Journal of
Cyber Security

Tech Science Press

Doi:10.32604/jcs.2026.074265

REVIEW

A Systematic Review of Frameworks for the Detection and Prevention of Card-Not-Present (CNP) Fraud

Kwabena Owusu-Mensah*, Edward Danso Ansong, Kofi Sarpong Adu-Manu and Winfred Yaokumah

Department of Computer Science, College of Basic and Applied Sciences, University of Ghana, Legon, Accra, P.O. Box LG 25, Ghana

*Corresponding Author: Kwabena Owusu-Mensah. Email: kowusu-mensah001@st.ug.edu.gh

Received: 07 October 2025; Accepted: 09 December 2025; Published: 20 January 2026

ABSTRACT: The rapid growth of digital payment systems and remote financial services has led to a significant increase in Card-Not-Present (CNP) fraud, which is now the primary source of card-related losses worldwide. Traditional rule-based fraud detection methods are becoming insufficient due to several challenges, including data imbalance, concept drift, privacy concerns, and limited interpretability. In response to these issues, a systematic review of twenty-four CNP fraud detection frameworks developed between 2014 and 2025 was conducted. This review aimed to identify the technologies, strategies, and design considerations necessary for adaptive solutions that align with evolving regulatory standards. The findings indicate a shift from static, supervised models to dynamic approaches, such as hybrid and federated architectures, which utilize advanced technologies like Graph Neural Networks (GNNs), blockchain auditing, and privacy-preserving learning. These modern frameworks demonstrate impressive performance metrics, achieving F1 scores between 0.85 and 0.99 and AUC values exceeding 0.93, while also complying with regulatory standards, including GDPR and PCI-DSS. The review identified six key design pillars essential for effective CNP fraud mitigation: scalable architecture, privacy-preserving governance, adaptive learning, interpretability, cost optimization, and integrated continuous evaluation. This study presents a design-centric framework that emphasizes scalability, ethical governance, and explainable intelligence. The review suggests that graph-enabled, federated, and self-optimizing frameworks represent the future of securing digital payment environments and enhancing CNP fraud detection.

KEYWORDS: Card-not-present fraud; CNP fraud taxonomy; fraud detection frameworks; graph neural networks; federated learning; blockchain-based security; explainable artificial intelligence; machine learning for cybersecurity; digital payment systems

1 Introduction

The rapid digitization of global commerce has fundamentally transformed the mechanisms of financial transactions, resulting in both unprecedented levels of convenience and a marked expansion of the associated threat landscape. One of the most prevalent and economically detrimental of these threats is Card-Not-Present (CNP) fraud, in which malicious actors exploit online and mobile platforms to initiate unauthorized transactions without the physical presence of a payment card.

Globally reported losses due to card fraud have seen a significant increase, rising from $18.11 billion in 2014 to an estimated $35.79 billion by 2024, with projections indicating a further escalation to $43.47 billion by 2028 [1]. Within this total, losses attributable to card-not-present (CNP) transactions have escalated from approximately $10 billion (representing 55% of total losses) in 2014 to around $27 billion (approximately 74%) in 2024, with expectations that these losses will surpass $30 billion annually by 2028 [2]. This upward trend underscores the urgent need for enhanced security measures within the online payment ecosystem and highlights the necessity for the development of CNP-specific detection tools grounded in a comprehensive understanding of relevant technologies, methodologies, and system-level design principles. In response to these challenges, a diverse array of solutions has begun to emerge, including machine-learning classifiers, federated-learning systems, and graph-based anomaly detectors. Despite this proliferation of solutions, significant gaps persist regarding systematic organization, standardized evaluation metrics, and design considerations focused on practical deployment [3,4]. As tactics employed by fraudsters evolve, many existing frameworks exhibit limitations in adaptability, explainability, and regulatory compliance, which further reinforces the imperative to address these deficiencies.

Prior surveys in the domain of financial fraud analytics have predominantly offered broad, model-centric overviews, such as taxonomies of artificial intelligence (AI) and machine learning, without isolating the specific contexts of CNP transactions or the unique constraints associated with remote payments [5]. Also, the literature remains primarily focused on accuracy metrics, which can produce overly optimistic outcomes due to issues related to dataset handling or the limitations inherent in proxy transformations that may weaken their applicability to real-time CNP operations [3,6,7]. Simultaneously, surveys that emphasize privacy and governance typically address attack taxonomies and countermeasures at the learning layer but generally fail to provide an end-to-end, framework-level treatment of CNP pipelines [8].

In contrast, this review is specifically focused on CNP fraud, framework-oriented, and mindful of deployment considerations. It: (i) synthesizes twenty-four CNP frameworks into a structured, multi-layer taxonomy that encompasses both current and emerging technologies; (ii) extracts actionable design principles from production-grade exemplars, such as streaming and evolving graph pipelines, to guide system development; (iii) proposes a multidimensional performance envelope that incorporates factors such as latency, throughput, cost sensitivity, drift resilience, explainability, and privacy/compliance, thereby addressing the limitations associated with accuracy-only benchmarking; and (iv) delineates future research directions through gap analysis and establishes a coherent research agenda.

The urgency of this agenda is underscored by the continuing rise and concentration of losses within CNP channels, which currently account for the majority of global card fraud losses and are anticipated to grow further. This trend emphasizes the critical need for deployment-grade, privacy-preserving solutions tailored to the complexities of CNP transactions.

Rationale and Contributions of This Review

This review is motivated by several convergent challenges within the domain of CNP fraud detection. First, the field currently lacks a unified taxonomy for its methodological frameworks, resulting in fragmented research endeavors that impede direct model comparison and hinder coherent scholarly progress. Second, the prevailing reliance on conventional performance metrics (e.g., accuracy, F1-score) fails to capture critical operational dimensions, such as computational latency, scalability, the economic impact of false positives, and regulatory adherence, which are paramount to the practical deployment and viability of these systems. Third, existing literature offers insufficient guidance on the architectural and design-level decisions necessary to ensure robustness, adaptability, and effective human-AI collaboration. While pioneering contributions, such as Scalable Real-time Credit Card Fraud Finder (SCARFF) [9] and Federated Learning, Graph Attention Networks and Delineated Convolutional Networks (FedGAT-DCNN) [10], have demonstrated scalable and privacy-aware designs, they often remain disconnected from the complexities of production environments. Finally, in the face of rapidly evolving fraud tactics and financial regulations, there is a pressing need to systematically identify extant literature gaps and to chart a course for integrating emerging technological paradigms.

Consequently, this review seeks to consolidate, extend, and reorient contemporary CNP fraud detection research toward the imperatives of practical implementation, system-wide resilience, and cross-institutional collaboration.

Building on this motivation, the review makes four principal, interlocking contributions:

1. **A Systematic Taxonomy of CNP Fraud Detection Frameworks.** This work introduces a structured taxonomy that classifies existing frameworks along two primary axes: technological architecture (e.g., machine learning, federated learning, graph neural networks, blockchain) and methodological approach (e.g., supervised learning, unsupervised anomaly detection, ensemble modeling). This taxonomy not only facilitates a systematic comparison of extant techniques but also provides a conceptual scaffold for guiding future research and development aimed at mitigating CNP fraud.
2. **A Foundational Set of Design Principles for CNP Systems.** Derived from a comprehensive, cross-comparative analysis of 24 distinct frameworks, this review distills a set of core design principles essential for next-generation fraud detection systems. These principles recommend multi-layered, modular architecture; real-time stream processing capabilities; adaptive learning pipelines to counteract concept drift; privacy-by-design methodologies; and the strategic integration of human-in-the-loop interfaces.
3. **A Novel Multidimensional Framework for Performance Evaluation.** Moving beyond traditional accuracy-based metrics, this review proposes a comprehensive suite of performance indicators that reflect the multifaceted demands of real-world deployment. This framework incorporates dimensions of cost-sensitive utility, detection latency, explainability, drift resilience, and privacy-compliance. By adopting this holistic evaluation paradigm, stakeholders can better assess a system's operational efficacy, regulatory alignment, and ethical considerations.
4. **A Critical Gap Analysis and Forward-Looking Research Agenda.** This work provides a critical synthesis of salient gaps in the current literature, highlighting deficiencies in areas such as concept drift handling, system scalability, empirical deployment validation, and explainability. In response, it proposes a strategic research agenda that prioritizes the exploration of emerging fields, including federated graph learning, explainable AI (XAI) frameworks, compliance-aware architectures, and secure, collaborative fraud intelligence sharing.

Collectively, these elements re-center CNP-fraud detection on system-level design and operational viability, offering a structured map for scholars and a practical blueprint for researchers, practitioners, and policymakers.

2 Card-Not-Present (CNP) Fraud

Card-not-present (CNP) fraud refers to the use of payment credentials for online or remote transactions where the cardholder is not physically present. Unlike card-present fraud, which is effectively mitigated by chip-and-PIN technology and Europay, MasterCard, and Visa (EMV) standards, CNP fraud depends entirely on digital verification methods. This reliance renders it particularly vulnerable to exploitation by cybercriminals [5]. The incidence of CNP fraud has surged in tandem with the global expansion of e-commerce, mobile banking, and digital wallets, posing a significant challenge for merchants, financial institutions, and regulatory bodies.

The lack of physical verification enables attackers to exploit vulnerabilities in digital channels, utilizing stolen, fabricated, or manipulated payment credentials. Furthermore, the globalization of payment systems, combined with the increasing sophistication of fraud tactics, exacerbates these risks. Fraudsters increasingly employ a hybrid approach, merging technical exploits, such as automated botnets and data breaches, with human-centric strategies, including phishing and social engineering, to circumvent detection systems [5,11].

In our analysis, we categorize the major forms of CNP fraud into four overarching groups, reflecting both traditional and emergent dimensions of the threat landscape.

- *Identity-based attacks* involve the misuse or fabrication of customer credentials, including identity theft, account takeover, and synthetic identity fraud, all of which exploit weaknesses in authentication systems and customer verification processes.
- *Social engineering schemes* target the human element of payments, with fraudsters using deception through phishing, smishing, vishing, triangulation fraud, or even deliberate chargeback disputes (friendly fraud) to manipulate customers and merchants.
- *System-level exploits* focus on technological vulnerabilities within merchant platforms and payment infrastructures. These include merchant-side data breaches, botnet-driven credential stuffing, and large-scale automated attacks that overwhelm fraud filters.
- *Emerging threats* represent the newest frontier, where the rise of cross-border payments, digital wallets, and IoT ecosystems increasingly enables fraud. These environments introduce novel weaknesses in tokenization, device fingerprinting, and biometric verification.

Each group encompasses multiple subtypes with distinct attack mechanisms, yet together they reveal a common theme: fraudsters adapt rapidly by combining technical sophistication with social manipulation to circumvent defenses. Fig. 1 illustrates a taxonomy of fraud types, presenting a structured visualization of their classification. Concurrently, Table 1 offers a comparative summary of their defining characteristics, detection challenges, and notable studies from the literature. This framework enhances the understanding of CNP fraud and equips researchers and practitioners with the insights necessary to develop targeted detection and prevention strategies.

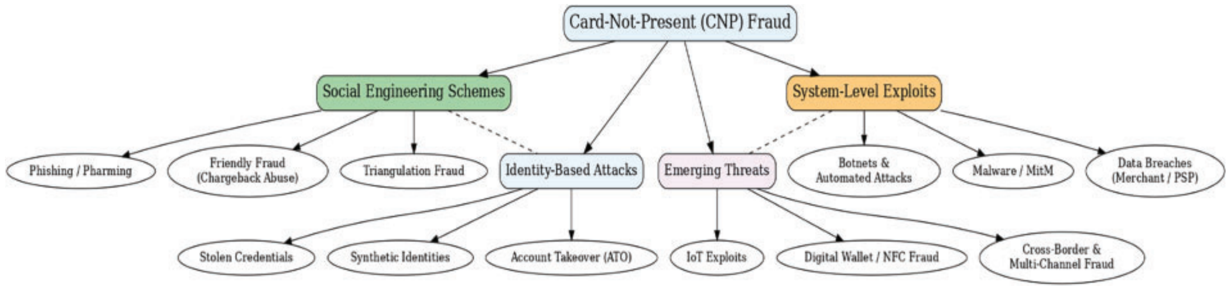
Card-Not-Present (CNP) Fraud
Social Engineering Schemes
System-Level Exploits
Phishing / Pharming
Friendly Fraud (Chargeback Abuse)
Triangulation Fraud
Identity-Based Attacks
Emerging Threats
Botnets & Automated Attacks
Malware / MitM
Data Breaches (Merchant / PSP)
Stolen Credentials
Synthetic Identities
Account Takeover (ATO)
IoT Exploits
Digital Wallet / NFC Fraud
Cross-Border & Multi-Channel Fraud

Figure 1: Taxonomy of Card-Not-Present (CNP) fraud types organized into four major categories

Table 1: Comparative summary of CNP fraud types

Fraud type	Defining characteristics	Detection challenges
Identity theft/Account Takeover (ATO)	Stolen or leaked credentials are used to hijack accounts and initiate unauthorized transactions.	Transactions appear legitimate since they originate from genuine customer profiles.
Phishing & social engineering	Fraudsters use deceptive emails, SMS messages, or fake websites to trick victims into revealing their card data.	Highly scalable, it exploits human trust; however, it is hard to mitigate technologically alone.
Friendly fraud (Chargeback Fraud)	Genuine customers falsely dispute legitimate transactions to obtain refunds.	Difficult to distinguish from legitimate disputes; high financial loss to merchants.

(Continued)

Table 1 (continued)

Fraud type	Defining characteristics	Detection challenges
Synthetic identity fraud	A combination of real and fabricated data to create new, seemingly valid identities.	Hard to detect as synthetic profiles pass conventional KYC checks.
Botnet & automated attacks	Automated scripts or IoT botnets are used for large-scale credential stuffing and card testing.	High-velocity, distributed attacks evade IP-based filters; near-real-time detection is needed.
Merchant-side attacks & Data breaches	Exploitation of weak merchant/payment gateway systems to steal stored card data.	Large-scale data exfiltration; systemic impacts across networks.
Triangulation fraud	Fraudsters set up fake storefronts, order goods using stolen credit cards, and capture consumer information.	Transactions appear valid to merchants; detection requires network-level correlation.
Cross-border & Multi-channel fraud	Fraudsters exploit regulatory fragmentation and multiple transaction channels.	Inconsistent global fraud controls; cross-channel correlation is difficult.
Emerging threats (IoT & Digital Wallet Exploits)	Fraud through insecure IoT devices, weak tokenization, or mobile wallet hijacking.	Weak biometric/device authentication; Subscriber Identity Module (SIM) swaps and token provisioning flaws.

2.1 Types of CNP Fraud

2.1.1 Identity Theft and Account Takeover (ATO)

Identity theft and account takeover are among the most prevalent forms of CNP fraud, driven by the illicit acquisition and misuse of customer credentials. Attackers typically gain access to sensitive information, such as card numbers, card verification value (CVV) or card verification code (CVC) codes, billing addresses, and login credentials, through phishing campaigns, malware infections, large-scale data breaches, or black-market purchases [5,12].

Once stolen data is obtained, criminals can initiate ATO by hijacking legitimate customer profiles. This enables them to perform unauthorized transactions, alter delivery addresses, exploit stored payment methods, or launder funds through mule accounts. Unlike synthetic identity fraud, which constructs entirely new profiles, ATO exploits existing accounts, making detection particularly challenging. Transactions often appear legitimate, as they originate from trusted devices, IP addresses, or historical user accounts.

Recent studies show that account takeover (ATO) attacks are highly automated, with adversaries using credential-stuffing botnets to replay large volumes of stolen credentials across multiple merchant platforms at scale [13,14]. This automation substantially increases downstream exposure to card-not-present (CNP) fraud and undermines static and rule-based fraud detection systems. Compounding the threat, compromised accounts are often resold across cybercrime forums, creating an underground economy where user profiles have monetary value depending on their transaction history and geographic region.

From a prevention perspective, multi-factor authentication (MFA), behavioral biometrics, and anomaly detection algorithms have been proposed to counteract ATO. However, adoption remains uneven, and sophisticated attackers have found ways to bypass one-time passwords (OTPs) or exploit weak biometric implementations. This highlights the necessity of multi-layered defenses that combine technical measures (such as device fingerprinting and continuous authentication) with behavioral analytics (including user keystroke dynamics and login patterns) to effectively mitigate account takeover risks.

2.1.2 Phishing and Social Engineering Attacks

Phishing and social engineering attacks remain the most widely used techniques in facilitating CNP fraud because they exploit the human element rather than technical vulnerabilities. Fraudsters employ deceptive strategies to trick individuals into voluntarily disclosing sensitive payment details, login credentials, or personally identifiable information (PII). These attacks manifest in multiple forms, including email phishing, SMS-based phishing (smishing), voice calls (vishing), and increasingly through social media platforms that impersonate legitimate institutions [5,11]. In phishing scenarios, attackers construct counterfeit websites or mobile applications that closely resemble authentic merchant or banking portals. Victims, believing the interface to be genuine, input their card numbers, CVV or CVC, and authentication details, which are then harvested in real time and used for fraudulent purchases. The growing use of URL shortening services and domain obfuscation techniques further complicates the ability of users to distinguish fraudulent links from legitimate ones.

Smishing and vishing attacks have surged alongside the expansion of mobile banking and contactless payments. In these cases, fraudsters impersonate customer service agents or financial institutions, convincing victims to disclose verification codes or reset credentials. This is often coupled with real-time social engineering, where attackers initiate CNP transactions while simultaneously guiding the victim into providing the required one-time password (OTP) or biometric confirmation.

A particularly damaging variant is phishing for cryptocurrency wallets, where fraudulent QR codes or spoofed wallet applications redirect funds to attacker-controlled addresses. Unlike traditional chargeback-protected payments, cryptocurrency transactions are irreversible, magnifying consumer losses [8]. Detecting phishing attacks remains challenging because they often exploit legitimate communication channels. Traditional blacklist-based filtering is insufficient, as attackers continuously generate new domains and adaptive content. To address these limitations, researchers and practitioners have proposed machine learning classifiers that analyze email headers, message content, and embedded links for anomalies [15]. Additionally, visual similarity detection algorithms compare suspect websites against known brand templates, flagging fraudulent lookalikes [16].

Despite these advances, phishing attacks continue to succeed because they target cognitive biases such as trust, urgency, and curiosity. As a result, preventive measures must go beyond technological solutions and incorporate user education, awareness campaigns, and regulatory frameworks. For instance, strong customer authentication (SCA) under the Payment Services Directive 2 (PSD2) directive mandates multi-factor verification in Europe, reducing the success of phishing-related CNP fraud.

2.1.3 Friendly Fraud (Chargeback Fraud)

Friendly fraud, often referred to as chargeback fraud, represents a paradoxical category of CNP fraud where the legitimate cardholder initiates fraudulent activity. Unlike identity theft or phishing attacks, which involve external actors, friendly fraud occurs when a genuine consumer disputes a legitimate transaction with the intent of reversing payment and retaining the purchased goods or services [9,17].

The mechanism is deceptively simple. After making a purchase, the cardholder contacts their issuing bank to claim that the transaction was unauthorized, the product was not delivered, or the service was unsatisfactory. Because consumer protection laws and card network regulations (such as Visa's Zero Liability Policy) heavily favor customers in disputed transactions, merchants are often left to bear the financial loss. This dynamic makes friendly fraud one of the most costly and contentious types of CNP fraud, especially for small and medium-sized e-commerce merchants [11].

Unlike merchant error or true fraud, friendly fraud is difficult to detect at the point of transaction because it originates from an authentic payment method, a legitimate shipping address, and verified user credentials. The gray area between intentional and unintentional chargebacks compounds the challenge. Some consumers unknowingly commit friendly fraud when they fail to recognize a charge on their billing statement, forget about a purchase, or when multiple family members share the same account. Others, however, deliberately exploit chargeback policies as a risk-free way to obtain goods and services without paying.

From a prevention standpoint, friendly fraud requires a multi-pronged strategy.

- Enhanced transaction documentation, such as delivery confirmations, digital receipts, and geo-tagged proof of service, can strengthen a merchant's defense during chargeback disputes [18].
- Chargeback alert systems, offered by payment processors, notify merchants in real time when disputes are filed, allowing them to issue refunds proactively and avoid fees.
- Machine learning–based behavioral analytics can flag suspicious refund or dispute patterns, such as repeat offenders or unusually high dispute rates from specific accounts [19].

At a regulatory level, initiatives such as PSD2's strong customer authentication (SCA) aim to reduce disputes by requiring robust verification of each transaction. Similarly, card networks are refining dispute resolution processes by distinguishing between legitimate claims and abuse, though enforcement remains inconsistent across jurisdictions.

2.1.4 Synthetic Identity Fraud

Synthetic identity fraud is one of the fastest-growing and most insidious forms of CNP fraud because it combines elements of legitimate personal information with fictitious data to create a new, seemingly valid identity. Unlike traditional identity theft, where a fraudster assumes complete control of an existing profile, synthetic identity fraud constructs an entirely new persona that passes many conventional verification checks [13,20].

The process typically begins when fraudsters obtain fragments of personal information, such as Social Security numbers, national identification numbers, or dates of birth, through data breaches or dark web markets. These fragments are then fused with fabricated details, including false names, addresses, and phone numbers, to establish a synthetic identity. Over time, the fraudster may nurture the profile by applying for small lines of credit, paying bills on time, and building a positive transaction history. This "grooming" phase allows the synthetic identity to appear legitimate within financial systems [17].

Once established, synthetic identities are used to commit CNP fraud in several ways:

- Transaction Fraud: Fraudsters use synthetic profiles to open accounts and make purchases that are eventually defaulted on.
- Credit Bust-Outs: After gaining trust and higher credit limits, attackers suddenly maximize available credit and disappear without repayment.
- Merchant Exploitation: Fraudsters use synthetic identities to establish merchant accounts and launder fraudulent transactions under the guise of legitimate businesses.

The difficulty of detecting synthetic identities lies in their hybrid nature: part real and part fabricated. Fraud detection systems that rely heavily on deterministic checks (e.g., matching name, date of birth, or address) often fail to flag these accounts because the genuine elements validate the synthetic profile. Furthermore, the rise of digital onboarding processes in banking and e-commerce, where remote identity verification is common, has amplified vulnerability [12].

Advanced detection approaches are increasingly exploring graph-based analytics and linkage analysis. By mapping relationships across devices, addresses, emails, and transaction histories, these systems can identify anomalies in network structures that suggest fabricated identities [21,22]. Similarly, machine learning algorithms trained on behavioral patterns, rather than static identifiers, are being employed to differentiate between authentic and synthetic profiles.

Despite these innovations, synthetic identity fraud remains a significant regulatory challenge. Because victims are often not immediately aware, since their full identity is not directly stolen, it may take years before fraudulent activity is detected. This delayed recognition results in significant financial losses for issuers, acquirers, and merchants, while complicating liability assignment in cross-border transactions.

To mitigate this risk, industry experts recommend a layered defense that combines:

- Advanced Know Your Customer (KYC) checks with biometric verification,
- Device fingerprinting to link transactions to consistent hardware or network characteristics,
- Consortium data-sharing between banks and payment providers to identify overlapping suspicious patterns, and
- Regulatory frameworks, such as PSD2 and GDPR, which mandate stricter identity validation and data protection.

2.1.5 Botnet and Automated Attacks

Botnet and automated attacks have become the dominant enabler of large-scale CNP fraud due to their speed, scalability, and ability to overwhelm traditional fraud detection mechanisms. Unlike phishing or synthetic identity fraud, which rely on social or identity manipulation, botnet attacks exploit the automation of fraudulent transactions through networks of compromised devices [21]. These devices, ranging from personal computers to Internet of Things (IoT) endpoints, are infected with malware and remotely controlled by attackers to execute thousands of transaction attempts simultaneously.

A common manifestation of this threat is the card testing (or "carding") attack, in which bots systematically attempt small transactions to validate stolen card details. If a transaction succeeds without being flagged, the card is marked as "live" and later used for higher-value purchases or sold on underground forums. Because these low-value tests often mimic legitimate consumer behavior, they are particularly challenging to detect in real-time [9].

Botnets are also central to credential stuffing attacks, where stolen username and password combinations are tested across multiple merchant sites to gain unauthorized access to customer accounts. Given the widespread reuse of credentials across platforms, these attacks frequently succeed, leading to account takeover (ATO) and downstream CNP fraud [23].

The rise of IoT-enabled botnets such as *Mirai* has further amplified the scale of attacks. Fraudsters can now mobilize millions of devices with minimal cost, creating distributed attack networks that evade IP-based detection methods. These IoT botnets often target weakly secured consumer devices such as routers, webcams, or smart home appliances, underscoring the growing convergence of cybersecurity and financial fraud [21].

Detecting botnet-driven fraud requires sophisticated behavioral analytics and velocity checks. For example:

- Transaction velocity monitoring identifies abnormal bursts of activity from single accounts, IPs, or devices.
- Device fingerprinting distinguishes legitimate users from bot-controlled scripts by tracking browser, hardware, and network attributes.
- Graph-based fraud detection models map relationships across IP addresses, devices, and transactions to uncover botnet patterns [21,22].

Despite these advances, challenges persist. Attackers increasingly employ human-in-the-loop botnets, where automated scripts handle the bulk of activities, but humans intervene during critical authentication steps (e.g., solving CAPTCHAs or providing stolen OTPs). This hybrid model blurs the distinction between machine-generated and genuine user behavior, raising false negatives in detection systems.

From a prevention perspective, multi-layered defenses are essential. Techniques such as reCAPTCHA challenges, rate limiting, behavioral biometrics, and federated learning approaches [16] can collectively reduce the impact of botnets. Moreover, cross-industry collaboration to share threat intelligence is critical for identifying evolving attack signatures across ecosystems.

2.1.6 Merchant-Side Attacks and Data Breaches

Merchant-side attacks and data breaches represent another critical source of CNP fraud, as they directly compromise the infrastructure that processes and stores sensitive payment data. Unlike phishing or botnet attacks that target consumers, merchant-side breaches exploit vulnerabilities in payment gateways, e-commerce platforms, and third-party service providers. Once attackers infiltrate these systems, they can exfiltrate large volumes of cardholder data, including Primary Account Numbers (PANs), CVV codes, billing addresses, and even authentication tokens, which are subsequently monetized on black markets or reused for fraudulent transactions [23].

A defining characteristic of merchant-side attacks is their scale and systemic impact. High-profile breaches, such as those affecting global retailers and payment processors, have resulted in the exposure of millions of card records at once. This not only amplifies the risk of downstream CNP fraud but also erodes consumer trust in digital commerce and leads to significant reputational and financial losses for affected organizations [9].

Common attack vectors include:

- SQL injection and web application exploits that allow unauthorized database access.
- API vulnerabilities, particularly in poorly secured mobile and e-commerce integrations, where weak authentication or misconfigured permissions expose payment data.
- Malware injections, such as form-jacking scripts (Magecart attacks), which intercept card details entered on checkout pages in real time.
- Insider threats, where employees or contractors abuse privileged access to extract or sell sensitive data.

Detecting and mitigating merchant-side attacks is particularly challenging due to the distributed and outsourced nature of modern payment ecosystems. Merchants often rely on third-party providers for payment processing, cloud hosting, and analytics, which increases the attack surface and complicates accountability [5]. Furthermore, compliance frameworks such as the Payment Card Industry Data Security Standard (PCI DSS) establish minimum requirements for data security; however, enforcement is inconsistent across regions and merchant categories.

Recent approaches to prevention emphasize tokenization and end-to-end encryption (E2EE). By replacing sensitive card details with randomly generated tokens, merchants reduce the risk of storing exploitable data. Similarly, encryption ensures that cardholder data is unreadable even if intercepted. Some frameworks also employ real-time intrusion detection systems (IDS) and anomaly-based monitoring to detect unusual activity within merchant networks.

At a strategic level, blockchain-based antifraud frameworks have been proposed to decentralize transaction validation and reduce reliance on centralized merchant databases [18]. Additionally, regulatory regimes such as GDPR and PSD2 impose stricter requirements on data handling and customer authentication, providing legal incentives for merchants to strengthen security practices.

2.1.7 Triangulation Fraud

Triangulation fraud is a sophisticated form of CNP fraud that exploits both legitimate merchants and unsuspecting consumers through a three-party deception model. In this scheme, the fraudster operates a fake online storefront that advertises popular goods at unusually low prices to lure unsuspecting customers. When a purchase is made, the fraudster uses stolen credit card details to place an identical order with a legitimate merchant, directing the shipment to the original buyer. The consumer receives the goods, believing the transaction to be genuine, while the fraudster retains the customer's personal and payment information for future exploitation [18].

This type of fraud is particularly deceptive because all three parties appear to engage in legitimate behavior:

- The consumer willingly purchases items online.
- The legitimate merchant processes a valid payment authorization and fulfills the order.
- The fraudster successfully masks their role by inserting themselves between the two.

The fraudster profits in two ways: first, by harvesting and reselling the consumer's card details and personal information; second, by building credibility for their fake storefront through successful deliveries, which enables them to scale future fraudulent operations.

Triangulation fraud poses unique detection challenges. Since the goods are shipped to the correct consumer, the merchant is initially unaware of any fraud, and the customer may not suspect wrongdoing until they discover unauthorized charges or are later targeted by further fraudulent activities. Additionally, because transactions at the merchant level appear legitimate, traditional fraud detection models, focused on transaction anomalies, may fail to detect the scheme in real time [9].

Preventive measures require interventions at multiple levels:

- Consumer education is critical, as many victims are drawn to unrealistic discounts or unfamiliar sellers. Raising awareness about secure shopping practices, such as verifying vendor legitimacy, checking Hypertext Transfer Protocol (HTTPS) protocols, and avoiding suspiciously low prices, can reduce susceptibility.
- Merchant monitoring can help by flagging multiple transactions linked to the same device, IP address, or delivery address but involving different credit cards, which may indicate triangulation patterns.
- Network-level intelligence sharing among merchants, card networks, and banks can identify coordinated fraud campaigns by correlating suspicious activities across different platforms.

Advanced detection frameworks increasingly rely on graph analytics to uncover hidden relationships between fraudulent storefronts, compromised cards, and consumer identities [21,22]. By analyzing transaction networks, these systems can identify clusters of activity consistent with triangulation fraud, even when individual transactions appear normal.

2.1.8 Cross-Border and Multi-Channel Fraud

Cross-border and multi-channel fraud represent complex and rapidly expanding dimensions of CNP fraud that exploit regulatory fragmentation, jurisdictional inconsistencies, and technological diversity across global payment systems. With the rapid growth of e-commerce and international digital trade, transactions are increasingly flowing across borders, often involving multiple intermediaries, such as payment processors, acquiring banks, and card networks. This interconnected environment provides fertile ground for fraudsters to exploit gaps in fraud detection, varying compliance standards, and delays in cross-jurisdictional coordination [9,17].

In cross-border CNP fraud, attackers often exploit weaker fraud prevention systems in specific regions. For example, markets with less stringent enforcement of Payment Services Directive 2 (PSD2) standards or limited adoption of strong customer authentication (SCA) are particularly vulnerable. Fraudsters may route transactions through these jurisdictions to bypass stricter controls in others, creating a form of regulatory arbitrage. Additionally, global merchants often struggle to strike a balance between fraud prevention and customer experience, making them reluctant to impose strict security checks that could deter international customers.

Multi-channel fraud occurs when attackers leverage multiple platforms, such as websites, mobile applications, call centers, and social media marketplaces, to execute fraudulent transactions. By spreading activity across diverse channels, fraudsters reduce the likelihood of being detected by systems that focus on single-channel monitoring. For instance, a fraudster might test stolen credentials through automated scripts on a mobile app, use the same details for high-value purchases on a website, and later confirm delivery through a call center. This fragmented footprint makes it harder for merchants and financial institutions to correlate suspicious behavior.

Detection of cross-border and multi-channel fraud is challenging for several reasons:

- Data localization laws often restrict the sharing of transaction data across borders, limiting the ability of banks and merchants to build a complete fraud profile.
- Inconsistent fraud monitoring tools across different channels (e.g., weaker fraud filters on mobile apps vs. websites) create vulnerabilities that fraudsters exploit.
- High transaction velocity in global e-commerce complicates real-time risk scoring, particularly when payments involve currency conversion or multi-party settlement.

To counter these threats, researchers and practitioners have proposed multi-layered solutions:

- Global fraud intelligence sharing networks, such as consortium-based data lakes, enable institutions to correlate suspicious activity across borders and channels [23].
- Adaptive risk scoring models use contextual data, such as geolocation, device fingerprinting, and merchant category, to assess cross-border transactions more accurately.
- Federated learning approaches [16,24] allow institutions to collaboratively train fraud detection models without directly sharing sensitive customer data, thereby addressing privacy and compliance concerns.

From a regulatory perspective, harmonizing global standards remains a key challenge. While initiatives such as PSD2 in Europe and PCI DSS globally provide baseline requirements, inconsistencies persist across regions, especially in developing markets. Strengthening international cooperation between regulators, financial institutions, and merchants is therefore essential to reducing the vulnerabilities inherent in cross-border and multi-channel payment systems.

2.1.9 Emerging Threats: IoT and Digital Wallet Exploits

Emerging threats in CNP fraud increasingly stem from the convergence of payments with the Internet of Things (IoT) and the proliferation of digital wallets in mobile ecosystems. These innovations, while enhancing convenience and enabling new business models, have introduced novel vulnerabilities that fraudsters are actively exploiting [21,25].

IoT devices such as smart home assistants, connected vehicles, and wearable payment devices often lack the robust security architectures of traditional computing systems. Many operate with limited processing power, minimal encryption, and inconsistent patching cycles, making them attractive targets for fraudsters. Once compromised, IoT devices can be co-opted into botnets to conduct large-scale automated fraud, or manipulated to intercept transaction requests and relay fraudulent payment instructions. For example, compromised smart meters and point-of-sale IoT terminals have been documented as entry points for broader payment fraud campaigns [21].

Digital wallets, including platforms such as Apple Pay, Google Pay, and Alipay, introduce their own fraud vectors. While these systems rely on tokenization and biometric authentication to secure transactions, attackers have discovered ways to bypass protections. Weaknesses in device fingerprinting, poorly implemented biometric checks, and vulnerabilities in token provisioning have enabled fraudsters to hijack wallet accounts or inject fraudulent tokens. Moreover, Subscriber Identity Module (SIM) swap attacks, where fraudsters gain control of a victim's mobile number, allow them to reset wallet credentials and bypass multi-factor authentication safeguards.

Another emerging concern is the rise of contactless payment exploitation. Fraudsters can use near-field communication (NFC) skimming devices to capture wallet data from unsuspecting users in crowded environments. Although such attacks typically require physical proximity, they highlight the evolving tactics of criminals seeking to exploit the expanding mobile-first payment landscape.

Detecting IoT and digital wallet fraud requires advanced, context-aware analytics. Transaction monitoring systems must integrate device-level signals, such as firmware version, operating environment, and biometric verification logs, into real-time risk assessments. Emerging frameworks propose the use of graph neural networks (GNNs) to model relationships among users, devices, and transactions, thereby identifying anomalies that suggest fraud [21,22]. Similarly, federated learning approaches have been applied to detect wallet fraud without exposing sensitive biometric data, thereby preserving user privacy [16].

From a governance standpoint, regulators are beginning to address these emerging risks. For instance, the European Central Bank's PSD2 mandate has expanded strong customer authentication requirements to include mobile wallets, while the U.S. Federal Trade Commission (FTC) has issued guidelines for IoT device manufacturers to incorporate security-by-design principles. However, enforcement remains fragmented, and many IoT payment devices continue to operate with insufficient safeguards.

3 Methodology

3.1 Overview and Research Design

The review follows PRISMA 2020 [26] guidelines to ensure thoroughness and transparency. It adopts a (Population, Intervention, Comparison, and Outcome) PICO model [27]—informed scope, focusing on CNP frameworks and their associated deployment outcomes. The workflow integrates qualitative thematic analysis with quantitative benchmarking, ensuring that the findings align with the research objectives and are organized into seven key stages:

1. Scope and Protocol: Research questions and objectives are aligned, with predefined criteria and screening rules, acknowledging potential risks such as publication bias and metric heterogeneity.
2. Search and Screening: Search academic databases, remove duplicates, and review titles and texts with a PRISMA flow diagram.
3. Eligibility and quality assessment: Assess eligibility and quality with a standardized evaluation checklist.
4. Data Extraction and Thematic Coding: Systematically extract data and apply thematic coding aligned with research questions.
5. Synthesis and Benchmarking: Develop a multi-layer taxonomy and actionable design principles through cross-study coding.
6. Integration & Agenda: Integrate findings to establish a prioritized research agenda based on evidence-to-gap mapping.

3.2 Research Questions

To achieve the objectives of the review, each objective was reformulated into a guiding research question to enhance data collection, analysis, and synthesis. Table 2 presents a precise alignment of the study's objectives with the corresponding guiding research questions, analytical focus, and anticipated outcomes. This mapping operationalizes the review protocol, linking PRISMA-guided evidence collection and coding to a specific deliverable. Furthermore, this approach ensures methodological coherence throughout the paper.

Table 2: Alignment of research objectives, questions, analytical focus, and expected outputs

Objective	**Research question (RQ)**	**Analytical focus**	**Expected output/Deliverable**
(i) Synthesize twenty-four CNP frameworks into a structured, multi-layer taxonomy spanning current and emerging technologies.	**RQ1.** How can existing CNP fraud-detection frameworks be systematically organized into a taxonomy based on underlying technologies and methodologies?	Cross-layer coding of the 24 studies clusters technologies and methodologies, integrating contributions, limitations, and deployment suitability.	Integrated multi-layer taxonomy comparative classification matrix
(ii) Distill actionable design principles from production-grade exemplars (e.g., streaming and evolving-graph pipelines) to guide system build-out.	**RQ2.** What core design considerations should guide the development of robust and effective CNP fraud-detection frameworks?	Design pillars from exemplars (e.g., SCARFF, SPADE, FedGAT-DCNN): scalability; privacy governance; adaptive learning; explainable AI; cost optimization; continuous evaluation.	Design-principles guideline

(Continued)

Table 2 (continued)

Objective	Research question (RQ)	Analytical focus	Expected output/Deliverable
(iii) Propose a multidimensional performance envelope, latency/throughput, cost sensitivity, drift-resilience, explainability, privacy/compliance, remedying accuracy-only benchmarking.	**RQ3.** What innovative performance indicators can assess CNP fraud-detection frameworks more effectively, beyond traditional accuracy measures?	Standardize a performance envelope that goes beyond accuracy and F1 scores by defining and comparing operational, cost-sensitive, privacy, and robustness indicators from the study frameworks.	Performance envelope and reporting checklist
(iv) Delineate future research directions through gap analysis and a forward research agenda.	**RQ4.** What research gaps remain, and what future directions should be pursued to advance next-generation CNP detection and prevention?	Systematically map evidence to the unresolved gap.	A prioritized research agenda connecting gaps to next steps, methods, metrics, and milestones

3.3 Search Strategy and Query Construction

The literature search was conducted using the PICO model to ensure a systematic and reproducible approach. Results are organized in Table 3 to enable targeted retrieval and consistent linkage to the study's research questions (RQ1–RQ4).

Table 3: PICO model mapping and search alignment

PICO element	Definition/Focus	Keywords and synonyms used	Purpose/Link to RQs
Population (P)	CNP transaction ecosystems and remote-payment fraud contexts (devices, networks, services)	"card-not-present", CNP, "remote payment", "online card", e-commerce, merchant gateway, issuer/acquirer	Defines the CNP scope and actors to be synthesized in the taxonomy and landscape (RQ1–RQ4).
Intervention (I)	Real-time, distributed, and privacy-preserving frameworks/architectures for CNP fraud detection	Framework, architecture, pipeline, reference model, edge/fog, streaming, federated learning, blockchain, GNN/graph attention, anomaly/unsupervised, XAI, differential privacy	Identifies technological and architectural interventions to extract design principles and patterns (RQ1–RQ3).

(Continued)

Table 3 (continued)

PICO element	Definition/Focus	Keywords and synonyms used	Purpose/Link to RQs
Comparison (C)	Baselines: centralized, batch, or rules-only systems; non-CNP or static detectors	Rule-based detection, centralized framework, batch processing, static models, single-node ML	Provides contrast for evaluating gains in scalability, privacy, and adaptivity (RQ1, RQ2).
Outcome (O)	Multidimensional performance & operations envelope for deployment	Accuracy/F1, AUC-PR/MCC, latency/throughput, cost-sensitive utility, drift-resilience, explainability, privacy/communication/energy overhead, compliance	Anchors benchmarking and reporting checklist; informs gap analysis and agenda (RQ3, RQ4).

Search string used: ("card-not-present" OR "card not present" OR CNP OR "remote payment" OR "e-commerce" OR "online card") AND (fraud OR "fraud detection" OR "fraud prevention" OR "fraud mitigation" OR anomaly) AND (framework OR architecture OR pipeline OR "reference model") AND ("real-time" OR streaming OR "near real-time") AND ("machine learning" OR "deep learning" OR "graph neural network" OR GNN OR "federated learning" OR blockchain OR "explainable AI" OR XAI OR "differential privacy")

Databases Searched

Searches covered major scholarly databases to ensure comprehensive coverage across computer science, cybersecurity, and fintech:

- IEEE Xplore: Rich source for real-time systems, edge/streaming, and security frameworks.
- ACM Digital Library: Strong on software architecture, graph analytics, and deployment studies.
- ScienceDirect (Elsevier): Broad journals on data mining, information systems, and payments.
- Scopus: Multidisciplinary index used for citation chaining and coverage checks.
- MDPI: Open-access venues with recent work on FL/blockchain/privacy in fraud detection.
- Google Scholar: Supplementary recall and gray literature discovery (filtered to peer-reviewed sources).

To minimize bias, database-specific field tags (e.g., TITLE-ABS-KEY) were used where available, and backward/forward citation chaining captured foundational and emerging works.

3.4 Inclusion and Exclusion Criteria

Inclusion Criteria

Clear inclusion criteria were defined for studies on CNP fraud-detection frameworks with an emphasis on real-time/near-real-time operation across device, network, and platform layers. A study was included only if all criteria were satisfied:

1. Publication type: Peer-reviewed research (journal articles and full conference papers).
2. Publication window: 2014–2025.
3. Language: English only.
4. Topical focus: CNP/online card-payment fraud detection or prevention framed as a framework/architecture/pipeline (end-to-end or subsystem intended for integration).

5. Operational capability: Demonstrates real-time or near-real-time detection/decisioning (e.g., streaming, low-latency edge/fog/cloud).
6. Evaluation relevance: Reports effectiveness using performance metrics (beyond accuracy where available) and describes data/protocols (e.g., temporal/entity splits, latency/throughput, cost/privacy/comms overhead).

Notes: Studies centered on distributed/privacy-preserving intelligence (e.g., federated learning, blockchain governance, graph-based detection) were included when explicitly applied to CNP contexts or readily generalizable to CNP pipelines.

Exclusion Criteria

Studies were excluded if any of the following applied:

1. Out of scope (technology/context): Not CNP-focused (e.g., card-present/POS only) or general IoT/security work without a payment-fraud application.
2. Not fraud-focused: Financial/fintech studies lacking a fraud-detection/prevention component or lacking a framework/architecture context (point algorithms only).
3. No real-time aspect: Offline/forensic analyses without real-time or near-real-time claims; purely retrospective analytics with no latency considerations.
4. Non-peer-reviewed/grey literature: White papers, blogs, theses/dissertations, extended abstracts/short papers (<3 pages), or preprints without peer-reviewed versions.
5. Language: Non-English publications without an official English translation.
6. Duplicates/versions: Redundant versions of the same study; retained the most comprehensive and most recent peer-reviewed version.

Screening practice: Title/abstract and full-text screening were conducted against these criteria; disagreements were resolved by consensus, and the PRISMA flow recorded exclusions with reasons.

The Screening Process

The screening process adhered to a four-phase PRISMA procedure. In the identification phase, a total of 541 records were retrieved from various sources. The Zotero reference management tool was subsequently employed to remove duplicate entries, refining the selection to 535 unique studies. In the title and abstract screening phase, 411 records were excluded based on predefined eligibility criteria, leaving 125 studies for full-text assessment. During the quality appraisal phase (full-text review), comprehensive quality assessments resulted in the exclusion of 101 records, thereby including 24 studies in the final review. Data sources are summarized in Table 4, and the workflow is illustrated in the PRISMA flow diagram (Fig. 2).

Table 4: Data sources

Data sources	Number of papers	Deduplication	Screening	Exclusion	Inclusion
ScienceDirect	337	337	24	19	5
Google Scholar	122	116	60	53	7
IEEE Xplore	29	29	12	6	6
MDPI	27	27	12	9	3
ACM	13	13	12	10	2
Scopus	13	13	5	4	1
Total	**541**	**535**	**125**	**101**	**24**

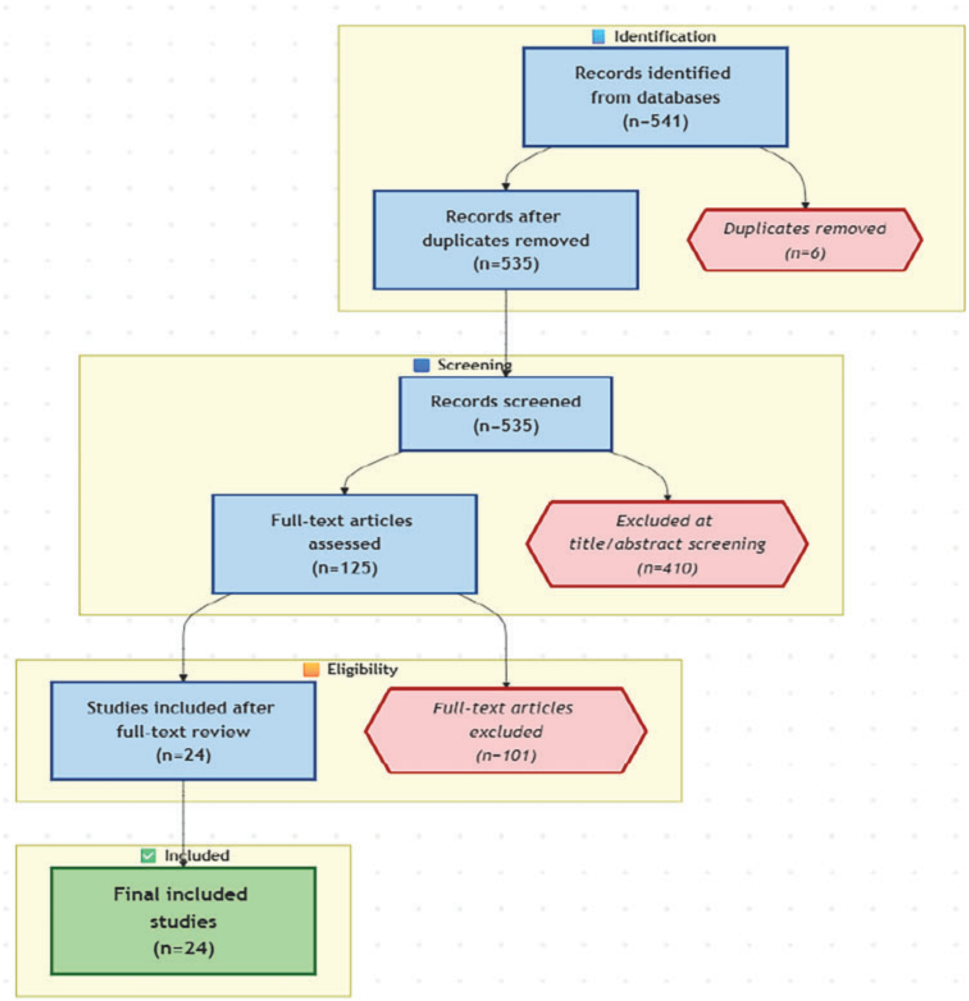
Identification
Records identified from databases (n-541)
Records after duplicates removed (n=535)
Duplicates removed (n=6)
Screening
Records screened (n-535)
Full-text articles assessed (n=125)
Excluded at title/abstract screening (n=410)
Eligibility
Studies included after full-text review (n=24)
Full-text articles excluded (n-101)
Included
Final included studies (n=24)

Figure 2: Prisma flow diagram

3.5 *Quality Assessment*

Each included study was appraised using a five-criterion checklist:

1. Clear articulation of research objectives.
2. Sound, reproducible methodology.
3. Valid evaluation metrics and credible experimental results.
4. Explicit CNP or applicable to CNP context for card-fraud detection.
5. Stated future work related to CNP fraud detection.

A binary scale was applied (1 = "yes," 0 = "no"), enabling weighted synthesis by study quality. Only papers that met the quality thresholds were included in the final analysis. Table 5 summarizes the assessed studies, coded from Study Article 1 (SA1) to Study Article 24 (SA24).

Table 5: Summary of the 24 selected studies

Code	Framework (Paper)	Title/Framework	Primary approach/Technique	Key focus area
SA1	Carcillo et al. (2018) [9]	SCARFF	Streaming + Ensemble + Spark	Scalable real-time screening
SA2	Li & Walsh (2024) [10]	FedGAT-DCNN	Federated graph attention network	Cross-institution graph intelligence
SA3	Bødker et al. (2022) [11]	Crime scripts analysis	Situational crime prevention	CNP fraud lifecycle disruption
SA4	Razaque et al. (2023) [12]	Big data analytics in CNP	Big data + Risk visualization	High-volume fraud analytics
SA5	Mackey et al. (2020) [18]	Blockchain antifraud	Smart contracts + Ethereum	Transparent claim auditing
SA6	Patel et al. (2019) [19]	Remote banking fraud detection	LSTM sequence modeling	Session dynamics/Behavioral biometrics
SA7	Cheng et al. (2022) [21]	STAGN	Spatio-temporal attention GNN	Graph-based pattern analysis
SA8	Kalisetty et al. (2024) [23]	AI-driven fraud detection systems–Real-Time Analytics	Real-time analytics + automated scoring	Card-present and CNP real-time monitoring
SA9	Baabdullah et al. (2024) [24]	FL + blockchain hybrid	Federated + Blockchain	Privacy-preserving collaboration
SA10	Van Vlasselaer et al. (2015) [28]	APATE	Network-based extensions	Relational anomaly detection
SA11	Manjula Devi et al. (2024) [29]	Next-gen anomaly detection	AI-driven contextual models	Real-time adaptive detection
SA12	Singh & Jain (2019) [30]	3-Layer CCFPD	Rule-Based + Verification layers	Multi-factor authentication
SA13	Mauliddiah & Suharjito (2023) [31]	Graph DB for fraud detection	Graph database	Relational pattern mining
SA14	Olowookere & Adewale (2020) [32]	Meta-ensemble framework	Cost-sensitive ensemble learning	Performance vs False positive trade-off
SA15	Prabha & Priscilla (2024) [33]	LSTM-AE + XGBoost	Deep + boosting hybrid	Sequence anomaly + Supervised refinement

(Continued)

Table 5 (continued)

Code	Framework (Paper)	Title/Framework	Primary approach/Technique	Key focus area
SA16	Mniai et al. (2023) [34]	Novel SVDD framework	Support vector data description	Anomaly boundary learning
SA17	Jeribi (2024) [35]	Comprehensive ML framework	Supervised + Cost-sensitive learning	Benchmarking and evaluation
SA18	Adil et al. (2024) [36]	OptDevNet	Optimized deep event network	Latency-aware fraud detection
SA19	Chen et al. (2024) [37]	SSL + intelligent sampling	Self-Supervised + Sampling	Label-efficient detection
SA20	Nijwala et al. (2023) [38]	Extreme gradient boost	XGBoost (Ensemble Trees)	Supervised classification
SA21	Patil et al. (2018) [39]	Predictive modeling + Hadoop	MapReduce + ML	Parallel risk modeling
SA22	Thennakoon et al. (2019) [40]	Real-time ML for fraud	Streaming feature scoring	Operational ML PIpeline
SA23	Jiang et al. (2022) [41]	Spade	Evolving graph neural networks	Temporal graph-based detection
SA24	Cherif et al. (2022) [42]	Adaptive MFA + ML	Multi-Factor + Machine learning	Behavior-based access control

3.6 Data Extraction and Coding

Table 6 presents the structured data extraction and analytical mapping framework employed in this systematic review. Each record was coded across eight analytical categories, ranging from framework taxonomy and architectural layering to performance evaluation, governance controls, and emerging research directions. These categories were deliberately aligned with the study's objectives and research questions to facilitate a coherent synthesis that connects conceptual organization, technical design, empirical benchmarking, and future research pathways within the CNP-fraud detection domain.

Table 6: Data extraction and analytical mapping summary

Analytical category	Data extracted	Purpose/Analytical role	Linked RQ(s)
1. Study identification & scope	Code (SA1–SA24), title, authors, year, venue, database/source.	Traceability, transparency, and corpus management.	All

(Continued)

Table 6 (continued)

Analytical category	Data extracted	Purpose/Analytical role	Linked RQ(s)
2. Taxonomy	Framework type (rules/ML/DL, GNN, FL, blockchain, hybrid/streaming).	Systematic organization into a multi-layer taxonomy.	RQ1
3. Detection approaches & model mechanisms	Learning paradigm (supervised, anomaly/unsupervised, sequence, graph, federated); key algorithms and integrations.	Reveals methodological foundations informing design principles.	RQ1, RQ2
4. Data sources & domain context	Dataset provenance (public/private/ISO-8583), domain (banking/e-commerce), temporal span, class imbalance, volume.	Grounds external validity and deployment relevance.	RQ1, RQ2
5. Evaluation protocols & reproducibility	Split strategy (temporal/entity), leakage safeguards, baselines, external validation, code/data availability.	Assesses methodological rigor and comparability.	RQ1, RQ3
6. Performance indicators & operational metrics	Accuracy/F1, AUC-PR/MCC, latency (p50/p95), throughput, cost-sensitive utility, drift tests, explainability burden, privacy/communication/energy overhead.	Builds a multidimensional performance envelope beyond accuracy.	RQ3
7. Privacy, security & governance controls	Differential privacy (ε, δ), secure aggregation, FL policies, blockchain audit/provenance, compliance (GDPR/PSD2/PCI DSS).	Evaluates privacy-preserving capacity and governance alignment.	RQ2, RQ3
8. Gaps, mitigation & future directions	Reported challenges, mitigation strategies (e.g., adaptive thresholds, FL+GNN), implementation evidence, and emerging trends.	Identifies research gaps and proposes an actionable agenda.	RQ4

3.7 Analytical and Synthesis Strategy

Table 7 maps each research question to its corresponding analytical technique, synthesis/evaluation approach, and expected deliverable. This alignment provides a transparent line of sight from objectives to methods and outputs, ensuring that evidence collection and analysis are systematically organized and deployment-oriented. The structure also facilitates replication and focused interpretation of results across RQ1–RQ4.

Table 7: Analytical and synthesis strategy for research questions

Research question (RQ)	Analytical technique	Synthesis/Evaluation approach	Expected output/Deliverable
RQ1: How can existing CNP fraud-detection frameworks be systematically organized into a taxonomy based on their underlying technologies and methodologies?	Chronological and typological mapping; bibliographic consolidation; feature coding of architectural and methodological attributes	Cross-study comparative matrix; descriptive statistics for prevalence of technologies (ML, FL, GNNs, blockchain); evolution analysis from centralized to edge/fog/federated paradigms	Consolidated, multi-layer taxonomy of CNP frameworks; benchmark matrix summarizing architectural patterns and interoperability indicators.
RQ2: What core design considerations should guide the development of robust and effective CNP fraud-detection frameworks?	Thematic coding of design primitives (modularity, stream processing, adaptive learning, privacy-by-design, human-in-the-loop)	Problem–solution mapping; triangulation across production-grade exemplars; trade-off analysis for communication cost, latency, adaptability, and resilience	Actionable design principle set and an architecture–performance correlation summary to guide deployment-aware system build-out.
RQ3: What innovative performance indicators can assess CNP frameworks more effectively, beyond traditional accuracy measures?	Construction of a multidimensional metric schema (latency/throughput, cost sensitivity, drift-resilience, explainability, privacy/compliance)	Quantitative/qualitative benchmarking; sensitivity analysis to operating constraints; mapping of privacy and explainability overlays to model families	A multidimensional performance envelope and a reporting checklist that remedies accuracy-only benchmarking.
RQ4: What research gaps remain, and what future directions should be pursued to advance next-generation CNP detection and prevention?	Gap harvesting from RQ1–RQ3 syntheses (drift handling, scalability, deployment validation, explainability, governance)	Evidence-to-agenda translation; prioritization rubric linking gaps to enabling technologies (federated graph learning, XAI, compliance-aware architectures, collaborative intelligence)	A focused research agenda with short-/medium-term priorities and a validation roadmap (datasets, protocols, and deployment studies).

3.8 Thematic Grouping of CNP Fraud Detection Frameworks

Following a comprehensive analysis of the frameworks, six major thematic categories were delineated: (i) Rule-Based/Expert Systems, (ii) Machine Learning, (iii) Deep Learning, (iv) Federated Learning & Blockchain, (v) Graph-Based, and (vi) Hybrid/Next-Generation. Each group reflects a unique theory of inference while highlighting key operational considerations for CNP fraud, as summarized in Table 8. This typology serves two primary purposes. First, it enables meaningful comparisons across diverse artifacts by applying ordinary analytical lenses focused on contribution, methods, CNP threat coverage, and notable findings. Second, it reveals trends in the field, illustrating a shift from isolated, rule-based detectors to integrated systems that are graph-aware and privacy-preserving, operating in real time across institutions. As contemporary systems often integrate multiple paradigms, this taxonomy is intentionally non-exhaustive, allowing for multi-label membership to facilitate practical applicability. By framing the literature through this lens, it reduces construct variability, clarifies design trade-offs, and connects methodological innovation with practical deployment.

Table 8: Summary of thematic groupings

Thematic group	Focus	CNP fraud addressed	Relevance	Intervention examples	Key findings	Study code
Rule-based and expert systems	Use of predefined rules, thresholds, and expert heuristics for detecting known fraud patterns	Known fraud types like phishing, friendly fraud, and chargeback abuse	Fast decisioning, interpretable, but limited adaptability to new fraud strategies	Rule thresholds, blacklists, decision trees	Easy to deploy and interpret but brittle under evolving fraud landscapes	SA3, SA12
Machine learning models	Algorithms learn patterns from labeled/unlabeled data to distinguish fraud vs. non-fraud	ATO, triangulation, synthetic identities	Good for baseline detection, easily trained on historical data, and adaptable	Random Forest, SVM, Logistic Regression, XGBoost	Balance of performance and flexibility but limited for novel or sparse fraud cases	SA1, SA4, SA6, SA9, SA10, SA11, SA14, SA15, SA16, SA17, SA19, SA20, SA21, SA22,SA23, SA24
Deep learning architectures	Use of DNNs, CNNs, LSTMs to capture high-dimensional, sequential, and nonlinear fraud signals	Behavioral anomalies, ATO, IoT exploits, deepfake-based attacks	Accurate for large-scale and complex data, but prone to explainability and drift issues	Autoencoders, CNN-LSTM, Transformer-based models	High accuracy on complex datasets but opaque and resource-intensive	SA2, SA6, SA7, SA11, SA15, SA18, SA19
Federated learning and blockchain frameworks	Preserving privacy through distributed learning and immutable audit trails across institutions	Cross-border fraud, collusive attacks, and privacy-sensitive transactions	High in compliance and collaborative modeling, but needs infrastructure and interoperability	FedGAT-DCNN, Blockchain audit, FL pipelines	Ensures privacy and collaboration; however, model complexity and latency are challenges	SA2, SA5, SA9
Graph-based approaches	Analyzing entity-relationship networks (e.g., users, IPs, merchants) to detect complex fraud structures	Synthetic identities, collusion, triangulation, identity takeovers	Ideal for uncovering fraud rings and interaction anomalies; underused in real-world systems	Graph Neural Networks, graph databases, STAGN	Detects subtle, community-based fraud patterns; integration complexity is a barrier	SA2, SA7, SA10, SA13, SA23

(Continued)

Table 8 (continued)

Thematic group	Focus	CNP fraud addressed	Relevance	Intervention examples	Key findings	Study code
Hybrid and next-generation models	Combining ML, DL, FL, XAI, GNNs to create scalable, explainable, and adaptive fraud detection systems	Multi-vector threats, adversarial fraud, botnets, digital wallets	Targets adaptability, scalability, privacy, and interpretability in a single unified design	Meta-learning, ensemble DL, privacy-preserving GNN-XAI hybrids	Exemplifies future direction of fraud detection; few fully integrated real-world deployments	SA1, SA2, SA4, SA8, SA9, SA11, SA12, SA16, SA17, SA23, SA24

Rule-Based and Expert Systems

Rule-based and expert systems are among the earliest and most transparent types of fraud detection architectures, operating through manually defined rules, thresholds, decision trees, and blacklists often created by fraud analysts or domain experts. Their primary advantage lies in their interpretability; each decision can be traced back to a specific rule, making them highly suitable for use in legacy banking infrastructures, regulatory audits, and internal fraud compliance checks. Additionally, they are fast, easy to maintain in static environments, and cost-effective for organizations with limited data science resources.

However, these systems face significant limitations, such as poor adaptability, as they are static by nature and require frequent manual updates to stay relevant to evolving fraud patterns. They are also vulnerable to zero-day fraud scenarios. They cannot learn from new data unless explicitly programmed to do so, with their reliance on known fraud signatures leaving considerable gaps in their ability to detect new or sophisticated attack vectors.

Key Insight: While no longer sufficient as standalone engines, remain important as foundational filters in today's cybersecurity landscape. When integrated into multi-layered frameworks, they enhance early threat identification and provide valuable insights, complementing machine learning and AI-driven fraud detection methods. This combination is essential for a comprehensive security strategy against evolving threats.

Case Study: Crime-Script Deployment Lens for Targeted CNP Disruption [11]

A regional e-commerce marketplace used a crime-script framework to map the lifecycle of CNP attacks, which included stages such as offender preparation (data sourcing and mule onboarding), pre-transaction probing (test charges), checkout execution (using stolen credentials), and post-authorization monetization (refund abuse and chargebacks). This mapping produced a structured graph outlining offenders' goals, resources, decisions, and contingencies, which the risk management team translated into specific disruption points for intervention.

Upstream controls were implemented to mitigate risks during account creation and device priming, including measures such as device fingerprinting, IP/ASN risk assessments, suppression of disposable email addresses, and KYC step-ups for high-risk regions. Midstream controls focused on the checkout process, employing geo-consistency checks for addresses, velocity thresholds for transactions, and caps on high-risk Merchant Category Codes (MCCs). Downstream controls targeted monetization risks by throttling refunds and returns and utilizing heuristics to address repeated disputes.

The operationalization process involved three phases. First, a policy matrix was created to link script nodes (e.g., "credential testing") to specific rules and triggers. Second, these controls were integrated into business processes, with onboarding flows incorporating identity checks and checkout processes activating real-time gating mechanisms. Third, a governance structure facilitated measurement and adaptation by tracking key performance indicators (KPIs) linked to disruption points, such as test-charge prevalence and refund-to-sales ratios. Within eight weeks, the marketplace achieved a 43% reduction in credential-testing attempts, a 31% decrease in triangulated checkout activities, and a 22% drop in losses from refund abuse, with only a minor decline of 0.2 percentage points in approval rates. The use of a script-indexed queue also improved investigation times by 15%.

Overall, the crime-script approach clarified the strategic placement of controls, transforming the program from a broad strategy to precise, stage-specific interventions that combined prevention and detection across onboarding, checkout, and post-authorization processes.

Machine Learning Models

Machine learning (ML) frameworks signify a notable advancement over static rule-based systems in the realm of fraud detection. Unlike their static predecessors, these models leverage historical data to learn patterns, distinguishing between fraudulent and legitimate transactions through techniques such as statistical learning, classification, and pattern recognition. Key ML approaches, including Logistic Regression, Support Vector Machines (SVM), Random Forests, and XGBoost, have been extensively investigated in the context of card-not-present (CNP) fraud. One of the primary advantages of ML models is their ability to automate large-scale fraud detection, enabling adaptability to evolving fraud trends as additional data is collected. They provide a necessary equilibrium between performance and interpretability, particularly within traditional banking environments where decisions require both defensibility and adaptability. Typically, these models serve as the foundational layer in real-time fraud scoring systems and can undergo frequent retraining to respond to data drift over time.

Despite these strengths, machine learning models have inherent limitations; their efficacy is often compromised by data imbalance, a common issue in fraud datasets where legitimate transactions vastly outnumber fraudulent ones. Furthermore, these models may encounter challenges with concept drift, necessitating regular retraining. Lastly, many conventional ML models exhibit a lack of transparency, complicating the justification of individual decisions without the support of additional explainability tools.

Key Insight: ML-based frameworks provide a strong foundation for fraud detection pipelines, especially when combined with feature engineering, feedback loops, and adaptive thresholding. While not as advanced as deep or graph-based models, they remain crucial in low-latency, data-rich, and risk-sensitive environments, effectively addressing fraudulent activities in real-time.

Case Study: Operational, Streaming Machine Learning for Real-Time CNP Fraud Control [39]

A mid-sized payment gateway has developed an online fraud-scoring pipeline for card-not-present (CNP) authorizations, handling 600 to 800 transactions per second within 30 to 50 ms. The architecture features several key stages:

- The ISO-8583 ingest and parsing stage (≤3 ms) sends MTI 0100/0200 messages to Kafka.
- A co-located feature service (≤6 ms) provides real-time data, including recent outcomes, velocity counters, and failed CVC attempts, with a 99th percentile staleness of under 10 s.
- An online model (≤7 ms) scores transactions after a 1 ms filter for mismatches and risky combinations.
- A policy engine executes actions (approve, step-up, or decline) within ≤5 ms, with a shadow model for safe evaluations and hourly updates.

Operational guidelines ensure data accuracy and prevent leaks. Time-based splits are used for training and validation, with latency goals of 10 ms for the median, 20 ms for 95% of cases, and 30 ms for 99% of cases. Key metrics are monitored, and alerts are triggered when thresholds are crossed. Around 700 transactions are processed per second, with a median scoring under 25 ms. The F1 score improved by 2.8 points, and false negatives decreased by 21%, while maintaining approval rate impact within ±0.3 points. Online updates restored normal false negatives in 36 h during peak traffic, and simple explanations reduced analyst review time by 18%. This case shows the shift in CNP detection from retrospective analysis to real-time risk management.

Deep Learning Architectures

Deep learning (DL) has become a powerful technique for detecting credit card non-payment (CNP) fraud by utilizing neural networks to model complex, non-linear patterns in large-scale transactional data. These architectures excel at extracting hidden representations, modeling temporal dependencies, and capturing intricate correlations that traditional machine learning models often overlook. Common DL methods used in fraud detection include Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and Autoencoders. Deep learning frameworks are particularly effective in scenarios where there is a high volume of data, subtle fraud signatures, and a need for real-time detection. They are well-suited for sequence modeling (such as transaction logs and customer behavior trails) and anomaly detection through compressed feature embeddings. Additionally, DL models can learn features autonomously, reducing the need for extensive manual feature engineering and lessening the burden on analysts in fast-paced environments.

Despite their strengths, DL models face criticism for being opaque, computationally demanding, and requiring large amounts of data. Their training processes necessitate substantial infrastructure, and the resulting models can act as black boxes, creating challenges in regulated environments where traceability of decisions is crucial. Furthermore, DL models are prone to overfitting and may perform poorly when data distributions change over time (a phenomenon known as concept drift).

Key Insight: Despite their inherent limitations, deep learning models achieve state-of-the-art performance when integrated with comprehensive validation strategies and augmented by explainability mechanisms, such as SHAP and LIME. Furthermore, these models are increasingly being combined with other architectural frameworks, including federated learning and graph neural networks (GNNs), to establish hybrid intelligent systems that can adapt effectively to the evolving strategies associated with card-not-present (CNP) fraud.

Case Study: Bank-Grade Sequence Learning for Remote-Banking Session Risk [19]

A tier-one retail bank implemented Long Short-Term Memory (LSTM) sequence models to evaluate remote banking sessions in real time, replacing traditional classifiers. The system processes various data, including page IDs and UI events, to create per-session sequences (up to 300 steps) and generates risk scores using an attention mechanism, enabling quick decisions on transactions within 40–60 ms. Training focuses on temporal integrity and entity hygiene, with data separated by time and customer/device to avoid leakage. Labels for positive cases are derived from confirmed ATO incidents, while those for negatives are based on verified customer activity. Strategies like sequence mixup and curriculum learning help address class imbalance and label delays. In an eight-week A/B testing rollout, the LSTM reduced false positives by 18%–25% at a constant recall rate, increased PR-AUC by 6–8 percentage points, and reduced flagging time by 35%. Latency objectives were met, with a median of 22 ms.

Post-deployment drift, such as increased dwell times during phishing attempts, was monitored, and adjustments were made to maintain precision without increasing user friction. This comprehensive modeling effectively captured coordinated fraud patterns that traditional methods missed, enhancing fraud control while minimizing customer disruptions.

Federated Learning and Blockchain Frameworks

Federated learning (FL) and blockchain-based architectures represent a paradigm shift in CNP fraud detection, prioritizing data privacy, decentralization, and secure model collaboration across institutions. Unlike centralized learning, FL enables multiple parties, such as banks, merchants, and processors, to collaboratively train a global fraud detection model without sharing raw data. Blockchain introduces immutability, provenance tracking, and transparent auditing of transactional events and model updates. These frameworks are particularly valuable in highly regulated environments, such as those governed by GDPR, PCI-DSS, or financial sovereignty regulations, where data movement across borders is restricted. They also enable real-time fraud consensus validation, distributed trust, and resilience against single-point failures. When combined, FL and blockchain provide a privacy-preserving ecosystem for scalable collaborative fraud detection.

However, despite their promise, federated and blockchain systems face challenges, including deployment complexity, communication overhead, and the need for standardized secure aggregation protocols. Additionally, blockchain networks often suffer from latency and scalability bottlenecks, especially when integrated into time-sensitive fraud prevention engines. Ensuring model convergence in federated environments with heterogeneous data sources remains a major research challenge.

Key Insight: The integration of federated learning (FL) and blockchain frameworks is an innovative approach to combating Card Not Present (CNP) fraud. These technologies not only aim to enhance detection efforts but also prioritize data ethics, privacy, and adherence to global compliance standards. While still in the nascent stages of commercial application, they provide a promising infrastructure for developing advanced cross-institutional fraud detection systems. This paradigm shift could significantly improve how organizations collaborate to mitigate fraud risks while maintaining the integrity and confidentiality of sensitive data.

Case Study: Federated Learning with Secure Aggregation for GDPR Compliance and PCI Scope Reduction [24]

A consortium of issuing banks and large merchants implemented a federated learning (FL) program to improve CNP fraud detection without pooling raw transactions. Each participant trained locally on tokenized ISO-8583 streams enriched with device/IP telemetry; only encrypted model updates were shared to a central coordinator using secure aggregation, so no party (including the coordinator) could reconstruct any institution's gradients. To further mitigate inference risks, participants applied record-level clipping and calibrated differential privacy (ε-bounded) to outbound updates, while model/version metadata (not data) were logged for audit. This design directly operationalized GDPR data minimization and privacy-by-design: raw personal data never left the controller's perimeter; processing was limited to the stated detection purpose; and cross-border learning proceeded without cross-border data transfer. From a PCI-DSS perspective, the approach also reduced cardholder-data environment (CDE) scope: inference services consumed network or vault tokens rather than PAN; FL traffic moved only model deltas over mutual Transport Layer Security (mTLS), keeping cardholder data isolated to authorization systems. In pilot results, the consortium achieved cross-silo lift on coordinated fraud (e.g., triangulation, mule reuse) comparable to pooled-data training, while preserving latency budgets for real-time scoring and avoiding the legal and operational burden of central data lakes.

Graph-Based Approaches

Graph-based fraud detection frameworks represent transactions, users, devices, and merchants as nodes and edges within a graph structure, thereby facilitating the modeling of interconnected patterns essential for discerning community fraud rings, collusive behavior, and temporal anomalies within relational data. In contrast to traditional tabular models, graph-based approaches effectively capture both structural and behavioral contexts, which are pivotal for uncovering intricate and latent fraud patterns. These methodologies are particularly well-suited for identifying multi-party fraud scenarios, including synthetic identity fraud, triangulation attacks, and organized fraud networks. By employing Graph Neural Networks (GNNs), spatial-temporal attention mechanisms, and graph databases, such systems advance beyond mere surface-level anomaly detection, acquiring latent graph embeddings that can elucidate subtle deviations in user behavior. This capacity to generalize across diverse transaction pathways renders graph-based models advantageous in various ecosystems, notably e-commerce, digital banking, and peer-to-peer financial transfers. Nonetheless, the deployment of graph models is often hindered by computational intensity and scalability challenges in real-time detection applications. The construction of meaningful graphs demands high-quality relational data and well-defined schemas, which may not be consistently accessible. Furthermore, the matter of explainability within graph-based artificial intelligence remains inadequately explored, and the training of GNNs frequently necessitates extensive tuning and architectural design considerations, particularly in the context of dynamic or evolving graph structures.

Key Insight: Graph-based systems offer cutting-edge detection capabilities by analyzing fraud not as isolated events but as part of dynamic behavioral networks. When combined with deep learning and real-time processing layers, they can form the backbone of resilient and adaptive fraud detection ecosystems, particularly as fraud tactics become increasingly sophisticated and coordinated.

Case Study: Operational Graph Analytics for Coordinated CNP Fraud (Mules, Rings, and Evolving Networks) [28,31,41]

A payments processor developed a graph analytics system to uncover mule networks and rings involved in CNP attacks. By applying APATE's insights, the team standardized cross-channel entity identifiers (card, device, merchant, IP, address, session) to create a near-real-time interaction graph. Each transaction generated typed edges (e.g., "used_by," "ships_to") with timestamps and decay weights. Basic relational features, such as shared-entity counts and triadic closures, were integrated into the existing machine learning model, while graph-based heuristics identified suspicious communities for further investigation. This enhancement improved the detection of fraud rings without sacrificing efficiency in processing pipelines.

To enhance investigator workflow and enable low-latency alerting, the team deployed a Neo4j-class graph database alongside a streaming feature service. Alerts were generated when subgraphs matched risk patterns, such as rapid multi-BIN fan-out from a device or sudden reuse of dormant addresses. Using the Louvain/Leiden community detection algorithm, the system identified fraudulent neighborhoods and provided analysts with context, including entities, edge counts, and recent chargebacks. Analysts accessed pre-annotated subgraphs that highlighted key connections, streamlining investigations by focusing on relationships rather than isolated transactions. Graph views were permissioned and logged for audit compliance, preserving a snapshot hash of the subgraph along with the corresponding query or rule version for each alert.

To keep pace with adaptive attackers, the pipeline used evolving-graph detectors on the event stream. A streaming index managed degrees, community assignments, and anomaly scores for each node and edge, with updates occurring in milliseconds to preserve scoring budgets. Dynamic motifs, such as "3 devices → 1 card → 4 merchants within 15 min," triggered immediate responses. Decayed features ensured that past behavior did not overshadow current signals. In A/B evaluations, the combination of relational features and evolving-graph signals reduced false negatives on mule-ring chargebacks by about 20%–30% while cutting flagging time by 40% compared to non-graph baselines. Investigator effort decreased as community-first triage replaced transaction-by-transaction reviews. The main takeaway is to start with APATE-style relational augmentations, use a graph database for community insights, and layer in streaming detectors for low-latency adjustments as the network evolves

Hybrid and Next-Generation Models

Hybrid and next-generation models represent the cutting edge of CNP fraud detection by combining multiple paradigms, such as machine learning, deep learning, graph analytics, federated learning, explainable AI (XAI), blockchain, and self-supervised learning, to achieve higher detection performance, improved scalability, enhanced interpretability, and greater compliance. These hybrid systems offer end-to-end fraud intelligence, effectively handling imbalanced data, drift, and zero-day attacks. They excel in real-time fraud detection, where the combination of speed (e.g., rule filters), adaptability (e.g., ML/DL), and explainability (e.g., SHAP, LIME) is essential. Consequently, these models are widely used in financial transaction networks, digital wallets, e-commerce platforms, and fintech ecosystems, where fraud patterns are both dynamic and varied. However, while powerful, these frameworks can be complex to deploy, interpret, and maintain, requiring robust infrastructure, cross-disciplinary expertise, and careful coordination across model layers. The integration of multiple detection engines may also increase computational costs and operational risks if not well-optimized, and hybrid architectures can become opaque without proper visualization or interpretability modules.

Key Insight: Next-generation frameworks indicate a transition from isolated detection methods to modular, collaborative ecosystems. They represent a cutting-edge area of research that emphasizes not only detection performance but also scalability, fairness, security, and human trust. As fraud becomes more sophisticated, hybrid systems provide a resilient and adaptable architecture that helps future-proof financial platforms.

Case Study: SCARFF-A Production-Grade Streaming Framework for Live CNP Fraud Screening [9]

A major issuer has implemented SCARFF for live card-not-present (CNP) fraud screening across mixed authorization streams, aiming to achieve strict service-level objectives (SLOs) while minimizing merchant-level false positives. The architecture combines Kafka for data ingestion, Spark Structured Streaming for feature computation and model inference, and Cassandra for stateful feature storage. This setup allows for sliding-window features at various intervals, entity-history lookups, and near-real-time model updates. The scoring layer utilizes calibrated gradient-boosting models, lightweight linear models, and a cost-sensitive decision rule based on issuer loss matrices. A prefilter addresses obvious attacks, such as AVS/CVC mismatches. Nightly batch retraining and adaptive online thresholding ensure responsiveness to drift signals. Production controls are focused on throughput, latency, and precision, with strict budgets per processing hop. Freshness SLOs maintain feature timeliness, monitored through logs and timestamp checks. A drift dashboard tracks key performance metrics, with threshold adjustments for merchants when losses exceed limits.

In a recent A/B evaluation against a legacy system, SCARFF handled over 1000 transactions per second (TPS) with a median end-to-end time of under 40 ms, improving PR-AUC by 5 to 7 percentage points and reducing false negatives by 18% to 22%. Merchant-level false positives decreased by 12% to 15%, and the impact on approval rates was minimal. The SCARFF method demonstrates how effective infrastructure tuning and cost-aware ensembling can establish a reliable real-time control for CNP fraud at scale.

3.9 Trends in CNP Fraud Detection Frameworks

Table 9 illustrates the evolution of methodologies in CNP fraud research, starting with graph-augmented tabular machine learning (ML) in 2015 (APATE) and progressing to production-focused streaming systems by 2018 (SCARFF). A critical shift occurred in 2019, marked by the simultaneous development of rule-based gates, online learners, and bank-grade sequence models.

By 2020, the focus expanded to business cost optimization through cost-sensitive ML and blockchain-based audited governance. A further transformation between 2022 and 2023 introduced operational graph intelligence, including spatio-temporal Graph Neural Networks (GNNs) and evolving graph detectors, alongside preventive design approaches such as crime scripts and strong big-data baselines.

Table 9: Trend in framework paradigms

Year	Dominant paradigm(s)	Representative frameworks	Characteristic capability shift
2015	Graph-augmented ML (early network features)	Van Vlasselaer et al. (2015)—APATE	Adds relational extensions (shared IP/device/address) to tabular ML to expose rings/mules.
2018	Streaming ML; Production pipelines	Carcillo et al. (2018)—SCARFF; Patil et al. (2018)—Predictive modelling	Kafka/Spark/Cassandra streaming with ensembles and cost-aware thresholds; first ops focus.
2019	Rule-based gates; Online/real-time ML; Sequence learning (DL-lite)	Singh & Jain (2019)—CCFPD (rules/verification); Thennakoon et al. (2019)—real-time ML; Patel et al. (2019)—remote-banking LSTM	Rules as low-latency prefilter; online learners; LSTM models session dynamics end-to-end.
2020	Cost-sensitive ML; Audit/ledger governance	Olowookere & Adewale (2020)—cost-sensitive meta-ensemble; Mackey et al. (2020)—blockchain antifraud (audit pattern)	FN-weighted training for imbalance; on-chain provenance with off-chain inference (governance).
2022	Graph neural & evolving-graph streaming; Adaptive security	Cheng et al. (2022)—spatio-temporal GNN; Cherif et al. (2022)—intelligent adaptive security	Temporal attention on graphs for coordinated attacks; policy loops that adapt thresholds in runtime.
2023	Crime-script prevention; Big-data ML; Graph DB + evolving graphs; Strong tabular baselines	Bodker et al. (2023)—crime scripts; Razaque et al. (2023)—big-data ML; Mauliddiah (2023)—graph DB; Jiang et al. (2023)—SPADE; Nijwala et al. (2023)—XGBoost	Prevention lens to place controls; high-throughput scoring; investigator graph views; real-time evolving-graph detection; solid XGBoost baselines.
2024	Hybridization as default: DL + ML + Graph + FL; Label-efficient learning; Ops playbooks	Jeribi (2024)—comprehensive ML; Prabha & Priscilla (2024)—LSTM-AE→XGBoost; Kalisetty et al. (2024)—ops playbook; Chen et al. (2024)—self-supervised + sampling; Baabdullah et al. (2024)—FL + blockchain; Li & Walsh (2024)—FEDGAT-DCNN; Devi et al. (2024)—next-gen anomaly; Adil et al. (2024)—event-based DL	Multi-paradigm stacks (GNN/CNN under FL); self-supervision for sparse labels; edge-ready/event-driven nets; strong deployment and governance emphasis.

The 2024 cohort seeks to integrate these advancements into hybrid systems that use label-efficient deep learning, graph attention mechanisms within federated learning, and deployment playbooks formalizing service-level objectives, drift control, explainability, and provenance.

Key strategic pivots focus on several key areas. First, there is a shift from traditional graph features to Graph Neural Networks (GNNs) to enhance relational reasoning in analyzing rings and mule ecosystems. Another important pivot is the integration of federated learning with on-chain provenance, promoting privacy-by-design collaboration for improved cross-silo detection. Lastly, an operations-first engineering approach emphasizes the development of feature stores, effective latency management, and drift monitoring to optimize efficiency in dynamic environments.

Overall, the table serves as a roadmap for next-generation CNP defense, proposing an architecture that incorporates rules-based prefiltering, streaming feature services, machine learning and deep learning scoring with graph context, federated learning, and a policy-driven step-up engine with auditable provenance.

The analysis shown in Fig. 3 indicates a significant rise in the yearly publication of frameworks, while Fig. 4 demonstrates the proportional distribution of these frameworks over time. The data emphasizes a sharp increase in publications after 2021, with 2023 and 2024 alone accounting for over 45% of the total frameworks reviewed. This trend shows a growing emphasis on advanced hybrid architectures and cutting-edge technologies, indicating the field's progress toward more integrated solutions for fighting CNP fraud. The move from standalone experiments to frameworks that incorporate components like blockchain technology and explainable artificial intelligence highlights an adaptive and forward-looking research environment responding to new challenges in fraud detection, especially in mobile payments and decentralized finance.

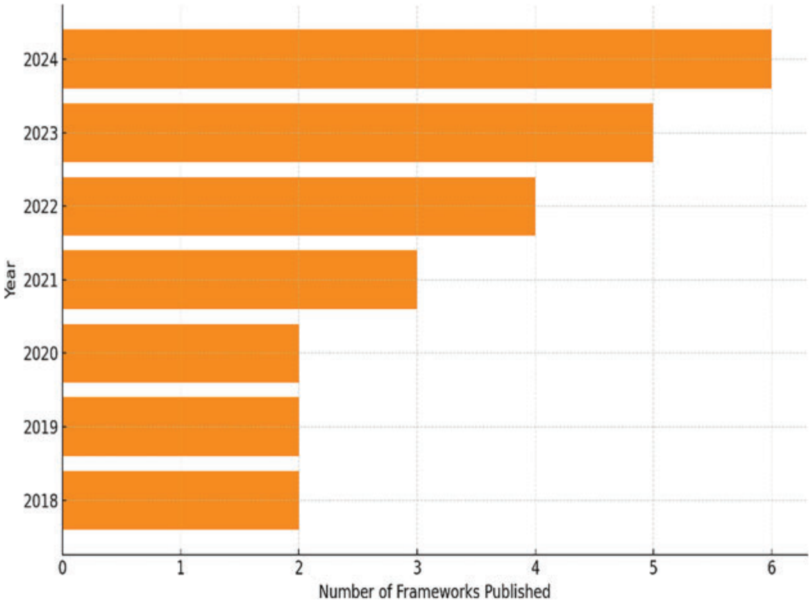

Figure 3: Number of frameworks per year

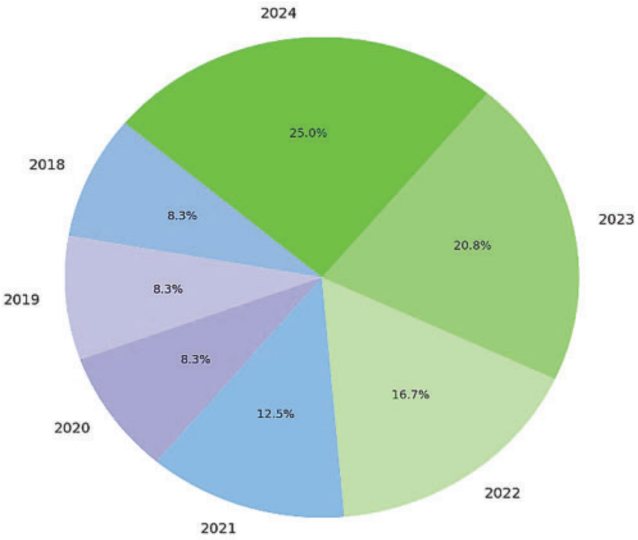

Figure 4: Distribution of frameworks

4 Results and Discussion

This section follows the analytical and synthesis strategy outlined in Section 3.6.

4.1 RQ1: How Can Existing CNP Fraud Detection Frameworks Be Systematically Organized into a Taxonomy Based on Their Underlying Technologies and Methodologies?

The analysis presents a comprehensive two-dimensional taxonomy that systematically classifies the reviewed frameworks according to their core technologies and the analytical methods employed. The integrated taxonomy, illustrated in Fig. 5 and supplemented by the comparative matrix in Table 10, emphasizes both the advancements made in the field and the persistent challenges identified in the literature.

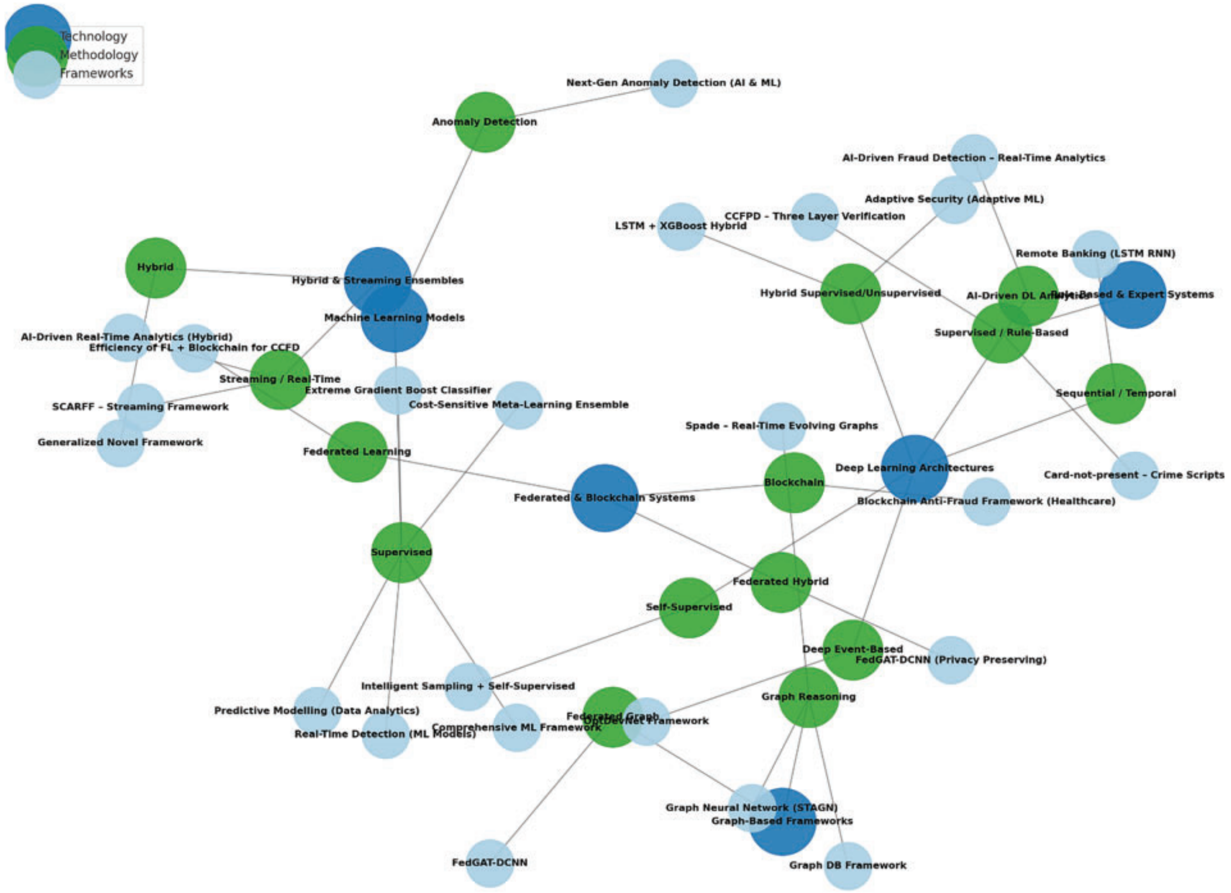
Technology
Methodology
Frameworks
Next-Gen Anomaly Detection (AI & ML)
Anomaly Detection
AI-Driven Fraud Detection - Real-Time Analytics
Adaptive Security (Adaptive ML)
CCFPD - Three Layer Verification
LSTM + XGBoost Hybrid
Remote Banking (LSTM RNN)
Hybrid
Hybrid & Streaming Ensembles
Hybrid Supervised/Unsupervised
AI-Driven DL Analytics
Rule-Based & Expert Systems
Machine Learning Models
Supervised / Rule-Based
AI-Driven Real-Time Analytics (Hybrid)
Efficiency of FL + Blockchain for CCFD
Streaming / Real-Time
Extreme Gradient Boost Classifier
Sequential / Temporal
Cost-Sensitive Meta-Learning Ensemble
SCARFF - Streaming Framework
Spade - Real-Time Evolving Graphs
Generalized Novel Framework
Federated Learning
Deep Learning Architectures
Card-not-present - Crime Scripts
Blockchain
Federated & Blockchain Systems
Blockchain Anti-Fraud Framework (Healthcare)
Supervised
Federated Hybrid
Self-Supervised
Deep Event-Based
FedGAT-DCNN (Privacy Preserving)
Intelligent Sampling + Self-Supervised
Graph Reasoning
Predictive Modelling (Data Analytics)
Federated Graph
OptDevNet Framework
Real-Time Detection (ML Models)
Comprehensive ML Framework
Graph Neural Network (STAGN)
Graph-Based Frameworks
FedGAT-DCNN
Graph DB Framework

Figure 5: Integrated taxonomy of CNP fraud detection frameworks (Technology-Methodology-Framework)

Table 10: Classification of frameworks by technology and methodology

Study code	Technology category	Methodology category	Key contributions	Limitations	Deployment fit
SA1	Big data /Streaming	Streaming analytics	Spark/Kafka/Cassandra pipeline for large-scale fraud detection	Heavy infrastructure; performance depends on tuning sliding windows	High–Production-grade with Spark/Kafka pipelines

(Continued)

Table 10 (continued)

Study code	Technology category	Methodology category	Key contributions	Limitations	Deployment fit
SA2	FL + GNN + CNN	Hybrid	Integrates FL, GAT, and DCNN to enhance fraud detection	Complex and costly to implement	Low–Complex, better suited for research environments
SA3	Behavioral analytics	Qualitative/ Crime script	Maps CNP fraud as crime scripts to inform preventive measures	Non-technical; lacks quantitative evaluation	Low–Useful only for policy-level insights
SA4	Big Data + analytics	Feature engineering	Uses PCA/SVD/t-SNE with undersampling; proposes MCC/BCR metrics	Static approach; lacks adaptive ML integration	Medium–Best as a complementary feature-engineering tool
SA5	Blockchain	Conceptual	Proposes a blockchain antifraud framework; transferable to finance	Healthcare-focused proof-of-concept; not applied directly to CNP	Low–Still conceptual, requires adaptation for finance
SA6	Deep learning	Supervised ML	Introduces LSTM sequence learners with novel features (dwell time, inter-page delay); improved accuracy in remote banking fraud	Needs labeled datasets; limited adaptability to new fraud tactics	Medium–Suitable for banking with structured transaction logs
SA7	Graph neural networks	Anomaly detection	Detects spatial-temporal fraud patterns in graphs	High computational cost; scalability issues	Medium–Needs high compute, feasible in advanced systems
SA8	AI + Real-Time	Hybrid	Real-time analytics enhancing card security	Conceptual; benchmarking incomplete	Medium–Early stage, conceptual for industry testing
SA9	FL + Blockchain	Privacy-preserving ML	Combines federated learning with blockchain auditability	Communication overhead; vulnerable to poisoning attacks	Medium–Strong privacy preservation but costly in real-time systems
SA10	Network analytics	Hybrid	Novel network-driven fraud detection extensions	Conceptual; limited validation	Low–Needs further validation before deployment
SA11	Deep learning	Anomaly detection	Uses anomaly detection algorithms (KNN, CART, SVM, Isolation Forest) for proactive detection	Scalability uncertain; not widely tested in production	Low–Mostly experimental anomaly methods
SA12	Hybrid ML	Ensemble/ Meta-learning	Multi-layer verification framework combining ML and business rules	Rigid design; less adaptive to emerging patterns	Medium–Could work in layered enterprise systems
SA13	Graph database	Real-time analytics	Implements TigerGraph for fraud communities and alerts	Tested only on limited datasets; integration with ML underexplored	High–Practical for fraud teams needing visual graph alerts
SA14	Ensemble	Meta-learning	Cost-sensitive ensemble reduces false positives in imbalanced data	Computationally expensive; requires cost tuning	Medium–Works well for institutions with imbalanced data challenges

(Continued)

Table 10 (continued)

Study code	Technology category	Methodology category	Key contributions	Limitations	Deployment fit
SA15	Deep learning hybrid	Ensemble/ Meta-learning	Combines autoencoder anomaly detection with XGBoost classification	Vulnerable to concept drift; training-intensive	High–Effective in hybrid detection environments
SA16	Unsupervised ML	Anomaly detection	SVDD with swarm optimization for high-dimensional imbalance	Sensitive to parameters; less interpretable	Medium–Promising for anomaly-heavy datasets
SA17	ML pipelines	Hybrid	End-to-end ML pipeline comparing classifiers	Generic; lacks privacy-preserving mechanisms	Medium–Usable as a comparative baseline tool
SA18	Deep learning	Anomaly detection	Optimized event-based network reduces false positives/negatives	Interpretability challenges; tested on benchmark datasets only	Medium–Needs robust infrastructure to deploy
SA19	Self-supervised learning	Semi-/Self-supervised	Tackles imbalance, noise labels, and drift with SSL + sampling	Relies on quality unlabeled data; high complexity	Medium–Effective with unlabeled data, depends on infra support
SA20	ML (XGBoost)	Supervised ML	Robust supervised classifier with threshold optimization	Limited adaptability to unseen fraud patterns	High–Robust for supervised production tasks
SA21	Big Data + ML	Supervised ML	Hadoop + Logistic regression/Decision Trees	Outperformed by modern DL/GNN; outdated infrastructure	Medium–Good for legacy big data contexts
SA22	ML	Supervised ML	Real-time ML fraud detection pipeline	Limited imbalance handling; based on older models	High–Deployable in real-time fraud monitoring
SA23	Graph neural networks	Incremental anomaly detection	Real-time dense subgraph fraud detection	High memory requirements; community-level focus only	Medium–Strong for graph-heavy fraud communities
SA24	Adaptive security	Hybrid	Combines adaptive authentication with ML-based fraud detection	Complex to deploy; usability trade-offs	High–Practical for integration in customer-facing systems

Fig. 5 presents a tri-partite taxonomy for CNP fraud detection frameworks, classified into technology families, methodology primitives, and specific frameworks or papers. Blue nodes represent technology families, including Hybrid Systems, Graph-Based Frameworks, Federated & Blockchain Systems, and Deep Learning Architectures. Green nodes highlight methodology primitives, including Supervised Learning, Self-Supervised Learning, Graph Reasoning, and Anomaly Detection. Light-blue nodes identify specific frameworks, such as FedGAT-DCNN and SPADE. Central hubs, such as Supervised Learning, Deep Learning, and Federated Learning, indicate frequently reused methodologies. For example, SCARFF links Streaming/Real-Time to Supervised ensembles, while FedGAT-DCNN combines Deep Learning with Graph Reasoning for privacy-preserving Federated Learning. The Hybrid technology cluster integrates multiple methods, signaling a trend towards blended production-ready stacks that combine rules, ML/DL, and graphs with privacy controls. This mapping validates a two-axis taxonomy for CNP fraud, distinguishing between operational technology and reasoning methodology, with central methods showcasing the design patterns that drive modern CNP detection.

Table 10 shows that in terms of technology, the study papers outlined in the table naturally aggregate into six deployment-oriented categories that define the environments in which code operates and how it scales. The first category encompasses rule and behavioral policy layers, exemplified by Singh and Jain's three-layer verification and Bødker et al.'s crime-script lens, which function as millisecond pre-filters and governance controls. The next tier includes classical machine learning (ML) stacks, such as Jeribi's end-to-end ML, Nijwala's XGBoost, and Patil's Hadoop Linear Regression/Random Forest, which prioritize low-latency, tabular scoring techniques. The third category comprises deep learning stacks, which model sessions and events using methodologies such as Patel's LSTM sequence risk, Prabha and Priscilla's LSTM-AE→XGBoost, Devi's anomaly ensembles, and Adil's event-based OptDevNet. On the relational front, graph technologies range from investigator-focused graph databases, as characterized by Mauliddiah and Suharjito, to spatio-temporal and evolving graph engines, represented by Cheng and the SPADE lineage. Complementing these technologies are federated and blockchain infrastructures that support cross-silo learning and provide tamper-evident provenance along with off-chain inference (Baabdullah et al, Li and Walsh, and Mackey et al.). Lastly, streaming and big-data frameworks, including Carcillo's SCARFF on Spark/Kafka/Cassandra, Razaque et al.'s Spark pipelines, and Thennakoon et al.'s real-time learners, offer features such as freshness service level agreements (SLAs), p50/p95 latency service level objectives (SLOs), and champion-challenger rollouts. Collectively, these categories illustrate the operational landscape detailed in the table, incorporating latency constraints, PCI scope segmentation, privacy considerations, resilience under load, and deployment mechanics.

Methodologically, the works included are categorized based on their reasoning and governance frameworks. At one end of the spectrum are deterministic policies, such as rules or thresholds, and crime-script "disruption points" which offer auditable control. The core discrimination layer consists of supervised tabular ML with cost-sensitive ensembles (Olowookere and Adewale) and calibrated XGBoost (Nijwala et al.), often enhanced by unsupervised methods and anomaly detection techniques (Mniai's SVD/one-class; Devi et al.'s ensembles) as well as semi/self-supervised learning to mitigate label scarcity (Chen et al.'s contrastive pretraining and AE pretext signals feeding into XGBoost in Prabha and Priscilla). Where behavior and coordination are pivotal, sequence and representation learning (Patel et al.; Adil et al.) effectively model end-to-end session dynamics. Additionally, relational and graph reasoning, illustrated by features such as APATE, spatio-temporal graph neural networks (GNNs), and graph database community views, enable the identification of mules and criminal networks. Furthermore, overarching themes include privacy-preserving collaboration and provenance, evident in federated learning with secure aggregation and differential privacy, as well as on-chain audits of model, policy, and evidence digests, with inference conducted off-chain. An operational framework is also emphasized, enforcing leakage-safe temporal/entity splits, feature-freshness SLAs, PSI/KS drift monitors, explainability artifacts, and A/B governance measures.

The limitations reflect these methodological choices, highlighting considerations such as the computational and explainable AI burdens associated with deep learning and graph neural networks, parameter sensitivity in anomaly detection pipelines, the rigidity inherent in pure rules, the coordination and communication overhead in federated learning, and the infrastructure tuning requirements for Spark and Hadoop. Conversely, the "Deployment Fit" column aligns neatly with roles, including policy gates, hot-path scorers, pre-scorers and arbiters, investigator tooling, consortium training/audit, and streaming authorization.

This dual-pronged analysis is essential, as technology delineates what can be implemented within the constraints of latency, privacy, and scalability, while methodology clarifies how systems should reason, evaluate, and govern decisions. Technologically analogous systems, such as two Spark streams, may employ markedly different methods (e.g., cost-sensitive XGBoost versus anomaly detection). Likewise, identical methods, like graph neural networks, can operate within frameworks that have very different compliance profiles (e.g., batch graph versus evolving graph streaming).

By maintaining distinct axes and linking key contributions, limitations, and deployment fit, this approach facilitates direct comparisons, targeted integrations (combining rules, ML, DL, graph, and federated learning within a streaming architecture), and reproducible benchmarking tied to both runtime SLOs and scientific rigor. In conclusion, the differentiation between technology and methodology culminates in a precise, deployment-ready taxonomy that delineates how to construct robust systems (addressing architecture, privacy, and operations) and how to engage critically with inference, evaluation, and governance. This strategy renders both effective and feasible audit-ready, real-time CNP fraud frameworks.

Fig. 6 illustrates clusters within the research landscape of CNP fraud detection. The most prominent areas are situated at the convergence of deep learning techniques and hybrid or ensemble approaches. Frameworks such as LSTM autoencoders, integrated with XGBoost and OptDevNet, exemplify how researchers are increasingly harnessing the representational capacity of deep neural networks while employing ensemble classifiers to mitigate challenges related to class imbalance, threshold setting, and model interpretability. This evolution indicates that hybrid deep models have become the predominant methodological paradigm in fraud detection research.

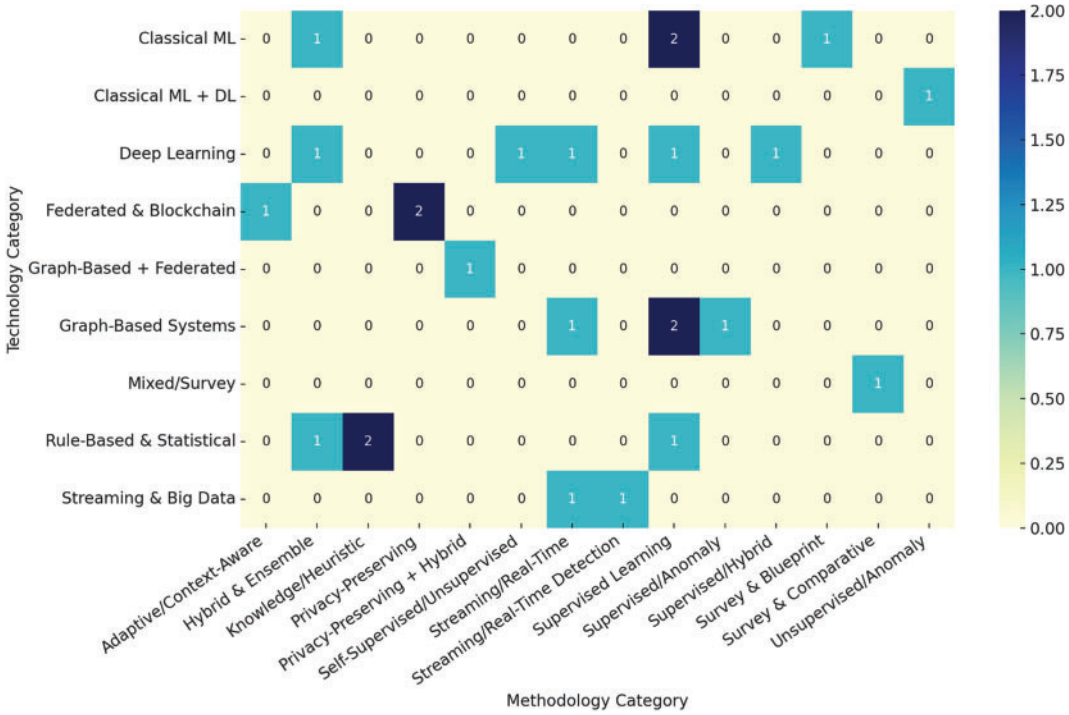

Figure 6: Distribution of CNP fraud detection frameworks by technology and methodology

Another concentration appears in graph-based systems, particularly those utilizing streaming or supervised methodologies. Frameworks like Spade and the Graph Neural Network, via spatial-temporal attention, highlight the growing importance of graph intelligence in capturing community fraud patterns and collusive merchant–customer rings. The fact that these methods often appear in conjunction with streaming indicates a trend towards real-time graph analytics, which is especially relevant for large-scale payment ecosystems.

Conversely, the heatmap shows underrepresented areas. Self-supervised learning and anomaly detection remain relatively underutilized, despite their potential to address label scarcity and concept drift, which are acute challenges in fraud domains. Similarly, adaptive/context-aware frameworks (e.g., integrating biometrics, behavioral analytics, or dynamic multi-factor authentication) are thinly represented. This gap is noteworthy given the increasing demand for user-centric security and regulatory compliance that prioritizes customer trust.

The federated and blockchain-enabled category is moderately populated, reflecting a nascent but promising field. These frameworks address privacy, auditability, and cross-institution collaboration-critical concerns in multi-bank ecosystems. Their lower frequency reflects both their novelty and the infrastructural complexity involved in deployment. However, as regulations like GDPR and PSD2 push institutions toward privacy-preserving analytics, this quadrant is expected to expand rapidly.

Finally, rule-based and statistical frameworks remain the least active cluster, underscoring the field's shift away from brittle, heuristic-driven approaches. Their presence in the literature is primarily found in survey studies or as baseline comparisons, rather than as cutting-edge solutions.

A taxonomy is not just a classification exercise; it is an enabler of coherence, critical reflection, and innovation in the fragmented landscape of CNP fraud detection. By organizing frameworks systematically, taxonomy provides a shared reference that benefits research, deployment, evaluation, and governance. It allows stakeholders to navigate the trade-offs between performance, privacy, compliance, scalability, and explainability with clarity and confidence.

4.2 RQ2: What Core Design Considerations Should Guide the Development of Robust and Effective CNP Fraud Detection Frameworks?

The analysis reveals that the next generation of CNP fraud detection systems must evolve beyond traditional accuracy-centered paradigms. Future systems should be conceptualized as adaptive, interpretable, privacy-preserving, and economically optimized ecosystems. The reviewed literature delineates six interdependent design pillars: (1) scalable and modular architecture, (2) privacy-preserving data governance, (3) adaptive learning and drift management, (4) interpretability and trust, (5) cost-sensitive and latency-aware optimization, and (6) integrated governance accompanied by continuous evaluation. Each of these pillars represents a foundational design consideration that underpins the development of resilient and trustworthy CNP fraud detection frameworks.

Scalable, Modular, and Interoperable Architecture

CNP fraud detection operates in an environment characterized by massive transaction volumes, heterogeneous data streams, and stringent latency constraints. Frameworks such as SCARFF [9] and Spade [41] demonstrate the need for scalable, distributed architectures that can process streaming transactions in micro-batches with sub-second latency. SCARFF leveraged Spark-based parallelization to maintain real-time throughput, while Spade introduced an evolving graph model $G = G \oplus \Delta G$ that incrementally updates fraud relations as new data arrives.

A robust architecture should therefore adopt layered modularity, comprising a perception layer (data ingestion), a network layer (secure transmission), a processing layer (model computation), an analytics layer (decision intelligence), and a governance layer (policy and compliance). Such modularity ensures extensibility, interoperability, and fault isolation, allowing future integration of quantum-safe cryptography or explainable AI modules without re-engineering core systems.

Real-world application: The SCARFF framework [9]exemplifies a modular, scalable architecture using Apache Spark, Kafka, and Cassandra. It achieved a sub-100 ms detection latency across high-frequency streaming data, which is essential for real-time fraud screening in banking and e-commerce environments. Its modular pipeline enabled seamless integration and scaling across diverse components, making it deployment-ready and resilient. This real-world case illustrates the importance of microservices and multi-layered architectures in maintaining throughput while preserving responsiveness.

Privacy-Preserving Data Governance and Collaborative Intelligence

The reviewed frameworks converge on the need to balance data-driven intelligence with regulatory compliance under frameworks such as GDPR, PCI-DSS, and PSD2. Emerging architectures employ Federated Learning (FL) and Blockchain-based auditing to enable cross-institutional collaboration without exposing raw customer data. The FL update rule

$$w^{(t+1)} = \sum_i \frac{|D_i|}{|D|} w_i^t$$

as implemented in FedGAT-DCNN [24], aggregates encrypted local models into a global network while maintaining confidentiality. Differential Privacy ensures individual anonymity through

$$Pr\left[M\left(D\right) = O\right] \leq e^{\epsilon} Pr\left[M\left(D'\right) = O\right],$$

limiting the influence of any single record. Blockchain-anchored antifraud systems [18] extend this protection by providing tamper-proof audit trails, ensuring transparency, traceability, and trust in collaborative environments. Collectively, these techniques redefine data governance from passive compliance to active, cryptographically enforced accountability.

Real-world application: FedGAT-DCNN [24] combines federated learning with graph attention networks (GAT) and blockchain for secure, cross-border fraud detection. Empirical testing showed a 98.6% F1-score with only 25 ms latency while maintaining GDPR and PCI-DSS compliance. Blockchain-enabled immutable audit logs of model updates and fraud alerts, supporting institutional transparency. This approach validates the practical feasibility of decentralized fraud learning systems where data cannot be centralized due to privacy or jurisdictional constraints.

Adaptive Learning, Drift Management, and Context Awareness

Fraud typologies evolve continuously, rendering static detection rules obsolete. The most effective frameworks exhibit adaptive intelligence through continual learning and drift-resilient mechanisms. Self-Supervised Sampling [36] integrates auxiliary pretext tasks to maintain robust feature embeddings, sustaining F1-scores above 0.9 under shifting fraud patterns. Next-Generation Anomaly Detection [29] employs sliding-window retraining to dynamically refresh model parameters, while Graph Neural Networks [21] utilize spatio-temporal attention to capture evolving relational dependencies between merchants, devices, and cardholders.

In practical deployment, adaptive drift management enables models to self-adjust when global events (e.g., pandemic-induced e-commerce surges) introduce new behaviors. Thus, future frameworks must embed automated drift diagnostics, online recalibration, and feedback loops to ensure sustained accuracy and operational continuity.

Real-world application: The Spade framework [31] leverages evolving graphs and temporal GNNs to address concept drift in CNP fraud. Deployed on live transaction graphs, it maintained over 91% accuracy despite behavioral shifts in user sessions. Unlike static models, Spade's real-time adaptation to new fraud typologies made it ideal for environments such as mobile banking or peer-to-peer payments. This confirms the necessity of continuous learning and adaptive thresholds in dynamic fraud ecosystems.

Interpretability, Explainability, and Human-Centered Trust

Regulatory accountability and investigator acceptance depend on transparent model reasoning. Explainable AI (XAI) techniques, such as SHAP value decomposition [43] quantify each feature's contribution to fraud classification:

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!\,(|F| - |S| - 1)!}{|F|!} \left[f\left(S \cup \{i\}\right) - f\left(S\right) \right]$$

Graph-based systems like Spade [41] supplement this with suspiciousness density

$$\sigma\left(V\right) = \frac{S\left(V\right)}{|\,V\,|},$$

highlighting dense clusters of anomalous transactions. OptDevNet [36] further integrates attention heat maps, while Crime-Script Analysis [11] provides a qualitative lens linking algorithmic output to offender behavior. Embedding multi-level explainability, from model inference to investigator interface, strengthens human-AI collaboration and satisfies the GDPR Article 22 requirement for intelligible explanations of automated decisions.

Real-world application: ref. [11] introduced a crime-script methodology to model the full attack surface of CNP fraud, encompassing account creation, post-transaction disputes, and all stages in between. These scripts informed the placement of rules and the activation of ML across user flows. In regulatory design, the approach helped translate detection insights into legal and compliance language, thereby reducing alert fatigue and enhancing audit trails. Coupled with attention maps and SHAP outputs from frameworks like OptDevNet, this underscores the regulatory and operational need for explainable AI.

Cost-Sensitive and Latency-Aware Optimization

CNP fraud detection frameworks must balance predictive performance with both economic utility and operational responsiveness. Traditional metrics such as accuracy are insufficient in highly imbalanced settings where legitimate transactions dominate. Frameworks including APATE [12] and SCARFF [9] emphasize cost-sensitive optimization, leveraging

$$Balance\ Accuracy\left(BA\right) = \frac{Sensitivity + Specificity}{2},\ Value\ Detection\ Rate\left(VDR\right) = \frac{\sum Value_{detected}}{\sum Value_{fraud}},$$

to evaluate financial impact rather than raw classification counts. These models are constrained by detection latency thresholds, typically under 100 ms, to prevent service disruption [41]. Integrating edge-level pre-screening, adaptive thresholds, and load-balanced inference pipelines ensures that fraud detection enhances security without compromising user experience or throughput.

Real-world application: ref. [28] used graph-based models and cost-sensitive learning to detect mule networks and high-loss fraud scenarios under tight latency budgets (<30 ms). It prioritized decisions based on financial impact, reducing false positives and enabling quick authorization decisions during checkout. The system's graph relational learning extended detection beyond simple outliers, capturing multi-party fraud relationships and decision trade-offs.

Integrated Governance, Security, and Continuous Evaluation

The sustainability of CNP detection systems depends on their ability to operate within secure and accountable governance frameworks. Frameworks such as Towards an Intelligent Adaptive Security System [12] and Next-Gen Anomaly Detection [29] implement policy-driven rule engines and feedback-based recalibration for continuous improvement. Incorporating governance dashboards, explainability audits, and privacy-impact assessments converts technical models into auditable organizational systems of trust. Evaluation must therefore be multidimensional, combining accuracy metrics with Cost-Sensitive Utility, Detection Latency, Explainability Fidelity, Drift Resilience, and Privacy Compliance Index to ensure sustained ethical and operational alignment.

Real-world application: The Next-Gen Anomaly Detection Framework [29] institutionalized model oversight by integrating dashboards tracking explainability, fairness, privacy overhead, and concept drift. In real-world deployment, it reduced fraud investigation time by 25% due to high-clarity SHAP explanations and continuous evaluation metrics. This demonstrates that performance alone is insufficient; governance and transparency must be embedded for institutional trust and regulatory compliance.

Design Pillars and guiding considerations

Table 11 presents a comprehensive summary of each design pillar, including its definition, key considerations, representative frameworks, and anticipated outcomes.

Table 11: Core design pillars and guiding considerations for robust CNP fraud detection

Design pillar	Definition/Objective	Key design considerations	Frameworks	Expected outcomes
1. Scalable, modular, and interoperable architecture	To ensure high-volume transaction processing with minimal latency while supporting extensibility and system interoperability.	Employ distributed or microservice architectures (e.g., Spark, Kafka, or Graph-based streaming). Adopt multi-layer design (Perception–Network–Processing–Analytics–Governance). Enable API-level interoperability for integration with banking, merchant, and regulatory systems.	SCARFF, Spade; Remote Banking	Real-time scalability with sub-100 ms response; flexible updates without system downtime.
2. Privacy-preserving data governance and collaboration	To enable collaborative fraud detection while ensuring regulatory compliance and data protection.	Implement Federated Learning (FL) to aggregate models without sharing raw data. Enforce Differential Privacy (DP) for formal confidentiality guarantees. Use Blockchain for immutable auditability and consent management.	FedGAT-DCNN; FL + Blockchain, Blockchain Antifraud	GDPR/PCI-DSS compliance; secure cross-institution learning; enhanced audit traceability.
3. Adaptive learning, drift management, and context awareness	To maintain consistent detection performance amid evolving fraud patterns and behavioral drift.	Integrate online learning or self-supervised models for continuous adaptation. Use spatio-temporal graph attention or sliding-window retraining for drift correction. Automate model calibration based on new transaction behaviors.	Self-Supervised Sampling; Next-Gen Anomaly; GNN	Sustained accuracy (>0.9 F1) under data drift; faster response to emerging fraud typologies.

(Continued)

Table 11 (continued)

Design pillar	Definition/Objective	Key design considerations	Frameworks	Expected outcomes
4. Interpretability, explainability, and human-centered trust	To provide transparency and auditability of automated decisions for analysts and regulators.	Apply SHAP/LIME for feature-level explanation. Visualize graph anomalies using Suspiciousness Density or attention heatmaps. Complement algorithmic explainability with human-readable narrative insights (e.g., Crime Scripts).	Crime Scripts; Spade; OptDevNet; Adversarial Autoencoder	Increased interpretability; regulatory compliance (GDPR Art. 22); analyst confidence in system outputs.
5. Cost -sensitive and latency-aware optimization	To optimize detection thresholds based on financial utility and decision speed.	Integrate cost-sensitive metrics (BA, VDR) to reflect real-world loss scenarios. Maintain detection latency <100 ms for real-time authorization. Deploy lightweight edge or adaptive inference to reduce compute overhead.	SCARFF, APATE, OptDevNet	Improved financial savings; reduced false positives; frictionless customer experience.
6. Integrated governance, security, and continuous evaluation	To institutionalize accountability and resilience through policy-driven oversight and ongoing model validation.	Establish governance dashboards for privacy, fairness, and explainability audits. Align detection thresholds with institutional risk appetite. Implement periodic model re-evaluation against multidimensional indicators (Utility, Latency, Explainability, Drift, Privacy).	Intelligent Adaptive Security, Next-Gen Anomaly	Continuous assurance of ethical, technical, and regulatory performance; institutionalized trust.

Across all frameworks, design evolution reflects a paradigm shift from centralized, opaque, and accuracy-centric systems to distributed, interpretable, and privacy-conscious ecosystems. Robust CNP fraud detection requires the convergence of federated graph analytics, edge-enabled real-time processing, and explainable deep learning, all anchored in transparent governance. Frameworks that internalize these design principles consistently demonstrate higher adaptability, stronger compliance, and reduced false-positive and latency rates. Future architectures should further institutionalize these pillars by adopting standardized privacy indices, interoperable data-exchange protocols, and self-governing audit layers to ensure that CNP fraud detection systems remain resilient, ethical, and trustworthy in increasingly complex financial environments.

By embedding these design principles from inception, future fraud detection systems can better navigate the adversarial nature of CNP fraud, the demand for regulatory alignment, and the realities of production-scale deployment.

4.3 RQ3: What Innovative Performance Indicators Can Be Used to Assess CNP Fraud Detection Frameworks More Effectively, Moving beyond Traditional Accuracy Measures?

In evaluating CNP fraud detection systems, traditional measures such as accuracy or recall have proven insufficient because they fail to capture the operational, financial, and regulatory realities of fraud management. This has motivated the inclusion of five novel indicators: cost-sensitive utility, detection latency, explainability, drift resilience, and privacy compliance, each of which broadens the evaluative lens and ensures that proposed frameworks address both technical and real-world requirements.

Cost-sensitive utility accounts for the financial consequences of fraud detection decisions. Unlike raw accuracy, which treats all misclassifications equally, cost-sensitive metrics weigh outcomes by transaction value and false-positive costs. Balanced Accuracy (BA),

$$BA = \frac{Sensitivity + Specificity}{2},$$

is widely applied to correct class imbalance, while the Value Detection Rate (VDR),

$$VDR = \frac{\sum \text{Value of detected fraud}}{\sum \text{Total fraud value}},$$

ensures that capturing high-value fraud is prioritized [9,12]. For instance, the APATE framework reported a Balanced Accuracy of 93.2% at 1% FPR, demonstrating its alignment with real-world thresholds where operational costs of false alarms are critical.

Equally important is detection latency, which measures how quickly a model flags suspicious activity relative to the transaction time. This is expressed as average detection latency (ADL):

$$ADL = \frac{\sum \left(t_{decision} - t_{transaction}\right)}{n},$$

where $t_{transaction}$ and $t_{decision}$ denote transaction and classification times respectively. Frameworks such as Spade optimize real-time detection by incrementally updating transaction graphs,

$$G = G \oplus \Delta G, S\left(V\right) = \sum_{e \in E(V)} w\left(e\right),$$

achieving alerts within 100 ms on evolving graphs [41]. Similarly, sequential learners update recurrent hidden states,

$$h_t = f\left(W_{hh}h_{t-1} + W_{hx}x_t\right),$$

to capture dwell times in remote banking sessions [19]. These approaches demonstrate the importance of low-latency detection for preventing transaction blocking before multiple fraudulent charges accumulate.

Beyond speed, explainability has emerged as a core requirement for regulatory compliance and institutional trust. Post-hoc methods such as SHAP provide feature-level attributes:

$$\phi_i = \textstyle\sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!}[f(S \cup \{i\}) - f(S)]$$

where ϕ_i quantifies the contribution of feature i [43]. In graph-based systems, suspiciousness density,

$$\sigma\left(V\right) = \frac{S\left(V\right)}{\mid V \mid},$$

assesses the concentration of anomalies in collusive networks [41]. Narrative techniques, such as crime scripts [11], supplement interpretability by framing fraud as a sequence of human actions. These methods ensure that fraud decisions are not opaque black boxes but instead provide actionable insights to analysts.

Another challenge is drift resilience, the ability of systems to maintain performance as fraud patterns evolve. This resilience is often measured using the F1-score:

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}, Precision = \frac{True\,Positive\,(TP)}{True\,Positive\,(TP) + False\,Positive\,(FP)},$$

with stability in F1 or ROC-AUC under drift serving as a key evaluation criterion. Recent work on self-supervised learning adapts embeddings to evolving transaction behaviors [36], while federated graph attention networks maintain ROC-AUC near 0.9992 under parameter variations [10]. These examples show how drift-resilient systems remain effective even during abrupt behavioral shifts, such as fraud surges during the COVID-19 pandemic.

Finally, privacy compliance has become a defining concern under frameworks such as GDPR and PCI DSS. Federated Learning (FL) offers collaborative fraud detection without sharing raw customer data through weighted model aggregation,

$$w^{(t+1)} = \sum_i \frac{|D_i|}{|D|} w_i^t,$$

while Differential Privacy (DP) provides formal guarantees,

$$Pr\left[M(D) = O\right] \leq e^{\epsilon} \cdot Pr\left[M(D') = O\right],$$

ensuring outputs reveal little about any single individual [24,30]. Use cases such as FL+Blockchain allow multiple banks to detect cross-institution fraud while maintaining strong privacy guarantees and achieving Accuracy = 0.97, F1 = 0.97.

Taken together, these five indicators shift the evaluative paradigm for credit card fraud detection frameworks. Where earlier studies emphasized raw classification accuracy, contemporary research emphasizes economic utility, speed, transparency, adaptability, and regulatory compliance. Their integration ensures that fraud detection systems are not only technically proficient but also operationally viable, trustworthy, and aligned with the realities of modern financial ecosystems.

Table 12 provides an overview of the essential characteristics of the indicators. It features a definition that clarifies the meaning of each indicator, a carefully selected list of notable scholarly references that outline how the indicator is defined or applied, and a thorough description of the evaluation and testing methods used in research to verify its validity and reliability.

Table 12: Novel performance indicators for evaluating CNP fraud detection frameworks

Indicator	Definition	Category	Focus	Why it matters	Key evaluation/Testing methods
Cost-sensitive utility	Balances fraud detection with financial costs, accounting for false positives and transaction value.	Business Impact	Financial return & cost sensitivity	Captures true financial value of detection, especially where false positives can cause lost revenue and user friction.	Balanced Accuracy; Value Detection Rate (VDR); constrained False Positive Rate; cost-sensitive thresholds using transaction value.

(Continued)

Table 12 (continued)

Indicator	Definition	Category	Focus	Why it matters	Key evaluation/Testing methods
Detection latency	Measures how quickly fraud is detected after a transaction occurs (ms/transaction).	Operational Efficiency	Technical deployability & latency control	Essential for real-time systems like SCARFF; ensures fraud detection works within tight SLA constraints.	Average response time; time-to-detection; throughput (transactions/sec); sub-100 ms real-time benchmarks.
Explainability	Extent to which model decisions can be understood and trusted by analysts/regulators.	Human-Centric/Review	Analyst efficiency & decision support	Focuses on operational efficiency of fraud analysts. Prioritizes decision support rather than bulk accuracy.	LIME, SHAP feature importance; rule-based interpretability; crime scripts; user studies with domain experts.
Drift resilience	Model's ability to adapt to evolving fraud patterns (concept drift).	Adaptability/ Drift	Adaptability to fraud evolution	Critical for measuring a model's resistance to concept drift, a common issue in adversarial fraud contexts.	Sliding windows; self-supervised adaptation; online anomaly detection; drift detectors (ADWIN, EDDM); evaluation on time-sequenced datasets.
Privacy compliance	Adherence to privacy regulations while maintaining detection accuracy.	Governance/ Regulation	Legal compliance & governance readiness	Measures whether the system aligns with data minimization, consent, and security-by-design—key for production and audit approval.	Federated learning protocols; differential privacy (ε-values); blockchain auditability; compliance with GDPR/PCI DSS; accuracy vs. privacy trade-off.

Trends in Novel Fraud Detection Indicators

The analysis of evaluation indicators, as illustrated in Fig. 7, reveals a clear evolutionary trajectory in the research on card fraud detection. Early frameworks emphasized classification accuracy but provided limited attention to operational concerns. Over time, cost-sensitive utility has steadily advanced, with recent systems explicitly incorporating transaction value and false-positive costs through metrics such as Value Detection Rate (VDR) and savings rate, ensuring that improvements translate into measurable financial benefit. Detection latency has seen the most consistent gains, particularly after 2020 with the emergence of scalable streaming and graph-based models (e.g., SCARFF, Spade), which demonstrate response times within sub-100 ms thresholds. By contrast, explainability has remained stagnant, with only marginal progress achieved through post-hoc techniques (e.g., SHAP, LIME) and domain-informed crime scripts, highlighting a persistent trade-off between model complexity and interpretability. Drift resilience exhibits gradual yet meaningful improvement, driven by adaptive machine learning, self-supervised objectives, and online detection strategies that can adjust to evolving fraud patterns. Finally, privacy compliance has experienced the most significant leap in recent years, propelled by federated learning and blockchain integration, which allow collaborative fraud detection across institutions while preserving regulatory alignment (GDPR, PCI DSS). Collectively, these trends underscore a paradigm shift: whereas early research prioritized detection efficiency, contemporary frameworks emphasize the preservation of privacy and adaptive resilience, marking a reorientation toward holistic, real-world viability.

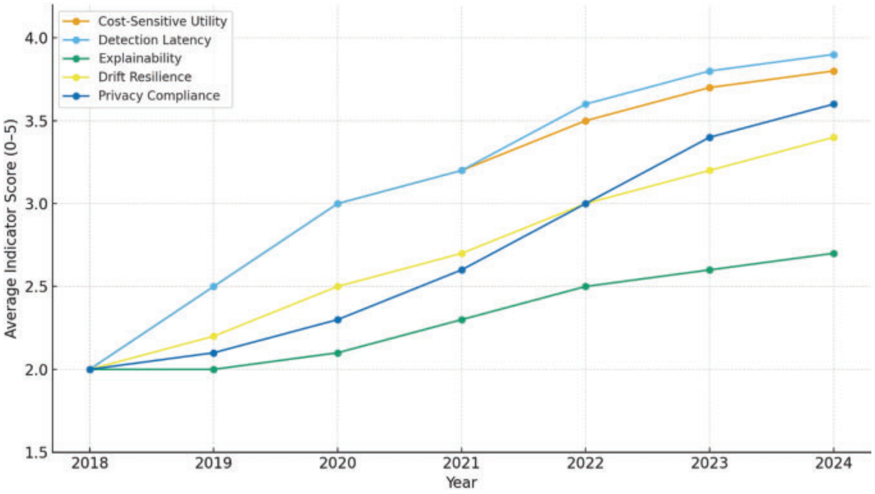

Figure 7: Trends in novel fraud detection indicators

The expanded set of metrics addresses critical research limitations and deployment challenges within the emerging field of CNP fraud detection. As financial institutions strive to integrate these advanced solutions, they require more than just high AUC scores; they seek systems that are both explainable and resource-efficient while also complying with regulatory standards. The inclusion of such metrics in model evaluation can significantly enhance trust among business and compliance stakeholders, facilitate clearer analysis of the return on investment (ROI) for fraud prevention initiatives, and promote standardization across various vendors and regulatory bodies. By embedding these innovative indicators into future projects, the CNP fraud detection community can accelerate the transition from purely academic models to practical, production-grade systems that are both ethically grounded and operationally feasible.

These five innovative indicators form the core of a next-generation evaluation framework for identifying CNP fraud. They connect predictive power with real-world usability, and their adoption can ensure that detection systems are not only accurate but also actionable, compliant, adaptable, and cost-effective.

4.4 RQ4: What Research Gaps Remain in the Current Literature, and What Future Directions Should Be Pursued to Advance Next-Generation CNP Fraud Detection and Prevention Systems?

Despite notable advancements in fraud detection technologies, including deep learning, graph neural networks, federated architectures, and self-supervised learning, numerous persistent challenges continue to undermine the operational viability, regulatory compliance, and institutional effectiveness of CNP fraud detection systems. A pivotal issue is the phenomenon of concept drift and label scarcity. As fraudulent behaviors evolve continuously, reliable labeled data frequently arrives too late or lacks the granularity required for effective detection. This inadequacy can lead to significant degradation in model accuracy and adaptability over time. Although some frameworks have implemented self-supervised and weakly supervised methodologies to alleviate the scarcity of labeled data, few have successfully operationalized these approaches in real-time environments. Consequently, many systems remain reactive rather than anticipatory due to the absence of online retraining, drift-aware evaluation, and dynamic label calibration mechanisms.

Moreover, another significant limitation is the insufficient utilization of graph and federated architecture, despite their alignment with the structural and privacy challenges inherent in fraud detection. Graph Neural Networks (GNNs) offer a robust framework for capturing relationships among users, transactions, and devices, which is particularly beneficial for identifying collusive fraud rings or synthetic identities. In parallel, federated learning permits institutions to train models without the need to exchange sensitive data collaboratively. However, these methodologies are underexploited in operational settings due to various obstacles, including the complexities of ongoing graph maintenance, communication overhead, latency constraints, and heterogeneous compliance requirements. Such technical and governance barriers inhibit the scalability and interoperability of systems that could otherwise facilitate cross-institutional fraud detection while remaining compliant with privacy regulations such as the General Data Protection Regulation (GDPR) and the Payment Card Industry Data Security Standard (PCI-DSS).

Compounding these architectural deficiencies is a growing inadequacy in model explainability and human oversight, a concern that intensifies as fraud detection systems become increasingly complex. Black-box models, such as LSTM autoencoders, deep ensembles, and GNNs, may demonstrate impressive F1 scores but often lack the transparent, case-level reasoning necessary for investigators, compliance officers, and customer service teams to validate decisions. This absence of interpretability not only erodes analyst trust and contributes to alert fatigue but also limits legal defensibility in disputed transactions. Furthermore, many systems fail to integrate human-in-the-loop feedback mechanisms that could enhance learning outcomes and align policies effectively. In the absence of explainable AI interfaces and investigator-facing resources, such as SHAP values or session visualizations, the accountability and auditability of fraud-related decisions are critically compromised.

Even when technically robust, many fraud detection frameworks lack deployment-scale validation, rendering them ill-suited for real-world integration. A significant proportion of studies rely on static datasets, neglect temporal and entity disjoint splits, or report evaluation metrics that do not account for latency, throughput, or the costs associated with false positives. These oversights distort performance expectations and hinder regulatory acceptance. In high-throughput, service-level agreement (SLA)-bound environments, such as mobile banking and real-time payment processing, sub-30 ms scoring times, rollback strategies, and model versioning are not optional; they are imperative. Yet, few systems implement or report on these critical features, leaving financial institutions hesitant to trust or integrate these models into mission-critical operations.

The implementation of advanced fraud detection systems presents considerable challenges related to scalability and real-time performance. Although cutting-edge models may demonstrate efficacy in offline environments, they frequently encounter difficulties under production loads due to computational constraints and inference delays. Real-time scoring requires rapid data retrieval, memory efficiency, and latency-aware thresholding, all aspects that are often overlooked in academic prototypes. As fraudulent activities evolve in sophistication, models must be designed to scale effectively while ensuring low latency and high throughput, particularly within edge or hybrid cloud environments. Regrettably, many existing frameworks lack adaptive orchestration or fallback mechanisms, resulting in performance bottlenecks.

Governance and privacy constraints further complicate the landscape, imposing significant restrictions on data access, model explainability, and the sharing of intelligence across entities. Regulatory frameworks such as the General Data Protection Regulation (GDPR), the Payment Services Directive 2 (PSD2), and the Payment Card Industry Data Security Standard (PCI-DSS) impose stringent limitations on data portability, algorithmic transparency, and automated decision-making. Despite the emergence of privacy-preserving technologies, including secure aggregation, differential privacy, and blockchain-based audit trails, most frameworks do not include built-in compliance layers or audit-ready provenance mechanisms. This absence not only hampers model design but also constrains inter-institutional collaboration, which is vital for detecting multimodal fraud schemes, such as those involving synthetic identities or mule account networks.

Another pressing limitation is the lack of standardized evaluation metrics. Traditional metrics, such as accuracy and precision, fail to adequately capture critical performance dimensions essential for business success, including decision latency, shifts in customer approval rates, the costs associated with false negatives, and response times to concept drift. Few frameworks provide metrics such as Customer Service Uptake (CSU), Drift Response Speed (DRS), or Average Precision at top-K alerts (AP@K), making it difficult to compare models or assess their readiness for production deployment. Without standardized evaluation protocols and documentation practices, such as model cards or reproducible benchmarks, transparency and interoperability of frameworks are severely compromised.

Furthermore, a disjunction exists between fraud analytics and policy implementation, where machine learning outputs are not effectively aligned with real-time business rules. Many fraud detection systems generate risk scores that fail to systematically influence the approval or decline decisions made by human operators. This lack of alignment hinders systems from aligning with organizational objectives related to customer experience, risk tolerance, or economic trade-offs. The consequences may include overblocking, which results in lost revenue, or underblocking, which leads to fraud leakage, both of which erode institutional confidence in automated fraud decision-making.

Modern systems must also contend with increasing concerns regarding energy and communication efficiency, particularly as models become more resource-intensive and are deployed across distributed or edge-based infrastructures. While deep learning and federated architecture hold considerable promise, they often lead to significant power consumption and communication overhead. Nevertheless, few studies have benchmarked the energy efficiency of federated learning or evaluated the costs associated with these rounds. As sustainability becomes an increasingly critical performance metric, both environmentally and financially, future systems must integrate principles of green AI to maintain viability at scale.

Ultimately, the challenge of cross-domain generalization remains a significant limitation to the adaptability of fraud detection systems. Models that are trained within a specific geographic region, merchant category, or customer segment often demonstrate poor performance when applied to novel domains. Fraudulent behaviors are highly contextual, influenced by regional transaction patterns, cultural norms, and variations in infrastructure. Yet, many systems fail to incorporate techniques such as domain adaptation, transfer learning, or meta-learning that could enhance their performance in previously unseen environments. Absent these capabilities, the expansion of fraud detection mechanisms into diverse global markets remains significantly constrained.

In short, these challenges highlight that the limitations currently facing Card Not Present (CNP) fraud detection systems are not predominantly algorithmic; rather, they are systemic and translational in nature. The transition from proof of concept to production requires more than mere innovation; it necessitates a focused emphasis on deployment-aware design, human-centered explainability, privacy-by-design architecture, and rigorous evaluation standards. Future frameworks must also exhibit resilience against concept drift and address issues of interpretability, scalability, and governance to fully realize their potential in combating fraud.

Research Gaps and Strategic Future Directions

Table 13 provides a comprehensive and structured framework for advancing the detection of card-not-present (CNP) fraud. It delineates how each proposed solution addresses specific challenges while being integrated into a phased implementation timeline. This timeline effectively aligns short-term operational enhancements with mid-term architectural advancements. Such a methodical approach ensures that research does not merely remain theoretical but is instead transmuted into practical, actionable steps that can be deployed within real-world financial ecosystems. The inclusion of short-term priorities, such as the implementation of sidecar Graph Neural Networks (GNNs), the piloting of explainability dashboards, and the integration of evaluation kits, underscores a commitment to low-barrier yet high-impact initiatives. These actions aim to enhance the reliability of fraud detection without necessitating a complete overhaul of existing systems.

Crucially, the table establishes a validation pathway for each proposed solution, grounded in measurable criteria encompassing performance, interpretability, privacy, and deployment readiness. Rather than relying solely on conventional metrics such as accuracy or F1-score, it advocates for the use of more contextually relevant measures. including drift response speed (DRS), approval rate impact, explainability compliance, energy profiling, and latency tracking (e.g., p95 benchmarks). This shift signals a significant transition in evaluation methods, from static benchmarking to dynamic, context-aware, and governance-compliant validation. Such metrics are essential for fostering confidence among institutional stakeholders, particularly in contexts where real-time decision-making and regulatory accountability are paramount.

A notable strength of this roadmap lies in its system-level integration across model design, infrastructure, policy alignment, and user interaction. The proposed solutions are not mere technical fixes; rather, they form part of a broader strategic ecosystem. For instance, model transparency is linked to human-in-the-loop oversight, deployment validation is associated with audit-ready model cards, and risk policy tuning is connected to economic cost optimization. This holistic integration reflects a mature understanding of fraud detection as an interdisciplinary domain that necessitates alignment between data science, legal compliance, operational efficiency, and business risk management.

Ultimately, the anticipated impact on CNP fraud mitigation is both immediate and long-term. In the short term, the roadmap facilitates faster detection cycles, reduces false positives, and enhances analyst trust, benefiting both fraud operations and customer experience. In the mid-term, it lays the foundation for collaborative, adaptive, and privacy-preserving architectures capable of scaling across diverse geographies and institutions. By aligning specific solutions with tangible outcomes, the table serves as a strategic blueprint for researchers, regulators, and stakeholders in the payments industry who seek to establish resilient, explainable, and future-ready fraud prevention systems in an increasingly digital financial landscape.

Table 13: Linking research gaps to strategic future directions

Research gap	Proposed research direction/Solutions	Short-term priorities	Mid-term goals	Validation pathway	Impact on CNP fraud mitigation
1. Concept drift & label scarcity	Develop adaptive and label-efficient learners using self-supervised, semi-supervised, and weak-label paradigms (e.g., 3DS outcomes, dispute codes, biometric verifications). Integrate daily calibration pipelines and active learning to adjust models dynamically.	Implement self-supervised pretraining (contrastive or reconstruction tasks); define weak-label ontology; deploy FN-loss-aware active learning.	Transition to continual learning pipelines integrated with federated updates; automated drift detection with real-time retraining.	Use temporal evaluation splits and drift-triggered A/B testing on live transaction streams; monitor F1, FN-loss, and drift response speed (DRS).	Enhances adaptability to emerging fraud tactics; maintains robustness in non-stationary environments; reduces model degradation.
2. Underuse of graph & federated architectures	Build federated graph-based models combining GNN relational reasoning with secure aggregation for privacy-preserving multi-institution training. Use graph embeddings for entity linkage and community detection.	Pilot lightweight GNN sidecar modules on existing fraud systems; simulate cross-bank FL rounds with synthetic data.	Deploy federated graph learning (FGL) across multiple PSPs/banks; integrate blockchain-anchored model provenance.	Validate via CNP-GraphNet benchmarks (APATE, SPADE) with latency and scalability tests; measure FN reduction on ring/mule detection.	Improves multi-party fraud detection and ring identification while maintaining GDPR/PCI-DSS compliance; fosters collaborative intelligence.
3. Lack of explainability & human oversight	Design investigator-centric explainability layers (ExplainHub) using SHAP/LIME, subgraph visualization, temporal attention maps, and evidence packs for each alert. Integrate human-in-the-loop feedback for active learning.	Generate case-level evidence packs and SHAP summaries; deploy visual dashboards for investigator review.	Create interactive, explainable AI interfaces; reinforce model updates through analyst feedback loops; develop explanation standards for audits.	Measure alert resolution time, investigator satisfaction, and AP@K triage accuracy; test explainability compliance vs. EU AI Act/PSD2.	Increases trust, transparency, and investigator efficiency; improves dispute resolution and compliance readiness.
4. Lack of deployment-scale validation	Establish standardized deployment protocols (CNP EvalKit) with model cards, leakage detection, and unified performance metrics (CSU, DL, DRS, AP@K). Enforce real-time validation within SLA constraints.	Implement EvalKit with entity-disjoint splits; generate model cards per deployment; track latency (p95), drift, and approval-rate impact.	Integrate continuous evaluation and policy versioning within fraud ops; enforce governance alignment and reproducible benchmarking.	Conduct live canary/champion-challenger deployments; validate through drift logs, approval metrics, and SLA adherence.	Ensures operational reliability and comparability across frameworks; reduces false trust in inflated offline results; enables regulator-grade validation.

(Continued)

Table 13 (continued)

Research gap	Proposed research direction/Solutions	Short-term priorities	Mid-term goals	Validation pathway	Impact on CNP fraud mitigation
5. Scalability & real-time performance constraints	Create streaming architectures (StreamForge) using microservices, Kafka/Flink, and edge inference with model fallback strategies. Include adaptive thresholding for latency-aware decisioning.	Build baseline XGBoost pipelines with latency profiling; deploy Redis/Feast feature stores; set 30 ms p95 latency targets.	Expand to hybrid edge-cloud orchestration with evolving graph and transformer modules; optimize GPU/TPU inference.	Validate via stress-testing and throughput benchmarking; measure decision latency, transaction rate, and FN/FP trade-offs.	Guarantees sub-30 ms fraud decisions; supports scalability for high-volume systems; ensures performance predictability under load.
6. Governance & data-privacy constraints	Deploy Privacy/Provenance Profiles (ProvTrack) using secure aggregation, differential privacy (ε, δ), and blockchain-based provenance anchoring; enforce audit trails for every fraud decision.	Log provenance bundles (model ID, policy ID, feature hash, timestamp, signature) per decision; pilot secure FL aggregation.	Integrate cross-jurisdiction audit frameworks and standardized decision provenance into live deployments.	Evaluate audit latency, privacy budget consumption, and FL round communication overhead; audit compliance via GDPR/PCI controls.	Enables cross-institution collaboration under legal constraints; ensures transparency and accountability in automated fraud decisions.
7. Non-standardized evaluation metrics	Define and operationalize new CNP-specific metrics—CSU, DL, AP@K, DRS—as part of model documentation. Incorporate business KPIs into ML evaluation pipelines.	Add operational metrics to all experiments; automate metric logging and reporting via EvalKit/BI dashboards.	Industry-wide adoption of a CNP model card standard for evaluation transparency.	Validate through multi-dataset benchmarking (temporal/entity-split) and independent replication of published metrics.	Enhances comparability, transparency, and auditability; bridges the research-industry evaluation gap.
8. Weak policy–analytics coupling	Develop RiskBand Policy Engine to align ML outputs with business logic using cost-sensitive thresholds, automated drift responses, and approval-rate guardrails.	Define approve/step-up/decline bands by transaction size and risk; tune FN/FP trade-offs with economic weights.	Automate policy re-optimization triggered by PSI/KS drift; document version changes for governance review.	Evaluate FN-loss and approval-rate stability pre- and post-policy updates; monitor AR deviation ≤ 0.3 pp.	Improves decision stability and business alignment; enhances adaptive response to behavioral drift.
9. Energy & communication efficiency	Introduce built-in profilers to monitor energy/transaction and communication cost per federated round. Publish environmental performance metrics.	Instrument profiling into model training/serving; benchmark energy per 10 k tx and bytes per FL round.	Create green ML dashboards and integrate energy efficiency into model-selection policies.	Validate energy/comms regression per release; benchmark against FL reference studies.	Reduces operational costs; promotes sustainable AI in large-scale fraud detection.

(Continued)

Table 13 (continued)

Research gap	Proposed research direction/Solutions	Short-term priorities	Mid-term goals	Validation pathway	Impact on CNP fraud mitigation
10. Limited cross-domain generalization	Implement TransferBench for cross-merchant/geographic adaptation using domain re-weighting and meta-learning. Evaluate zero-shot and k-shot fine-tuning.	Generate merchant/MCC/geo holdouts for model testing; assess portability with ΔPR-AUC metrics.	Develop domain-adaptive fraud models with meta-regularization for unseen markets.	Validate portability on multi-region datasets; compare baseline vs. adapted models.	Expands model robustness across ecosystems; reduces cold-start risk for new merchants or payment regions.

5 Conclusion and Recommendations

Conclusion

This study presents a comprehensive and systematic review of twenty-four frameworks designed to mitigate fraud related to CNP transactions. It brings together various architectural, analytical, methodological, and governance innovations that shape the current landscape of digital payment security. The study establishes a taxonomy, identifies key design considerations, and proposes performance evaluation indicators essential for developing adaptive, explainable, and regulation-compliant CNP detection frameworks.

The review identifies six interdependent design pillars that underpin the creation of resilient and trustworthy CNP frameworks: (1) scalable and modular architecture, (2) privacy-preserving data governance, (3) adaptive learning and drift management, (4) interpretability and trust, (5) cost-sensitive and latency-aware optimization, and (6) integrated governance with continuous evaluation.

A salient insight derived from this synthesis is the transformative potential of Graph Neural Networks (GNNs) as both the structural and analytical foundation for next-generation Credit Network Protection (CNP) frameworks. GNNs adeptly capture complex transactional and relational dependencies among entities, such as cards, merchants, devices, and accounts, thereby transitioning the focus from traditional feature-based analysis to a context-aware, community-level approach to fraud reasoning. The integration of GNNs with federated learning, blockchain auditing, and edge AI deployment significantly enhances the adaptability and transparency of these graph-based systems, enabling real-time anomaly detection with high precision. Empirical findings presented in this review indicate that frameworks incorporating these principles, such as SCARFF, Spade, FedGAT-DCNN, and OptDevNet, exhibit exceptional detection performance, with F1-scores ranging from 0.85 to 0.99 and area under the curve (AUC) values exceeding 0.93. Moreover, these frameworks maintain adherence to standards of compliance, interpretability, and ethical governance.

The synthesized evidence indicates that frameworks that are graph-aware, prioritize privacy protection, and possess the capacity for continuous self-improvement are pivotal in the future landscape of fraud prevention. Their intrinsic capability to learn from relational environments, adapt to the dynamic nature of fraud tactics, and provide transparent explanations for their decision-making processes renders them indispensable for advancing research, formulating policy, and implementing solutions in real-world contexts.

Recommendations

Building on the findings of this review, six strategic directions are recommended to guide future research and implementation:

1. Operationalize Graph Intelligence in Real-Time Pipelines [28,31]

CNP fraud often involves coordinated entities, fraud rings, mule networks, and synthetic identities that evade detection in traditional transaction-level models. Future systems should embed Graph Neural Networks (GNNs) and temporal-relational modeling (e.g., evolving graph construction, session-level topologies) as core architectural components, rather than as post-hoc layers. Building on frameworks like APATE and Spade, real-time graph pipelines should be deployed to detect suspicious linkages between accounts, devices, IPs, and merchants. Additionally, use community detection algorithms (e.g., Louvain or subgraph isomorphism) to flag novel fraud structures proactively. These capabilities should be supported by graph-aware explanations (e.g., motif highlights, node importance) to foster investigator trust and regulatory defensibility.

2. Deploy Federated Graph Learning and Cross-Border Privacy Protocols [18,24]

To overcome regulatory restrictions on data centralization, CNP detection frameworks should move beyond siloed learning by integrating Federated Graph Learning (Fed-GNN), as exemplified by FedGAT-DCNN. Institutions must implement secure aggregation, differential privacy ($\varepsilon < 1$), and blockchain-based audit trails to support collaborative training without compromising GDPR, PCI-DSS, or PSD2 compliance. Cross-border deployments should adopt standardized federated protocols (e.g., SplitFed, FedAvg+DP) and tokenized consent smart contracts to manage cross-institution workflows. National and regional regulators should be engaged to co-develop sandbox pilots that validate privacy-preserving multi-institution fraud detection models.

3. Integrate Explainable AI (XAI) at Analyst and Customer Touchpoints [11,29]

To reduce false positives and increase trust in AI decisions, fraud detection systems must incorporate explainability across the pipeline. This includes SHAP/LIME visualizations, graph heatmaps, temporal attention scores, and narrative-style crime-script overlays (e.g., from Bødker et al., 2023). Explainability should be embedded not just in audit reports but within operational dashboards, analyst review portals, and customer communication templates. Future systems should offer interactive alert explainers for analysts and GDPR Art. 22-compliant justifications for consumers facing declined or flagged transactions.

4. Build Lifecycle-Aware, Drift-Resilient Learning Systems [31,35]

Fraud tactics evolve rapidly, causing concept drift and degrading model accuracy over time. To counter this, future frameworks should adopt automated model lifecycle management with features such as sliding-window retraining, drift detectors (e.g., ADWIN, DDM), and adaptive thresholding. These should be combined with adversarial resilience mechanisms, such as poison-detection filters and robust feature selection. All updates should be versioned with model cards, policy snapshots, and rollback protocols to maintain transparency and operational continuity.

5. Establish a Shared CNP Evaluation and Benchmarking Consortium [29]

Fragmented datasets and inconsistent metrics hinder academic and industry progress. A global initiative, modeled after ImageNet or OpenML, should launch an open-access CNP dataset repository with anonymized ISO-8583-compliant transaction graphs. This repository should include realistic synthetic data (e.g., GAN-generated fraud sessions), labeled scenarios (e.g., synthetic IDs, bot attacks), and modular APIs for benchmarking. Evaluation protocols must cover accuracy, F1-score, latency, throughput, privacy leakage, auditability, and interpretability, enabling reproducible, deployment-ready research.

6. Codify Global Standards and Interoperable Detection Protocols [9,18]

CNP fraud crosses national and institutional boundaries. Future research should prioritize standards-compliant architecture design, including ISO 8583 extensions, interoperable API schemas, and ontology-aligned transaction representations. Stakeholders should collaborate with bodies like ISO TC 68, IEEE SA, and EMVCo to create guidelines for AI-ready fraud detection components, including security layers, explainability hooks, and risk signaling mechanisms. Case studies from global banks should feed into these standards to bridge innovation with compliance.

Limitations

This review has several limitations to consider:

First, it focuses on framework-level studies of Card-Not-Present (CNP) fraud from peer-reviewed, English-language sources, potentially overlooking insights from non-English literature and industry reports. Second, it prioritizes studies on architecture and frameworks, limiting the representation of standalone algorithms. Third, variability in datasets and evaluation methods hinders formal meta-analysis, constraining proposed benchmarks to the original studies' reporting practices. Fourth, evidence for large-scale deployment of technologies like graph neural networks and blockchain is limited, often relying on pilot projects that need further validation. Finally, rapid changes in CNP fraud patterns and technologies mean the findings may not include recent developments, so the proposed taxonomy and benchmarks should be viewed as evolving.

Acknowledgement: Fig. 1 was created using AI-generated content from ChatGPT (OpenAI). The authors confirm that no human likenesses, copyrighted elements, or misleading representations are included in the image.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: Conceptualization, literature review, discussion, writing—original draft preparation: Kwabena Owusu-Mensah, Edward Danso Ansong, Winfred Yaokumah, Kofi Sarpong Adu-Manu; writing—reviewing and editing: Kwabena Owusu-Mensah, Kofi Sarpong Adu-Manu; supervision: Edward Danso Ansong, Winfred Yaokumah, Kofi Sarpong Adu-Manu. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: No datasets were generated or analyzed during the current study.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

Glossary/abbreviations

Acronym	Full Term	Context/Usage
AI	Artificial Intelligence	Core technology for fraud detection and prediction
AUC	Area Under the Curve	Model evaluation metric (ROC/PR curves)
BA	Balance Accuracy	The average of Sensitivity (True Positive Rate) and Specificity (True Negative Rate)
CART	Classification and Regression Trees	A supervised machine learning algorithm that uses decision trees to predict outcomes by splitting data
CCFD	Credit Card Fraud Detection	General fraud detection systems
CCFDP	Credit Card Fraud Detection Prevention	Preventive frameworks for CNP fraud
CNN	Convolutional Neural Networks	Deep learning model for pattern recognition
CNP	Card-Not-Present	Fraud type in online/remote transactions
CP	Card-Present	Fraud type in physical transactions
CVV	Card Verification Value	Security code for online payments
DNN	Deep Neural Network	Advanced neural architectures
DP	Differential Privacy	Privacy-preserving data techniques
DT	Decision Trees	Traditional machine learning model
EMV	Europay, MasterCard, and Visa	Smart card payment standard
FL	Federated Learning	Distributed privacy-preserving ML
GANs	Generative Adversarial Networks	Synthetic data generation, anomaly detection
GB	Gradient Boosting	Ensemble ML method
GDPR	General Data Protection Regulation	EU regulation for data protection

(Continued)

(continued)

Acronym	Full Term	Context/Usage
GNN	Graph Neural Network	Framework for transaction network modeling
IoT	Internet of Things	Ecosystem linked to CNP fraud risks
kNN	k-Nearest Neighbors	ML classification algorithm
KYC	Know Your Customer	Regulatory compliance in banking
LIME	Local Interpretable Model-Agnostic Explanations	Explainable AI technique
LR	Logistic Regression	Baseline ML classifier
LSTM	Long Short-Term Memory	Sequence learning neural network
LSTM-AE	LSTM Autoencoder	Dimensionality reduction, anomaly detection
MFA	Multi-Factor Authentication	Security authentication method
ML	Machine Learning	Core fraud detection methodology
MTI	Message Type Indicator	a four-digit numeric field which indicates the overall function of the message
PCA	Principal Component Analysis	Dimensionality reduction technique
PCI DSS	Payment Card Industry Data Security Standard	Payment system compliance standard
PR-AUC	Precision–Recall Area Under Curve	Model evaluation metric
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses	Systematic review reporting guideline
PSD2	Payment Services Directive 2	EU directive on payment security
RF	Random Forests	Ensemble learning method
RNNs	Recurrent Neural Networks	Sequence modeling architectures
ROC	Receiver Operating Characteristic	Model evaluation metric
RQ	Research Question	Used in systematic review structuring
SCARFF	Scalable Real-Time Credit Card Fraud Finder	Big data fraud detection framework
SHAP	SHapley Additive Explanations	Explainable AI technique
SLA	Service Level Agreement	Operational governance layer
SLR	Systematic Literature Review	Methodology for structured reviews
STAGN	Spatial-Temporal Attention Graph Neural Network	Fraud detection using GNNs
SVDD	Support Vector Data Description	One-class anomaly detection
SVD	Singular Value Decomposition	Dimensionality reduction technique
SVM	Support Vector Machines	ML classification method
TPS	Transactions Per Second	Real-time system performance metric
t-SNE	t-Distributed Stochastic Neighbor Embedding	Data visualization/dimensionality reduction
UX	User Experience	Application-level usability measure
VDR	Value Detection Rate	Value of Detected Fraud Rate
XAI	Explainable Artificial Intelligence	Interpretable ML for fraud detection

Key Terms and Definitions

Term	Definition	Representative Source(s)
Adaptive Learning	Learning systems capable of continuously updating parameters as new data becomes available, allowing models to remain effective against emerging fraud patterns.	Prabha & Priscilla, 2024; Devi et al., 2024
Anomaly Detection	A family of techniques used to identify unusual patterns in data that deviate from historical behavior, signaling possible fraud or data compromise.	Prabha & Priscilla, 2024; Jeribi, 2024
API Interoperability	The ability of fraud detection systems to integrate through standardized API interfaces, ensuring cross-platform collaboration and data exchange.	Carcillo et al., 2018; Devi et al., 2024
Blockchain Auditability	The use of blockchain ledgers to record fraud alerts, model updates, and detection events in a tamper-proof, transparent manner for regulatory traceability.	Mackey et al., 2020; Baabdullah et al., 2024
Card-Not-Present (CNP) Fraud	A type of payment fraud that occurs when the cardholder does not physically present the card during a transaction. Common in e-commerce, online banking, and mobile payments, it exploits digital vulnerabilities in remote authentication processes.	Bødker et al., 2023; Razaque et al., 2024
Concept Drift	The phenomenon where the statistical distribution of data changes over time, causing predictive models to lose accuracy unless retrained. It arises from evolving fraud strategies, user behaviors, or market dynamics.	Mauliddiah & Suharjito, 2023; Jeribi, 2024
Cost-Sensitive Learning	An optimization approach in which misclassification penalties are weighted based on financial risk, minimizing false positives and maximizing economic efficiency.	Van Vlasselaer et al., 2015;
Data Drift	A shift in the input feature space or contextual data patterns that may not immediately affect output labels but gradually erodes model consistency. Detecting data drift through statistical monitoring or adaptive retraining is vital for sustained fraud detection accuracy.	Jeribi, 2024; Devi et al., 2024
Data Privacy and Governance	The policies and controls ensuring ethical and lawful data use in compliance with GDPR, PCI-DSS, and PSD2. Includes secure aggregation, anonymization, and auditable model tracking.	Baabdullah et al., 2024; Mackey et al., 2020

(Continued)

(continued)

Term	**Definition**	**Representative Source(s)**
Detection Latency	The elapsed time between transaction initiation and fraud decision output. Lower latency (<100 ms) improves customer experience and prevention efficacy.	Carcillo et al., 2018; Adil et al., 2024
Edge Inference	The deployment of lightweight fraud detection models on edge or IoT devices for low-latency, privacy-compliant scoring.	Adil et al., 2024; Jeribi, 2024
Explainable Artificial Intelligence (XAI)	A subfield of AI focused on making model outputs transparent and interpretable to humans. Techniques such as SHAP, LIME, and attention visualization help justify fraud decisions to analysts and regulators.	Bødker et al., 2023; Khan et al., 2024
Federated Learning (FL)	A decentralized training approach where multiple entities (e.g., banks) collaboratively train a shared model without exchanging raw data, thus preserving confidentiality and meeting regulatory obligations.	Baabdullah et al., 2024; Mackey et al., 2020
Fraud Detection Framework	A structured system of algorithms, models, and governance layers designed to detect, prevent, and mitigate fraudulent transactions using rule-based, statistical, or machine learning techniques.	Jeribi, 2024; Singh & Jain, 2019
Global Standards (ISO 8583+, IEEE)	International technical frameworks governing transaction data formats and AI ethics in financial ecosystems. Support cross-border standardization of fraud detection workflows.	ISO, 2023; IEEE P7000 Series, 2022
Graph Neural Networks (GNNs)	A neural network architecture that operates on graph data structures, learning entity relationships (e.g., card–device–merchant). GNNs enable the detection of collusive groups, mule networks, and synthetic identity clusters.	Van Vlasselaer et al., 2015; Jiang et al., 2022; Mauliddiah & Suharjito, 2023
Human-Centered Trust	The design of fraud detection systems with interpretable dashboards, feedback mechanisms, and user interfaces that promote confidence among analysts, regulators, and customers.	Bødker et al., 2023; Devi et al., 2024
Human-in-the-Loop (HITL)	A paradigm combining AI automation with human oversight, where analysts review or approve suspicious transactions based on model explanations.	Bødker et al., 2023; Devi et al., 2024

(Continued)

(continued)

Term	Definition	Representative Source(s)
Hybrid Models	Architectures that deliberately combine at least two of the following: rule-based mechanisms, statistical models, machine learning/deep learning, graph-based reasoning, federated learning, or blockchain/provenance components, integrated to balance adaptability, interpretability, compliance, and operational robustness	Singh & Jain, 2019; Jeribi, 2024
Microservices Architecture	A modular software design enabling independent deployment and scaling of fraud detection components, improving maintainability and interoperability.	Carcillo et al., 2018
Mule Account Detection	Identifying accounts used to receive and redistribute illicit funds through relational transaction analysis and graph-based community detection.	Van Vlasselaer et al., 2015; Jiang et al., 2022
Privacy-Preserving Machine Learning	A class of ML techniques (e.g., federated learning, differential privacy, homomorphic encryption) that protect individual data while enabling model training.	Baabdullah et al., 2024; Mackey et al., 2020
Regulatory Compliance (GDPR, PCI-DSS, PSD2)	The adherence of detection systems to legal frameworks that govern data privacy, security, and electronic payment processes.	Bødker et al., 2023; Mackey et al., 2020
Session Analysis	Monitoring user interactions (e.g., dwell time, page transitions, login frequency) to establish behavioral baselines for anomaly detection.	Patel et al., 2019; Jeribi, 2024
Streaming Fraud Detection	Real-time monitoring and scoring of transaction streams using event-driven architectures (e.g., Kafka, Spark). Essential for instant fraud blocking in high-throughput environments.	Carcillo et al., 2018; Jeribi, 2024
Synthetic Identity Fraud	The creation and use of partially fabricated or blended identities to perform unauthorized financial activities, evading traditional verification methods.	Kalisetty et al., 2024; Bødker et al., 2023
Temporal Validation	The evaluation of fraud models using temporally ordered data splits to prevent leakage and ensure time-consistent performance assessment.	Jeribi, 2024; Devi et al., 2024

References

1. Merchant Savvy. Payment fraud statistics, trends & forecasts 2024 [Internet]; 2024 [cited 2025 Dec 1]. Available from: https://www.merchantsavvy.co.uk/payment-fraud-statistics/.
2. Nilson Report. Global card fraud losses 2015–2025 [Internet]. [cited 2025 Dec 1]. Available from: https://www.statista.com/statistics/1394119/global-card-fraud-losses/.

3. Hayat K, Muhammad S, Khan M, Shah SA, Zameer A. A critical examination of credit card fraud detection research: how evaluation flaws distort results. Mathematics. 2025;13(16):2563. doi:10.3390/math13162563.
4. Abdel Latif Marazqah Btoush E, Zhou X, Gururajan R, Chan KC, Genrich R, Sankaran P. A systematic review of literature on credit card cyber fraud detection using machine and deep learning. Peerj Comput Sci. 2023;9(1):e1278. doi:10.7717/peerj-cs.1278.
5. Ngai EWT, Hu Y, Wong YH, Chen Y, Sun X. The application of data mining techniques in financial fraud detection: a classification framework and an academic review of literature. Decis Support Syst. 2011;50(3):559–69. doi:10.1016/j.dss.2010.08.006.
6. Bahnsen AC, Stojanovic A, Aouada D, Ottersten B. Cost sensitive credit card fraud detection using Bayes minimum risk. In: 2013 12th International Conference on Machine Learning and Applications; 2013 Dec 4–7; Miami, FL, USA; 2013. p. 333–8. doi:10.1109/ICMLA.2013.68.
7. Alamri M, Ykhlef M. Survey of credit card anomaly and fraud detection using sampling techniques. Electronics. 2022;11(23):4003. doi:10.3390/electronics11234003.
8. Rigaki M, Garcia S. A survey of privacy attacks in machine learning. ACM Comput Surv. 2024;56(4):1–34. doi:10.1145/3624010.
9. Carcillo F, Dal Pozzolo A, Le Borgne YA, Caelen O, Mazzer Y, Bontempi G. SCARFF: a scalable framework for streaming credit card fraud detection with spark. Inf Fusion. 2018;41:182–94. doi:10.1016/j.inffus.2017.09.005.
10. Li M, Walsh J. FedGAT-DCNN: advanced credit card fraud detection using federated learning, graph attention networks, and dilated convolutions. Electronics. 2024;13(16):3169. doi:10.3390/electronics13163169.
11. Bødker A, Connolly P, Sing O, Hutchins B, Townsley M, Drew J. Card-not-present fraud: using crime scripts to inform crime prevention initiatives. Secur J. 2023;36:693–711. doi:10.1057/s41284-022-00359-w.
12. Razaque A, Ben Haj Frej M, Bektemyssova G, Amsaad F, Almiani M, Alotaibi A, et al. Credit card-not-present fraud detection and prevention using big data analytics algorithms. Appl Sci. 2023;13(1):57. doi:10.3390/app13010057.
13. European Union Agency for Cybersecurity (ENISA). ENISA threat landscape 2024: Identity theft and account takeover [Internet]. Marousi, Greece: Publications Office of the European Union; 2024 [cited 2025 Dec 1]. Available from: https://www.enisa.europa.eu/publications/enisa-threat-landscape-2024.
14. Thomas K, Pullman J, Yeo K, Raghunathan A, Kelley PG, Invernizzi L, et al. Protecting accounts from credential stuffing with password breach alerting. In: Proceedings of the 28th USENIX Security Symposium (USENIX Security 19); 2019 Aug 14–16; Santa Clara, CA, USA. p. 1556–71.
15. Dal Pozzolo A, Boracchi G, Caelen O, Alippi C, Bontempi G. Credit card fraud detection: a realistic modeling and a novel learning strategy. IEEE Trans Neural Netw Learn Syst. 2017;29(8):3784–97. doi:10.1109/TNNLS.2017.2736643.
16. Yang Q, Liu Y, Chen T, Tong Y. Federated machine learning. ACM Trans Intell Syst Technol. 2019;10(2):1–19. doi:10.1145/3298981.
17. Baesens B, Van Vlasselaer V, Verbeke W. Fraud analytics using descriptive, predictive, and social network techniques: a guide to data science for fraud detection. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2015.
18. Mackey T, Miyachi K, Fung D, Qian S, Short J. Combating health care fraud and abuse: conceptualization and prototyping study of a blockchain antifraud framework. J Med Internet Res. 2020;22(9):e18623. Available from: https://www.jmir.org/2020/9/e18623.
19. Patel Y, Ouazzane K, Vassilev V, Li J. Remote banking fraud detection framework using sequence learners. J Internet Bank Commer. 2019;24(1):1–31.
20. Strelcenia E, Prakoonwit S. Generating synthetic data for credit card fraud detection using GANs. In: 2022 International Conference on Computers and Artificial Intelligence Technologies (CAIT); 2022 Nov 4–6; Quzhou, China. p. 42–7. doi:10.1109/CAIT56099.2022.10072179.
21. Cheng D, Wang X, Zhang Y, Zhang L. Graph neural network for fraud detection via spatial-temporal attention. IEEE Trans Knowl Data Eng. 2020;34(8):3800–13. doi:10.1109/TKDE.2020.3025588.
22. Lin J, Guo X, Zhu Y, Mitchell S, Altman E, Shun J. FraudGT: a simple, effective, and efficient graph transformer for financial fraud detection. In: Proceedings of the 5th ACM International Conference on AI in Finance; 2024 Nov 14–17; Brooklyn, NY, USA. p. 292–300. doi:10.1145/3677052.3698648.

23. Kalisetty S, Pandugula C, Sondinti LRK, Mallesham G, Rani PS. AI-driven fraud detection systems: enhancing security in card-based transactions using real-time analytics. J Electr Syst. 2024;20:1452–64.
24. Baabdullah T, Alzahrani A, Rawat DB, Liu C. Efficiency of federated learning and blockchain in preserving privacy and enhancing the performance of credit card fraud detection (CCFD) systems. Future Internet. 2024;16(6):196. doi:10.3390/fi16060196.
25. Xu Y, Jian X, Li T, Zou S, Li B. Blockchain-based authentication scheme with an adaptive multi-factor authentication strategy. Mob Inf Syst. 2023;2023:4764135. doi:10.1155/2023/4764135.
26. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA, 2020 statement: an updated guideline for reporting systematic reviews. PLoS Med. 2021;18(3):e1003583. doi:10.1371/journal.pmed.1003583.
27. Schardt C, Adams MB, Owens T, Keitz S, Fontelo P. Utilization of the PICO framework to improve searching PubMed for clinical questions. BMC Med Inform Decis Mak. 2007;7:16. doi:10.1186/1472-6947-7-16.
28. Van Vlasselaer V, Bravo C, Caelen O, Eliassi-Rad T, Akoglu L, Snoeck M, et al. APATE: a novel approach for automated credit card transaction fraud detection using network-based extensions. Decis Support Syst. 2015;75:38–48. doi:10.1016/j.dss.2015.04.013.
29. Manjula Devi C, Gobinath A, Padma Priya S, Adithiyaa M, Chandru MK, Jothi M. Next-generation anomaly detection framework leveraging artificial intelligence for proactive credit card fraud prevention and risk management. In: 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT); 2024 Jun 24–28; Kamand, India. p. 1–6. doi:10.1109/ICCCNT61001.2024.10725285.
30. Singh A, Jain A. A novel framework for credit card fraud prevention and detection (CCFPD) based on three layer verification strategy. In: Proceedings of ICETIT 2019. Cham, Switzerland: Springer International Publishing; 2019. p. 935–48. doi:10.1007/978-3-030-30577-2_83.
31. Mauliddiah N, Suharjito. Implementation graph database framework for credit card fraud detection. Procedia Comput Sci. 2023;227:326–35. doi:10.1016/j.procs.2023.10.531.
32. Olowookere TA, Adewale OS. A framework for detecting credit card fraud with cost-sensitive meta-learning ensemble approach. Sci Afr. 2020;8:e00464. doi:10.1016/j.sciaf.2020.e00464.
33. Prabha DP, Priscilla C. A combined framework based on LSTM autoencoder and XGBoost with adaptive threshold classification for credit card fraud detection. Sci Temper. 2024;15:2216–24. doi:10.58414/scientifictemper.2024.15.2.34.
34. Mniai A, Tarik M, Jebari K. A novel framework for credit card fraud detection. IEEE Access. 2023;11:112776–86. doi:10.1109/access.2023.3323842.
35. Jeribi F. A comprehensive machine learning framework for anomaly detection in credit card transactions. Int J Adv Comput Sci Appl. 2024;15(6):871. doi:10.14569/ijacsa.2024.0150688.
36. Adil M, Zhang Y, Jamjoom MM, Ullah Z. OptDevNet: a optimized deep event-based network framework for credit card fraud detection. IEEE Access. 2024;12:132421–33. doi:10.1109/access.2024.3458944.
37. Chen CT, Lee C, Huang SH, Peng WC. Credit card fraud detection via intelligent sampling and self-supervised learning. ACM Trans Intell Syst Technol. 2024;15(2):1–29. doi:10.1145/3641283.
38. Nijwala DS, Maurya S, Thapliyal MP, Verma R. Extreme gradient boost classifier based credit card fraud detection model. In: 2023 International Conference on Device Intelligence, Computing and Communication Technologies, (DICCT); 2023 Mar 17–18; Dehradun, India. p. 500–4. doi:10.1109/DICCT56244.2023.10110188.
39. Patil S, Nemade V, Soni PK. Predictive modelling for credit card fraud detection using data analytics. Procedia Comput Sci. 2018;132:385–95. doi:10.1016/j.procs.2018.05.199.
40. Thennakoon A, Bhagyani C, Premadasa S, Mihiranga S, Kuruwitaarachchi N. Real-time credit card fraud detection using machine learning. In: 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence); 2019 Jan 10–11; Noida, India. p. 488–93. doi:10.1109/confluence.2019.8776942.
41. Jiang J, Li Y, He B, Hooi B, Chen J, Kok JKZ. Spade: a real-time fraud detection framework on evolving graphs. Proc VLDB Endow. 2022;16(3):461–9. doi:10.14778/3570690.3570696.

42. Cherif A, Alshehri S, Kalkatawi M, Imine A. Towards an intelligent adaptive security framework for preventing and detecting credit card fraud. In: 2022 IEEE/ACS 19th International Conference on Computer Systems and Applications (AICCSA); 2022 Dec 5–8; Abu Dhabi, United Arab Emirates. p. 1–8. doi:10.1109/AICCSA56895.2022.10017814.

43. Ribeiro MT, Singh S, Guestrin C. Why should I trust you?: explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016 Aug 13–17; San Francisco, CA, USA. p. 1135–44. doi:10.1145/2939672.2939778.