world settings[59]. Hosawi and Stone also caution that an unclear explanation can increase cognitive load or lead to overconfidence in the system[60].

Finally, RQ5 shows that there are still limitations to evaluating XAI in real-time fraud detection scenarios. Mill et al. indicate that factors such as false positives, class imbalance, concept drift, and real-time analysis should be considered together [61]. It is not enough to measure the algorithm's speed; it is also necessary to consider how much time the analyst needs to understand the alert and make a decision.

VI. CONCLUSION

Based on the reviewed studies, it was found that machine learning and deep learning can help detect fraud that rule-based systems alone do not always recognize. Machine learning models are generally easier to implement and explain, while deep learning models can identify more complex relationships among transactions.

It also became clear that high accuracy alone does not guarantee a good outcome. In this type of problem, it is important to know how many instances of fraud the model detects, how many false positives it generates, and whether it can respond quickly.

XAI techniques can help explain why a transaction was flagged as suspicious. However, several studies have not demonstrated whether these explanations actually assist the analyst in making a decision. Another limitation is that many studies use public or synthetic data, so it cannot be guaranteed that the same performance will be achieved in a real bank or payment platform.

Based on the review, the combination of machine learning, deep learning, and XAI can be useful for improving fraud detection. However, further testing with real data, larger transaction volumes, and the involvement of financial specialists

is still needed. This would make it possible to verify whether the models perform well outside of laboratory tests and whether their results can be put to practical use.

REFERENCES

- [1] D. Ampumuza, C. Katushabe, and M. Tamale, "A systematic review and future directions for AI-driven detection of fraud patterns in SACCO transactions," *Front. Artif. Intell.*, vol. 8, Jan. 2026, doi: 10.3389/frai.2025.1690482.
- [2] S. S. and C. S., "Comparative Study on Anomaly-Based Intrusion Detection using Deep Learning Techniques," *EAI Endorsed Transactions on Internet of Things*, vol. 11, Nov. 2024, doi: 10.4108/eetiot.7178.
- [3] B. Fu, Y. Tong, Y. Li, Z. Tang, Z. Shang, and A. Li, "Financial Fraud Anomaly Detection of Listed Companies Based on Probabilistic Perspective Machine Learning Models," *Procedia Comput. Sci.*, vol. 266, pp. 979–986, 2025, doi: 10.1016/j.procs.2025.08.121.
- [4] Y. Wang, D. Lian, and E. Chen, "Fraud Detection Based on Credit Review Texts with Dual Channel Memory Networks," *Applied Artificial Intelligence*, vol. 38, no. 1, Dec. 2024, doi: 10.1080/08839514.2024.2385854.
- [5] M. R. Raja, M. A. Hosen, M. F. Kabir, S. Sultana, S. A. Ashraf, and R. Islam, "Detecting and Preventing Money Laundering Using Deep Learning and Graph Analysis," *International Journal of Advanced Computer Science and Applications*, vol. 16, no. 6, 2025, doi: 10.14569/IJACSA.2025.0160601.
- [6] Y. Chen, C. Zhao, Y. Xu, C. Nie, and Y. Zhang, "Deep Learning in Financial Fraud Detection: Innovations, Challenges, and Applications," *Data Science and Management*, Aug. 2025, doi: 10.1016/j.dsm.2025.08.002.
- [7] Y. Li, B. Fu, Y. Tong, Z. Tang, Z. Shang, and A. Li, "Abnormal Detection of Financial Fraud in Listed Companies Based on Deep Learning," *Procedia Comput. Sci.*, vol. 242, pp. 1402–1409, 2024, doi: 10.1016/j.procs.2024.08.112.
- [8] J. Nicholls, A. Kuppa, and N.-A. Le-Khac, "Financial Cybercrime: A Comprehensive Survey of Deep Learning Approaches to Tackle the Evolving Financial Crime Landscape," *IEEE Access*, vol. 9, pp. 163965–163986, 2021, doi: 10.1109/ACCESS.2021.3134076.
- [9] J. Černevičienė and A. Kabašinskas, "Explainable artificial intelligence (XAI) in finance: a systematic literature review," *Artif. Intell. Rev.*, vol. 57, no. 8, p. 216, Jul. 2024, doi: 10.1007/s10462-024-10854-8.
- [10] B. Lebichot, W. Siblini, G. M. Paldino, Y.-A. Le Borgne, F. Oblé, and G. Bontempi, "Assessment of catastrophic forgetting in continual credit card fraud detection," *Expert Syst. Appl.*, vol. 249, p. 123445, Sep. 2024, doi: 10.1016/j.eswa.2024.123445.
- [11] H. M. R. Al Lawati et al., "An Integrated Preprocessing and Drift Detection Approach With Adaptive Windowing for Fraud Detection in Payment Systems," *IEEE Access*, vol. 13, pp. 92036–92056, 2025, doi: 10.1109/ACCESS.2025.3569609.
- [12] S. K. Aljunaid, S. J. Almheiri, H. Dawood, and M. A. Khan, "Secure and Transparent Banking: Explainable AI-Driven Federated Learning Model for Financial Fraud Detection," *Journal of Risk and Financial Management*, vol. 18, no. 4, p. 179, Mar. 2025, doi: 10.3390/jrfm18040179.
- [13] V. Backiyalakshmi and B. Umadevi, "MRV-RSA: Developed Modified Random Value Reptile Search Algorithm and Deep Learning based Fraud Detection Model in Banking Sector," *Data Knowl. Eng.*, vol. 163, p. 102557, May 2026, doi: 10.1016/j.datak.2026.102557.
- [14] V. Chang, L. M. T. Doan, A. Di Stefano, Z. Sun, and G. Fortino, "Digital payment fraud detection methods in digital ages and Industry 4.0," *Computers and Electrical Engineering*, vol. 100, p. 107734, May 2022, doi: 10.1016/j.compeleceng.2022.107734.
- [15] U. G. Ketenci, T. Kurt, S. Onal, C. Erbil, S. Akturkoglu, and H. S. Ilhan, "A Time-Frequency Based Suspicious Activity Detection for Anti-Money Laundering," *IEEE Access*, vol. 9, pp. 59957–59967, 2021, doi: 10.1109/ACCESS.2021.3072114.
- [16] Md. S. Mia, S. Roy, M. A. Ihsan, S. Hossain, and Md. K. U. Ahamed, "Data-driven financial fraud detection using hybrid artificial and quantum intelligence," *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, vol. 5, no. 4, p. 100252, Dec. 2025, doi: 10.1016/j.tbench.2025.100252.
- [17] N. K. Narasapuram, S. A. A. Shah, M. F. Shiratuddin, and F. Soheli, "Explainable artificial intelligence models for detecting suspicious bank transactions," *International Journal of Machine Learning and Cybernetics*, vol. 17, no. 3, p. 111, Mar. 2026, doi: 10.1007/s13042-026-02994-w.
- [18] E. Y. Wirawan, M. A. Rahman, and E. Kurniawan, "Artificial Intelligence Implementation in Fraud Detection for Financial Due Diligence Project Financing in Indonesia Company," 2026, pp. 273–284. doi: 10.1007/978-3-032-01133-6_19.
- [19] P. Kulkarni, P. Pathak, S. Pillai, and V. Tigga, "Role of Generative AI for Fraud Detection and Prevention," in *Generative Artificial Intelligence in Finance*, Wiley, 2025, pp. 175–198. doi: 10.1002/9781394271078.ch10.
- [20] E. A. L. Marazqah Btoush, X. Zhou, R. Gururajan, K. C. Chan, R. Genrich, and P. Sankaran, "A systematic review of literature on credit card cyber fraud detection using machine and deep learning," *PeerJ Comput. Sci.*, vol. 9, p. e1278, Apr. 2023, doi: 10.7717/peerj-cs.1278.
- [21] V. L. Narayana Gorle and S. Panigrahi, "An efficient heuristic optimization-based fraudulent activities detection in the financial sector using adaptive machine learning and deep learning system," *Expert Syst. Appl.*, vol. 302, p. 130551, Mar. 2026, doi: 10.1016/j.eswa.2025.130551.
- [22] W. Y. Leong, Y. Z. Leong, and W. S. Leong, "Artificial Intelligence-Driven Fraud Detection: Enhancing Security in Digital Age," 2026, pp. 65–78. doi: 10.1007/978-981-95-2113-5_7.
- [23] H. Ahmad, E. Rawashdeh, A. AlTawil, and N. Al-Ramah, "EFC-Tomek: An effective undersampling technique for credit card fraud detection," *International Journal of Data and Network Science*, vol. 9, no. 4, pp. 845–852, 2025, doi: 10.5267/j.ijdns.2025.7.003.
- [24] M. J. Page et al., "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," *BMJ*, p. n71, Mar. 2021, doi: 10.1136/bmj.n71.
- [25] A. A. Veroniki et al., "Updating the PRISMA reporting guideline for network meta-analysis: a scoping review," *J. Clin. Epidemiol.*, vol. 188, p. 111985, Dec. 2025, doi: 10.1016/j.jclinepi.2025.111985.
- [26] M. G. Belu and A. M. Marinioiu, "AI-Enabled Supply Chain Management: A Bibliometric Analysis Using VOSviewer and RStudio Bibliometrix Software Tools," *Sustainability*, vol. 17, no. 5, p. 2092, Feb. 2025, doi: 10.3390/su17052092.
- [27] M. Rodrigues, A. D. Daniel, and M. Franco, "What is important to know about mumpreneurship? A bibliometric analysis," *International Journal of Organizational Analysis*, vol. 31, no. 7, pp. 3413–3435, Nov. 2023, doi: 10.1108/IJOA-05-2022-3293.
- [28] D. Ampumuza, C. Katushabe, and M. Tamale, "A systematic review and future directions for AI-driven detection of fraud patterns in SACCO transactions," *Front. Artif. Intell.*, vol. 8, Jan. 2026, doi: 10.3389/frai.2025.1690482.
- [29] J. Vazallo-Bautista, A. Villalobos-Peña, and J. Soria-Quijate, "Intelligent Systems, Machine Learning, and Deep Learning Algorithms for Detecting Banking Fraud: A Review," *International Journal of Advanced Computer Science and Applications*, vol. 16, no. 12, 2025, doi: 10.14569/IJACSA.2025.0161266.
- [30] M. N. Alatawi, "Detection of fraud in IoT based credit card collected dataset using machine learning," *Machine Learning with Applications*, vol. 19, p. 100603, Mar. 2025, doi: 10.1016/j.mlwa.2024.100603.
- [31] L. Abd-Ellatif, M. Abrar, and A. A. K. Ismaeel, "ATAD-Net: An Adaptive Deep Learning Framework for Real-Time Financial Fraud Detection," *Advances in Artificial Intelligence and Machine Learning*, vol. 05, no. 02, pp. 3988–4003, 2025, doi: 10.54364/AAIML.2025.52225.
- [32] N. Thair Ali, S. Jasim Hasan, A. Ghandour, and Z. Salam Al-Hchimy, "Improving credit card fraud detection using machine learning and GAN technology," *BIO Web Conf.*, vol. 97, p. 00076, Apr. 2024, doi: 10.1051/bioconf/20249700076.
- [33] W. Y. Leong, Y. Z. Leong, and W. S. Leong, "Artificial Intelligence-Driven Fraud Detection: Enhancing Security in Digital Age," 2026, pp. 65–78. doi: 10.1007/978-981-95-2113-5_7.
- [34] R. El Youbi, F. Messaoudi, M. Loukili, and R. Loukili, "From Click to Checkout: Deep Learning for Real-Time Fraud Detection in E-Payment

- Systems,” *Statistics, Optimization & Information Computing*, vol. 14, no. 6, pp. 3398–3408, Sep. 2025, doi: 10.19139/soic-2310-5070-2891.
- [35] H. M. R. Al Lawati et al., “An Integrated Preprocessing and Drift Detection Approach With Adaptive Windowing for Fraud Detection in Payment Systems,” *IEEE Access*, vol. 13, pp. 92036–92056, 2025, doi: 10.1109/ACCESS.2025.3569609.
- [36] O. Pavlova, V. Askerov, B. Tomchyshen, and T. Hovorushchenko, “Intelligent AML systems: using deep learning to detect financial fraud,” *TalTech Journal of European Studies*, vol. 15, no. 3, pp. 102–122, Dec. 2025, doi: 10.2478/bjes-2025-0032.
- [37] S. Iqbal, K. M. Awan, S. Kamal, and Z. U. Rehman, “Interpretable Ensemble Learning Models for Credit Card Fraud Detection,” *Applied Sciences*, vol. 15, no. 22, p. 12073, Nov. 2025, doi: 10.3390/app152212073.
- [38] R. R. Devi, J. E. Raja, and Y. B. Chin, “Reinforcement learning with graph neural network (RL-GNN) fusion for real-time financial fraud detection: a context-aware community mining approach,” *Sci. Rep.*, vol. 15, no. 1, p. 42953, Dec. 2025, doi: 10.1038/s41598-025-25200-3.
- [39] J. Černevičienė and A. Kabašinskas, “Explainable artificial intelligence (XAI) in finance: a systematic literature review,” *Artif. Intell. Rev.*, vol. 57, no. 8, p. 216, Jul. 2024, doi: 10.1007/s10462-024-10854-8.
- [40] B. S. Prashanth et al., “Prediction of bank transaction fraud using TabNet—an adaptive deep learning architecture,” *International Review of Economics & Finance*, vol. 106, p. 104916, Mar. 2026, doi: 10.1016/j.iref.2026.104916.
- [41] C. O. Kungu, K. Senagi, and E. Omondi, “Hybrid deep learning for anti-money laundering: Unsupervised detection of emerging schemes via feature fusion and explainable artificial intelligence,” *Machine Learning with Applications*, vol. 23, p. 100856, Mar. 2026, doi: 10.1016/j.mlwa.2026.100856.
- [42] E. Karnavou, G. Cascavilla, G. Marcelino, and Z. Geradts, “I know you’re a fraud: Uncovering illicit activity in a Greek bank transactions with unsupervised learning,” *Expert Syst. Appl.*, vol. 288, p. 128148, Sep. 2025, doi: 10.1016/j.eswa.2025.128148.
- [43] R. K. Gupta et al., “Enhanced framework for credit card fraud detection using robust feature selection and a stacking ensemble model approach,” *Results in Engineering*, vol. 26, p. 105084, Jun. 2025, doi: 10.1016/j.rineng.2025.105084.
- [44] M. Hasan, M. S. Rahman, M. J. Morshed Chowdhury, and I. H. Sarker, “CNN Based Deep Learning Modeling with Explainability Analysis for Detecting Fraudulent Blockchain Transactions,” *Cyber Security and Applications*, vol. 3, p. 100101, Dec. 2025, doi: 10.1016/j.csa.2025.100101.
- [45] Y. Chen, C. Zhao, Y. Xu, C. Nie, and Y. Zhang, “Deep Learning in Financial Fraud Detection: Innovations, Challenges, and Applications,” *Data Science and Management*, Aug. 2025, doi: 10.1016/j.dsm.2025.08.002.
- [46] G. Murugaraj, R. Krishnan, L. Hari Krishnan, and M. Sellapan Kandasamy, “Systematic Review on the Influence of Classical and Quantum Machine Learning in the Financial Sector,” *Data Science and Management*, Jan. 2026, doi: 10.1016/j.dsm.2026.01.001.
- [47] Md. S. Mia, S. Roy, M. A. Ihsan, S. Hossain, and Md. K. U. Ahamed, “Data-driven financial fraud detection using hybrid artificial and quantum intelligence,” *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, vol. 5, no. 4, p. 100252, Dec. 2025, doi: 10.1016/j.tbench.2025.100252.
- [48] N. Uddin, “Role of AI in preventing financial crime: A comprehensive analytical review,” *Journal of Economic Criminology*, vol. 13, p. 100200, Sep. 2026, doi: 10.1016/j.jeconc.2025.100200.
- [49] A. Usman, N. Naveed, and S. Munawar, “Intelligent Anti-Money Laundering Fraud Control Using Graph-Based Machine Learning Model for the Financial Domain,” *Journal of Cases on Information Technology*, vol. 25, no. 1, pp. 1–20, Jan. 2023, doi: 10.4018/JCIT.316665.
- [50] E. A. L. Marazqah Btoush, X. Zhou, R. Gururajan, K. C. Chan, R. Genrich, and P. Sankaran, “A systematic review of literature on credit card cyber fraud detection using machine and deep learning,” *PeerJ Comput. Sci.*, vol. 9, p. e1278, Apr. 2023, doi: 10.7717/peerj-cs.1278.
- [51] H. Ahmad, E. Rawashdeh, A. AlTawil, and N. Al-Ramah, “EFC-Tomek: An effective undersampling technique for credit card fraud detection,” *International Journal of Data and Network Science*, vol. 9, no. 4, pp. 845–852, 2025, doi: 10.5267/j.ijdns.2025.7.003.
- [52] Z. Miao, “Financial Fraud Detection and Prevention,” *Journal of Organizational and End User Computing*, vol. 36, no. 1, pp. 1–27, Sep. 2024, doi: 10.4018/JOEUC.354411.
- [53] A. A. Khan et al., “A drone-based data management and optimization using metaheuristic algorithms and blockchain smart contracts in a secure fog environment,” *Computers and Electrical Engineering*, vol. 102, p. 108234, Sep. 2022, doi: 10.1016/j.compeleceng.2022.108234.
- [54] Y. Wang, Q. Zheng, X. Li, L. Wang, and L. Lin, “CoSemiGNN: Blockchain fraud detection with dynamic graph neural networks based on co-association of semi-supervised,” *Expert Syst. Appl.*, vol. 298, p. 129853, Mar. 2026, doi: 10.1016/j.eswa.2025.129853.
- [55] B. Lebichot, W. Siblini, G. M. Paldino, Y.-A. Le Borgne, F. Oblé, and G. Bontempi, “Assessment of catastrophic forgetting in continual credit card fraud detection,” *Expert Syst. Appl.*, vol. 249, p. 123445, Sep. 2024, doi: 10.1016/j.eswa.2024.123445.
- [56] V. Chang, L. M. T. Doan, A. Di Stefano, Z. Sun, and G. Fortino, “Digital payment fraud detection methods in digital ages and Industry 4.0,” *Computers and Electrical Engineering*, vol. 100, p. 107734, May 2022, doi: 10.1016/j.compeleceng.2022.107734.
- [57] C. S. Kolli, V. Jagannathan, D. R. Lanka, C. Rajyalakshmi, V. K. R. Radha, and V. V. K. Reddy, “SCWO-SNN: Snow carpet weaver optimization based spiking neural network for fraud detection in UPI transactions using federated learning,” *Comput. Stand. Interfaces*, vol. 98, p. 104166, Aug. 2026, doi: 10.1016/j.csi.2026.104166.
- [58] V. Backiyalakshmi and B. Umadevi, “MRV-RSA: Developed Modified Random Value Reptile Search Algorithm and Deep Learning based Fraud Detection Model in Banking Sector,” *Data Knowl. Eng.*, vol. 163, p. 102557, May 2026, doi: 10.1016/j.datak.2026.102557.
- [59] J. Diaz et al., “Dynamic Trust Modulation and Human Oversight in AI-Driven AML Systems: A Conceptual Framework for Compliance,” *International Journal of Advanced Computer Science and Applications*, vol. 16, no. 12, 2025, doi: 10.14569/IJACSA.2025.0161209.
- [60] A. Hosawi and R. Stone, “Enhancing Trust in Human-AI Collaboration: A Conceptual Review of Operator-AI Teamwork,” *International Journal of Advanced Computer Science and Applications*, vol. 16, no. 7, 2025, doi: 10.14569/IJACSA.2025.0160701.
- [61] E. Mill, W. Gam, N. Ryman-Tubb, and C. Turner, “Opportunities in Real Time Fraud Detection: An Explainable Artificial Intelligence (XAI) Research Agenda,” *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 5, 2023, doi: 10.14569/IJACSA.2023.01405121.