Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 09/08/2026 - **Accepted:** 10/09/2026
Vol: 01.04
DOI: 10.66104/t68phf90
Pages: 1-37

OPERATIONALLY RELIABLE REAL-TIME FINANCIAL FRAUD DETECTION: DEEP LEARNING, DATA DRIFT, EXPLAINABILITY, AND MODEL GOVERNANCE

Matheus Leonardo Rodrigues Almeida
Faculdade Presidente Antônio Carlos, Brazil
Corresponding author: matheusleonardo951@gmail.com
Orcid ID: https://orcid.org/0009-0009-4407-9287

Abstract
Deep learning expands the capacity of financial institutions to detect transactional, temporal, and relational patterns, but benchmark discrimination alone does not establish safe or effective production performance. This structured integrative narrative review treats real-time fraud detection as a sociotechnical decision system. Forty analytical and contextual sources are synthesized across data and labels, architecture and representation, rare-event evaluation, latency, drift, explainability, privacy, customer impact, and governance. The review formally proposes the Lifecycle Operational Assurance Framework for Real-Time Fraud Detection (LOAF-RTFD). Its central thesis is that a fraud model should be evaluated not by benchmark discrimination alone, but by risk-adjusted system utility under realistic latency, capacity, fairness, and governance constraints. No architecture dominates across fraud typologies or payment environments. Reliable operation requires point-in-time-correct data, calibrated scores, capacity-aware thresholds, layered interventions, customer safeguards, drift monitoring, recovery, and controlled change. Updated analysis of Pix links detection to transaction limits, cautious blocking, fraud markings, inter-participant notification, the Special Return Mechanism, and traceable recovery. The framework is a proposed governance and reporting synthesis, not an empirically validated operational standard. Because one author performed selection, appraisal, and interpretive coding, no inter-rater reliability is claimed.
Keywords: financial fraud; deep learning; real-time detection; graph neural networks; concept drift; explainable artificial intelligence; Pix; model governance

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 09/08/2026 - **Accepted:** 10/09/2026
Vol: 01.04
DOI: 10.66104/t68phf90
Pages: 1-37

1. Introduction

Digital payments reduce transaction friction but also expand the scale and speed at which fraudulent activity can be attempted. Payment fraud includes unauthorized transactions, account takeover, identity misuse, merchant fraud, social-engineering-enabled transfers, and coordinated networks that distribute activity across accounts and devices. The European Banking Authority and European Central Bank reported that strong customer authentication remained effective against several forms of fraud, while fraudsters increasingly adapted through manipulation of legitimate users and other techniques outside simple credential theft (European Banking Authority & European Central Bank [EBA & ECB], 2024).

Financial institutions have long used expert rules, statistical models, manual review, and shared negative lists. These mechanisms remain valuable because they encode policy and known patterns, but static rules may become brittle when behavior changes and can generate large alert volumes. Machine learning adds the ability to learn interactions from historical cases, whereas deep learning can represent nonlinear, sequential, and relational structures with less dependence on manually specified combinations. The operational objective, however, is not to replace every rule or analyst; it is to allocate prevention and investigation capacity more effectively (Abdallah et al., 2016; Hilal et al., 2022).

The distinction between experimental performance and operational reliability is central. A model may obtain a high area under the curve on a static dataset yet fail when transaction patterns shift, labels arrive late, fraud prevalence changes, or latency exceeds the authorization budget. False positives can block legitimate activity and impose disproportionate burdens on particular groups, while false negatives create direct losses and secondary harms. The relevant unit of analysis is therefore the complete decision pipeline, including data collection, feature computation, scoring, thresholds, human review, customer communication, appeals, and feedback (NIST, 2023; Sculley et al., 2015).

This article asks how neural methods can support real-time fraud detection without allowing benchmark metrics to obscure deployment constraints. It proposes the Lifecycle Operational Assurance Framework for Real-Time Fraud Detection (LOAF-RTFD) and foregrounds the following principle: “A fraud model should be evaluated not by benchmark discrimination alone, but by risk-adjusted system utility under realistic latency, capacity, fairness, and governance constraints.” The objectives are to compare architecture families with explicit evidence labels, evaluate data and metric limitations, distinguish operational latency regimes, analyze drift and privacy-preserving collaboration, expand the Brazilian Pix case, and connect model risk governance with customer protection.

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 09/08/2026 - **Accepted:** 10/09/2026
Vol: 01.04
DOI: 10.66104/t68phf90
Pages: 1-37

2. Literature Review

2.1 Fraud Detection as Rare-Event Decision-Making

Fraud detection differs from ordinary balanced classification because harmful events are uncommon, heterogeneous, strategically concealed, and incompletely labeled. A trivial classifier can achieve impressive accuracy by predicting that every transaction is legitimate. For this reason, precision, recall, F1-score, precision-recall curves, calibration, alert volume, expected monetary loss, and investigation capacity are often more informative than accuracy alone. The appropriate metric depends on whether the operational priority is preventing loss, avoiding customer disruption, finding networks, or supporting investigation (Dal Pozzolo et al., 2018; Saito & Rehmsmeier, 2015).

Labels are not neutral ground truth. Some fraudulent transactions are never reported, chargeback outcomes arrive after substantial delay, investigators select which alerts receive scrutiny, and successful prevention can remove the very outcome that would confirm a prediction. Training data therefore reflect past policies and observation processes. If a model learns only from investigated alerts, it may reproduce earlier selection biases and remain blind to behaviors that prior rules did not surface (Baesens et al., 2015; Hand, 2006).

Cost asymmetry further complicates threshold selection. Missing a high-value coordinated scheme may be more serious than approving a low-value suspicious payment, but blocking thousands of legitimate payments can damage trust and exclude customers. Expected-value models must include direct losses, reimbursement, investigation expense, customer attrition, regulatory obligations, and operational capacity. A single global threshold may be inferior to differentiated policies by product, channel, risk tier, and available verification step (Baesens et al., 2015; EBA & ECB, 2024).

2.2 Neural Representation Families

Multilayer perceptrons are a practical baseline for structured transaction features such as amount, time, merchant category, device attributes, and historical aggregates. They can model nonlinear interactions but do not inherently encode order or relationships. Their value therefore depends heavily on feature pipelines that summarize recent behavior. Gradient-boosted trees frequently remain competitive on tabular data, so neural superiority should be demonstrated against strong non-neural baselines rather than assumed (Borisov et al., 2022; Hilal et al., 2022).

Recurrent neural networks, including long short-term memory and gated recurrent units, represent sequences of customer or account activity. They can identify changes in timing, amount, location, and merchant progression that are difficult to express through isolated transactions. Their limitations include sequential

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 09/08/2026 - **Accepted:** 10/09/2026
Vol: 01.04
DOI: 10.66104/t68phf90
Pages: 1-37

computation, sensitivity to sequence construction, and the risk that long histories increase latency without adding stable signal. Temporal convolution and attention provide alternative sequence representations (Hochreiter & Schmidhuber, 1997; Jurgovsky et al., 2018).

Autoencoders learn compressed representations and flag observations with unusually high reconstruction error. They are attractive when confirmed fraud labels are scarce, but anomaly is not equivalent to fraud. Legitimate travel, major purchases, new devices, seasonal behavior, and changes in income can all look unusual. Autoencoder scores are therefore more defensible as one signal within an ensemble or review process than as autonomous proof of wrongdoing (Pang et al., 2021).

Graph neural networks model accounts, cards, merchants, devices, addresses, and transfers as nodes connected by relationships. This structure is particularly relevant to coordinated fraud, mule networks, synthetic identities, and laundering patterns that may appear ordinary when transactions are assessed separately. Graph methods can propagate contextual information across neighborhoods, but performance depends on graph construction, time windows, edge definitions, sampling, and safeguards against information leakage (Dou et al., 2020; Weber et al., 2019).

Transformers use attention to represent relationships across sequences or sets and can combine heterogeneous event histories. They offer parallel training and flexible context modeling, but long histories and large embeddings increase computational cost. Transformer-based results should be interpreted in relation to latency, memory, data volume, and baseline strength. In many payment systems, a smaller calibrated model with stable features may deliver greater operational value than a large model whose marginal gain is difficult to reproduce (Vaswani et al., 2017; Sculley et al., 2015).

Architecture	Best-suited signal	Operational advantage	Principal limitation	Evidence basis
MLP	Engineered transaction and profile features	Fast inference and simple serving	Depends on feature engineering; ignores structure unless encoded.	Direct fraud + methodological evidence

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 09/08/2026 - **Accepted:** 10/09/2026
Vol: 01.04
DOI: 10.66104/t68phf90
Pages: 1-37

Architecture	Best-suited signal	Operational advantage	Principal limitation	Evidence basis
LSTM/GRU	Ordered account or card histories	Captures temporal dependence	Sequence construction and latency can be costly.	Direct fraud + adjacent sequence evidence
Autoencoder	Deviation from learned legitimate patterns	Can operate with limited fraud labels	Anomaly is not necessarily fraud.	Adjacent anomaly-detection evidence
GNN	Shared devices, accounts, merchants, and transfer networks	Represents coordinated behavior	Graph freshness, leakage, and sampling are difficult.	Direct fraud evidence
Transformer	Long heterogeneous sequences and contextual events	Flexible attention-based representation	Compute, memory, calibration, and explainability burdens.	Adjacent ML + operational inference
Hybrid ensemble	Complementary transactional, temporal, and relational signals	Improves resilience across typologies	Integration and governance complexity.	Author synthesis across heterogeneous evidence

2.3 Real-Time Systems and the Meaning of Latency

Real time should be defined by the decision deadline of the use case. Card authorization may require a response within a narrow service-level objective, whereas anti-money-laundering investigation can tolerate longer processing. End-to-end latency includes event transport, feature retrieval, graph lookup, model inference, policy execution, logging, and network overhead. Reporting model

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 09/08/2026 - **Accepted:** 10/09/2026
Vol: 01.04
DOI: 10.66104/t68phf90
Pages: 1-37

inference alone can therefore misrepresent whether the complete system meets its deadline (Kreps et al., 2011; Sculley et al., 2015).

Latency reporting should separate event transport, feature retrieval, graph lookup, inference, policy execution, logging, and network overhead. Median latency alone is insufficient; tail behavior, timeouts, degraded-mode behavior, and the end-to-end p95 and p99 are material to authorization safety.

Regime	Decision horizon	Permissible action	Evidence to report
Authorization-time blocking	Milliseconds to a few seconds	Approve, decline, hold	End-to-end p50/p95/p99, availability, timeout fallback
Step-up authentication	Seconds to minutes	Challenge or confirm	Completion, abandonment, accessibility, added latency
Near-real-time monitoring	Minutes	Warn, suspend, escalate	Detection-to-action time and queue capacity
Post-transaction recovery	Minutes to hours	Trace, block, notify, recover	Notification and recovery times; funds located
Batch investigation	Hours to days	Prioritize cases and networks	Backlog, investigator throughput, case yield
AML extended review	Risk-based and variable	Escalate suspicious behavior	Case quality, timeliness, documentation

Streaming systems must preserve consistency between training and serving. A feature computed differently offline and online can produce training-serving skew, while delayed events can make real-time aggregates incomplete. Feature stores, event-time processing, versioned transformations, and point-in-time-correct training

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 09/08/2026 - **Accepted:** 10/09/2026
Vol: 01.04
DOI: 10.66104/t68phf90
Pages: 1-37

sets reduce these risks. Reliability engineering is not secondary to model design because unavailable or stale features can change predictions more than a small architectural improvement (Sculley et al., 2015).

3. Methodology / Materials and Methods

3.1 Review design and purpose

This study uses a structured integrative narrative review with single-author interpretive thematic synthesis. Evidence spans machine-learning experiments, streaming systems, methodological studies, surveys, books, and regulatory guidance. The design does not claim exhaustive retrieval, protocol registration, statistical representativeness, or pooled effect estimation. It supports framework development by connecting model evidence with point-in-time correctness, latency, class imbalance, calibration, drift, explainability, customer burden, and accountable governance. The resulting LOAF-RTFD is proposed as a transparent reporting and decision structure rather than a validated technical standard.

3.2 Information sources and final search date

The final search update was conducted on September 6, 2026 through IEEE Xplore, ACM Digital Library, Google Scholar, Crossref-linked publisher platforms, Scopus-indexed discovery, and official repositories of Banco Central do Brasil, EBA, ECB, NIST, and the U.S. Federal Reserve. Backward citation tracing retained foundational works, while focused updating prioritized 2020-2026 evidence on graph fraud models, online and delayed-label learning, calibration, drift, financial model risk, customer impact, and instant-payment controls. Discovery was purposive and iterative rather than a registered systematic search.

3.3 Search strategy

Platform-adapted searches used title, abstract, keywords, or full metadata where available. Core query families were: (financial fraud OR payment fraud OR credit card fraud) AND (deep learning OR neural network OR transformer OR autoencoder OR graph neural network); fraud detection AND (real time OR streaming OR latency OR concept drift OR delayed labels); and fraud AND (calibration OR precision recall OR explainability OR fairness OR governance). Focused queries addressed Pix, the Special Return Mechanism, cautious blocking, fraud markings, privacy-preserving collaboration, independent validation, stress testing, and rollback. Appendix A records the platform, fields, filters, representative syntax, and search purpose. Exact historical retrieval totals were not preserved and are not reconstructed.

3.4 Eligibility and exclusion criteria

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 09/08/2026 - **Accepted:** 10/09/2026
Vol: 01.04
DOI: 10.66104/t68phf90
Pages: 1-37

Eligible sources contributed directly to at least one review domain through empirical evidence, a survey or evidence synthesis, a foundational technical or methodological contribution, a documented benchmark or dataset, a scholarly treatment of operational risk, or authoritative regulatory and governance guidance. Priority was given to peer-reviewed research and primary institutional documents. Sources were excluded when they consisted mainly of promotional claims, anonymous comparisons, unverifiable methods, duplicate propositions, or demonstrations that reported only aggregate accuracy without enough information to interpret their relevance. Institutional reports were used within their regulatory or contextual scope rather than treated as independent evidence that a particular model is effective.

3.5 Selection process and corpus composition

The sole author conducted relevance screening, full-text eligibility assessment, extraction, appraisal, interpretive coding, and synthesis. The revised corpus contains 40 sources, separating empirical studies, surveys, books, methodological work, benchmark or dataset papers, conference papers, and institutional or regulatory materials. Appendix B reports source characteristics and evidence roles. The flow record in Appendix A identifies the stages and transparently marks unavailable retrospective counts rather than presenting invented PRISMA-style numbers.

3.6 Data extraction and thematic coding

For each retained source, the author recorded publication type, fraud context, dataset type, observation period when reported, prevalence, label definition or delay, model family, temporal split, comparator, metrics, calibration, threshold analysis, latency, drift evaluation, explainability, operational relevance, and principal limitation. Evidence was coded into fraud typology; data and labels; architecture and representation; metrics and validation; production infrastructure; and governance. Architecture statements were labeled as direct fraud evidence, adjacent machine-learning evidence, methodological inference, institutional guidance, operational assumption, or author synthesis.

3.7 Critical appraisal and synthesis rules

Evidence was appraised source by source according to temporal realism, point-in-time correctness, leakage controls, representativeness, label quality, prevalence, comparator strength, uncertainty, calibration, threshold analysis, external or prospective validation, reproducibility, latency, drift, explainability, and correspondence with an operational decision. Appendix C makes these judgments auditable. “Not reported” is distinguished from “not applicable.” Random splits and

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 09/08/2026 - **Accepted:** 10/09/2026
Vol: 01.04
DOI: 10.66104/t68phf90
Pages: 1-37

aging public datasets receive cautious weight, accuracy alone is inadequate for rare events, and regulatory documents establish obligations or context rather than comparative model effectiveness.

3.8 Single-author procedure and methodological limitations

All review stages were completed by one author. No second reviewer independently screened, coded, or appraised a subset; therefore Cohen's kappa, Krippendorff's alpha, and percentage agreement are not reported. The thematic classifications and appraisal ratings are explicitly single-author interpretive judgments and are not reliability-tested. Search documentation, evidence-role labels, source-characteristics reporting, and source-level appraisal improve auditability but cannot remove confirmation bias. The article consequently limits its claims to a structured narrative synthesis and proposed framework.

4. Results

4.1 Benchmark Data and External Validity

Public datasets enable replication but often represent narrow and aging slices of fraud. The widely used European credit-card dataset contains 284,807 transactions and 492 labeled frauds, a prevalence of approximately 0.172%. Its anonymized principal components protect confidentiality but limit substantive interpretation and transfer to other products. Results obtained on this dataset remain useful for method comparison, yet they cannot establish readiness for current instant-payment, identity, or account-takeover environments (Dal Pozzolo et al., 2018).

Synthetic datasets permit controlled experiments and public release but reproduce assumptions embedded in the simulator. PaySim, for example, models mobile-money transactions from patterns derived from a specific context. A system can learn artifacts of the generator rather than adaptive human behavior. Researchers should therefore disclose whether data are real, synthetic, or semi-synthetic and avoid presenting performance on simulated distributions as equivalent to independent production validation (Lopez-Rojas et al., 2016).

Random train-test splits create a serious risk of temporal leakage. Near-duplicate behavior, linked entities, or later information may appear in both sets, producing optimistic estimates. Time-based evaluation better approximates deployment: a model is trained on earlier events and tested on a later period. Multiple rolling windows can reveal degradation and variance that a single holdout conceals (Žliobaitė et al., 2016).

4.2 Class Imbalance, Calibration, and Decision Cost

Resampling methods change the effective class distribution and can improve learning, but they also alter the meaning of raw predicted probabilities.

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 09/08/2026 - **Accepted:** 10/09/2026
Vol: 01.04
DOI: 10.66104/t68phf90
Pages: 1-37

Oversampling, undersampling, synthetic minority generation, focal loss, and class-weighted objectives should be evaluated on an untouched test distribution. Otherwise, apparent precision may depend on an artificial prevalence that does not exist in production (Chawla et al., 2002; Saito & Rehmsmeier, 2015).

Calibration asks whether events assigned a risk of 0.10 occur approximately ten percent of the time among comparable cases. Well-ranked but poorly calibrated scores complicate threshold selection, staffing forecasts, and expected-loss analysis. Calibration must be rechecked after resampling, model updates, and prevalence shifts. Reliability diagrams, Brier scores, and segment-level checks complement ranking metrics (Guo et al., 2017).

Precision-recall evaluation is especially important when fraud is rare because it focuses on positive predictions. Receiver operating characteristic curves can appear favorable even when false-positive counts are operationally overwhelming. Reports should include the confusion matrix at realistic thresholds, fraud prevalence, alerts per thousand transactions, value-weighted outcomes, and confidence intervals (Saito & Rehmsmeier, 2015).

Evaluation question	Necessary evidence	Failure if omitted
Does the model rank risk?	PR-AUC and ROC-AUC with uncertainty	A single score hides threshold behavior.
Can operations use the alerts?	Precision, alerts per period, review capacity	A statistically good model can overload investigators.
Does it prevent material loss?	Value-weighted recall and expected cost	Transaction counts treat very different losses as equal.
Will it generalize over time?	Temporal and rolling-window validation	Random splits conceal drift and leakage.
Are scores decision-ready?	Calibration by time and segment	Thresholds become unstable or misleading.
Who bears errors?	False-positive/negative analysis across relevant groups	Aggregate performance conceals unequal burden.

4.3 Concept Drift and Adversarial Adaptation

Fraud distributions change because customers adopt new products, merchants alter behavior, institutions revise controls, seasons affect spending, and adversaries respond to detection. Covariate drift changes input patterns, prior-probability shift

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 09/08/2026 - **Accepted:** 10/09/2026
Vol: 01.04
DOI: 10.66104/t68phf90
Pages: 1-37

changes prevalence, and concept drift changes the relationship between behavior and fraud. Monitoring only overall score distributions is insufficient; systems need performance estimates when delayed labels arrive and proxy indicators before labels mature (Gama et al., 2014; Žliobaitė et al., 2016).

Feedback creates additional complexity. Blocking a transaction prevents observation of what would have happened, while step-up authentication changes customer behavior. Analysts preferentially review high-scoring cases, so confirmed labels are not an unbiased sample. Champion-challenger testing, controlled exploration where lawful and safe, counterfactual evaluation, and careful logging can reduce the resulting blind spots (Sculley et al., 2015).

Adversarial risk includes evasion rather than only mathematically crafted perturbations. Fraudsters can distribute activity, imitate legitimate rhythms, rotate devices, or exploit social engineering so that a transfer is technically authorized by the victim. Robustness testing should therefore use scenario-based red teams, graph manipulation, feature availability failures, and policy circumvention alongside conventional adversarial examples (EBA & ECB, 2024; NIST, 2023).

4.4 Explainability, Fairness, and Human Oversight

Explanations serve different audiences. Investigators need actionable reasons and related evidence; customers need understandable communication and a path to contest errors; validators need global behavior, sensitivity, stability, and documentation; regulators may require evidence of lawful and controlled decision-making. A single feature-attribution chart cannot satisfy every purpose (Lundberg & Lee, 2017; NIST, 2023).

Post hoc methods such as SHAP can summarize feature contributions, but an attribution is not a causal explanation and may be unstable when features are correlated. Explanations should be tested for fidelity, consistency, usefulness, and disclosure risk. Revealing exact thresholds or sensitive features can help adversaries, creating a legitimate tension between transparency and security (Lundberg & Lee, 2017; NIST, 2023).

Fairness analysis must reflect the operational harm being studied. Higher false-positive rates can interrupt access to funds, while lower recall can expose groups to greater victimization. Protected attributes may be unavailable, but location, device, income-related behavior, and access patterns can operate as proxies. Institutions should examine error rates, appeal outcomes, and customer friction across legally and ethically relevant segments without assuming that removal of a sensitive field removes bias (Mehrabi et al., 2021; NIST, 2023).

4.5 Privacy-Preserving Collaboration

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 09/08/2026 - **Accepted:** 10/09/2026
Vol: 01.04
DOI: 10.66104/t68phf90
Pages: 1-37

Cross-institutional information can reveal fraud networks that no organization can observe alone, yet financial and identity data are highly sensitive. Federated learning keeps raw data within participating institutions while sharing model updates, but it does not automatically provide privacy. Updates can leak information, participants may have incompatible distributions, and malicious clients can poison training. Secure aggregation, differential privacy, governance agreements, and auditability remain necessary (Kairouz et al., 2021).

Differential privacy provides a formal bound on the influence of individual records, but added noise can reduce detection of rare events. The privacy-utility trade-off is particularly difficult in fraud because the minority class is both important and small. Privacy should therefore be designed across retention, access control, minimization, encryption, purpose limitation, and model outputs, rather than delegated to one training technique (Dwork & Roth, 2014; NIST, 2023).

Decision dimension	Required evidence before deployment
Use case	What cross-institutional fraud signal cannot be obtained locally?
Participants	Which institutions contribute and who operates coordination?
Heterogeneity	Are populations, labels, features, and prevalence comparable?
Legal basis	What lawful purpose, notices, retention, and rights apply?
Secure aggregation	How are updates protected from other parties and the coordinator?
Poisoning controls	How are anomalous or malicious updates detected and contained?
Privacy budget	What privacy accounting and cumulative exposure are accepted?
Utility loss	How does protection affect rare-event recall and calibration?
Auditability	How are code, data lineage, rounds, and decisions recorded?

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 09/08/2026 - **Accepted:** 10/09/2026
Vol: 01.04
DOI: 10.66104/t68phf90
Pages: 1-37

Decision dimension	Required evidence before deployment
Exit and responsibility	How can a participant withdraw and who answers for failures?

4.6 Pix and the Brazilian Instant-Payment Context

Pix provides continuous instant transfers under rules established by Banco Central do Brasil. Operational assurance must combine behavioral and network scoring with participant risk management, device and account controls, user-configurable and institution-defined transaction limits, differentiated safeguards for unusual activity, customer notification, and response procedures. The receiving institution may apply cautious blocking for up to 72 hours when fraud is suspected while it performs a more robust analysis and informs the recipient. These controls create decision points with different deadlines and evidentiary requirements (Banco Central do Brasil [BCB], 2021, 2026a).

Pix fraud should not be modeled as one homogeneous label. Unauthorized access, manipulation of a legitimate payer, fraudulent recipients, account rental, device compromise, coercion, and coordinated movement of funds generate different signals and intervention opportunities. Transaction-only models may miss the behavioral process preceding an authorized transfer, whereas graph analysis may identify linked receiving accounts, devices, keys, and subsequent transfers. The DICT notification-of-infringement and fraud-marking processes allow founded suspicion to be communicated and associated with the relevant user and, where applicable, the Pix key (BCB, 2025a, 2026a).

The Brazilian case demonstrates why prediction must connect to coordination and recovery. The payer's provider receives the customer report and can initiate the appropriate procedure; the recipient's provider evaluates the receiving account and can block resources; and DICT records notifications and fraud markings. The improved Special Return Mechanism and Recovery of Values process can trace suspicious resources beyond the original destination into subsequent transactions, increasing opportunities for blocking and return. Digital self-service contestation reduces reporting delay, but automation must preserve customer information, auditable reasons, proportionality, and correction of false suspicion (BCB, 2025b, 2025c, 2026a).

Stage	Operational mechanism	Primary responsibility
Prevention	Limits, device/account risk, authentication,	Payer and recipient PSPs

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 09/08/2026 - **Accepted:** 10/09/2026
Vol: 01.04
DOI: 10.66104/t68phf90
Pages: 1-37

Stage	Operational mechanism	Primary responsibility
	behavioral and network signals	
Detection and communication	Notification of infringement and governed risk indicators	Participant identifying suspicion
Containment	Cautious blocking for up to 72 hours and recipient notification	Recipient PSP
Fraud marking	Accepted notice or fraud identified in own customer; record in DICT	Responsible Pix participant
Contest and recovery	Digital contestation, MED and Recovery of Values	Payer PSP, recipient PSP and subsequent PSPs
Governance	Timelines, evidence, customer communication, audit and correction	All participants under BCB rules

4.7 Lifecycle Operational Assurance Framework for Real-Time Fraud Detection

The Lifecycle Operational Assurance Framework for Real-Time Fraud Detection (LOAF-RTFD) organizes assurance across design, data, validation, deployment, monitoring, change, and retirement. Governance begins with a defined use case, accountable owner, fraud typology, decision deadline, acceptable harms, and fallback procedure. Data documentation records provenance, label construction, missingness, temporal coverage, proxies, and blind spots. Validation reproduces the production feature pipeline and includes temporal, stress, calibration, cost, fairness, customer-impact, and latency tests (NIST, 2023; Sculley et al., 2015).

Theoretical foundation	Mapping to LOAF-RTFD
Financial model risk management	Independent challenge, documentation, change control, monitoring and accountability

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 09/08/2026 - **Accepted:** 10/09/2026
Vol: 01.04
DOI: 10.66104/t68phf90
Pages: 1-37

Theoretical foundation	Mapping to LOAF-RTFD
Sociotechnical systems	Model, people, processes, customers, institutions and infrastructure as one decision system
Rare-event decision theory	Prevalence, asymmetric cost, capacity-aware thresholds and calibrated utility
Human oversight	Defined escalation, evidence, authority, appeal and reviewer-performance evaluation
Consumer protection	Proportionality, accessibility, notification, contestation, resolution and exclusion risk

Lifecycle stage	Required control	Illustrative evidence
Design	Define decision, deadline, affected parties, risk appetite, and fallback.	Approved use-case statement and responsibility map.
Data	Document provenance, labels, retention, leakage, and representativeness.	Data sheet and point-in-time correctness tests.
Validation	Test temporal generalization, calibration, cost, fairness, and latency.	Independent report with reproducible thresholds.
Deployment	Version model, features, policy, and dependencies.	Release record, rollback plan, and service-level tests.
Monitoring	Track drift, alert burden, delayed outcomes, segments, and incidents.	Dashboards with escalation criteria and owners.

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 09/08/2026 - **Accepted:** 10/09/2026
Vol: 01.04
DOI: 10.66104/t68phf90
Pages: 1-37

Lifecycle stage	Required control	Illustrative evidence
Change/retirement	Revalidate material changes and remove unsafe or obsolete models.	Change log, challenger evidence, and retirement record.
Customer impact	Measure burden, accessibility, contestation and resolution.	Segment metrics, challenge completion, complaints, reversals, resolution time.

Production monitoring should connect technical and business signals. Input drift, missing features, score distributions, latency, and service failures provide early warnings; confirmed fraud, value recovered, false-positive complaints, review outcomes, and appeal reversals provide delayed outcome evidence. Alerts require explicit thresholds, owners, and response times. Monitoring without an action protocol merely records deterioration (NIST, 2023; Žliobaitė et al., 2016).

4.8 Layered Decisions, Customer Friction, and Recovery

A fraud score does not determine one inevitable action. Institutions can approve, decline, delay, request additional authentication, reduce a limit, route an event to review, warn the customer, monitor subsequent activity, or investigate connected entities. This action space permits differentiated responses to uncertainty. A medium-risk payment may justify proportionate verification rather than automatic rejection, while a high-confidence network signal may justify immediate containment and specialist review. Policy design should connect each score range to an action, expected benefit, customer cost, legal basis, and escalation path (Baesens et al., 2015; EBA & ECB, 2024).

Customer friction is both a cost and a source of information. An authentication challenge can prevent loss and generate evidence about device possession or user intent, but excessive challenges encourage abandonment and may disproportionately affect people with limited connectivity, older devices, disabilities, or less familiarity with digital services. Friction metrics should therefore be analyzed by channel and relevant customer segment alongside fraud outcomes. Optimizing only the fraud rate can shift costs onto legitimate users without making the complete service safer (EBA & ECB, 2024; NIST, 2023).

Customer-outcome domain	Required metric
False-positive burden	False-positive rate and precision by relevant segment

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 09/08/2026 - **Accepted:** 10/09/2026
Vol: 01.04
DOI: 10.66104/t68phf90
Pages: 1-37

Customer-outcome domain	Required metric
Challenge burden	Challenge rate, completion, abandonment and added time
Correction	Appeal volume, reversal rate and time to resolution
Service quality	Complaints, repeat contacts and time to unblock
Accessibility	Failures by channel, device capability and accessibility need
Exclusion risk	Abandoned payments, account restrictions and downstream loss of access

Recovery mechanisms remain necessary because prediction is imperfect. Rapid reporting, transaction tracing, account containment, inter-institutional communication, evidence preservation, and customer support can reduce harm after an event. Models can prioritize recovery activity by identifying linked recipients or rapid fund movement, but automated suspicion should not be treated as conclusive evidence. Clear procedures, auditable records, and human authorization are particularly important when interventions affect several accounts (BCB, 2025; NIST, 2023).

This layered perspective changes the economic objective. The value of a model is not simply the amount associated with detected transactions. Institutions must estimate losses actually prevented, displacement to other channels, reimbursement, authentication expense, manual-review cost, customer attrition, complaints, and recovery. Controlled policy experiments and carefully matched comparisons can provide stronger evidence than multiplying recall by transaction value, which assumes that every flagged event would otherwise have produced the full observed loss (Baesens et al., 2015; Hand, 2006).

4.9 Independent Validation and Stress Testing

Independent validation should challenge the assumptions used during development. Validators need access to data definitions, code or executable artifacts, feature lineage, model versions, hyperparameters, threshold rationale, and known limitations. They should reproduce core results and test alternative time windows, prevalence levels, missing-feature conditions, delayed inputs, data-quality failures, and capacity constraints. Independence does not require organizational isolation

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 09/08/2026 - **Accepted:** 10/09/2026
Vol: 01.04
DOI: 10.66104/t68phf90
Pages: 1-37

from domain knowledge; it requires authority to question evidence and prevent deployment when material risks remain unresolved (NIST, 2023; Sculley et al., 2015).

Stress testing should include plausible operational extremes rather than only random noise. Examples include holiday volume, service degradation, sudden merchant migration, new payment products, mass credential compromise, incomplete graph updates, unusually delayed labels, and adversarial concentration just below policy thresholds. The purpose is not to predict every future attack but to learn how the system fails, whether safeguards remain available, and how quickly operations can recover (Gama et al., 2014; NIST, 2023).

Rollback is an essential control for continuously updated systems. A new model can pass offline validation while interacting unexpectedly with live features, rules, or customer behavior. Shadow deployment, canary release, champion-challenger comparison, versioned policies, and tested rollback procedures limit exposure. Emergency controls should be simple enough to operate when normal model dependencies are unavailable and should generate records for later review (NIST, 2023; Sculley et al., 2015).

Documentation must be treated as operational infrastructure. Model cards, data documentation, validation reports, decision logs, monitoring definitions, incident records, and change approvals allow institutions to reconstruct why a decision system behaved as it did. Documentation also reduces key-person dependency and supports regulatory examination. It should record uncertainty and unresolved limitations rather than presenting governance as evidence that risk has disappeared (NIST, 2023).

5. Discussion

5.1 No Universally Superior Architecture

Architecture rankings are unstable because studies differ in data, preprocessing, validation, resampling, features, thresholds, and fraud typology. GNNs have a structural advantage when relationships carry signal, recurrent models when order matters, and autoencoders when labels are limited. Hybrid systems are attractive because they combine complementary evidence, but added complexity increases maintenance, latency, and validation burden. The appropriate model is the least complex system that satisfies the documented operational objective under realistic testing (Borisov et al., 2022; Hilal et al., 2022; Sculley et al., 2015).

Rules remain valuable within such systems. They can encode legal restrictions, emergency controls, known compromised identifiers, and transparent policy. Machine learning can rank ambiguous cases, while rules enforce non-negotiable constraints and provide fallbacks. A layered architecture also reduces dependency

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 09/08/2026 - **Accepted:** 10/09/2026
Vol: 01.04
DOI: 10.66104/t68phf90
Pages: 1-37

on any single model and can make incident response faster (Baesens et al., 2015; NIST, 2023).

5.2 From Model Accuracy to System Utility

The dominant research narrative often rewards small gains on shared benchmarks, but institutional utility depends on avoided loss, manageable review volume, customer friction, resilience, and compliance. A slower or less accurate model may be preferable if its probabilities are well calibrated, explanations are stable, features are available, and performance degrades gracefully. Conversely, a model with a superior AUC may create negative value when it depends on leaked information or floods operations with alerts (Saito & Rehmsmeier, 2015; Sculley et al., 2015).

This shift also changes publication practice. Authors should report dataset dates, prevalence, label delay, split strategy, leakage controls, preprocessing, baselines, uncertainty, calibration, inference environment, and threshold outcomes. Proprietary constraints may prevent release of raw data, but they do not justify omitting methodological detail. Reproducibility can be strengthened through code, synthetic test fixtures, feature definitions, and independently reviewable evaluation protocols (Sculley et al., 2015).

5.3 Governance as a Technical Requirement

Governance is sometimes presented as external paperwork applied after model development. In fraud detection it shapes technical design: explanation requirements affect architecture; privacy limits feature collection; response deadlines constrain inference; appeal rights require logging; and monitoring duties require label pipelines. The NIST AI RMF organizes risk management around governance, mapping, measurement, and management, offering a useful general structure that institutions can adapt to financial regulation and internal model-risk frameworks (NIST, 2023).

Human oversight should be specific rather than symbolic. An analyst cannot meaningfully supervise thousands of opaque alerts without time, evidence, and authority. Effective oversight defines which decisions are automated, which require review, what information is shown, how disagreement is recorded, when escalation occurs, and how customer challenges feed back into validation. Human involvement can introduce inconsistency or automation bias, so reviewer performance also requires evaluation (NIST, 2023).

5.4 Research Priorities

Future research should emphasize temporally realistic, cross-domain, and cost-sensitive evaluation. Valuable directions include dynamic graphs, uncertainty estimation, delayed-label learning, privacy-preserving collaboration, causal analysis

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 09/08/2026 - **Accepted:** 10/09/2026
Vol: 01.04
DOI: 10.66104/t68phf90
Pages: 1-37

of interventions, and robust calibration under prevalence change. Studies should test failure modes and negative results, not only best-case benchmark performance (Gama et al., 2014; Kairouz et al., 2021).

Brazilian research needs transparent evaluation of instant-payment typologies, institutional coordination, customer impacts, and regional or accessibility-related differences. Public datasets that preserve privacy while representing modern behavior would improve independent scrutiny. Regulatory sandboxes and controlled collaborations may help test new methods without confusing experimentation with unrestricted deployment (BCB, 2025; NIST, 2023).

5.5 Organizational Capability and Implementation Maturity

Model reliability depends on organizational capability as much as on architecture. A mature program requires fraud specialists, data engineers, model developers, platform engineers, cybersecurity teams, legal and privacy professionals, customer-service staff, independent validators, and accountable executives. These roles interpret different forms of evidence and should not be collapsed into a nominal model owner. Responsibility maps are particularly important during incidents, when feature failures, emerging fraud, customer complaints, and regulatory reporting may require simultaneous action (NIST, 2023; Sculley et al., 2015).

Implementation maturity can be assessed progressively. An experimental stage may use static datasets and retrospective analysis; a controlled pilot adds point-in-time features, shadow scoring, documented thresholds, and manual review; limited production adds monitoring, rollback, incident response, and customer safeguards; mature operation adds independent validation, segment analysis, adversarial testing, continuity planning, and periodic reconsideration of the use case. Advancement should depend on evidence that controls work, not on pressure to deploy a technically impressive model (NIST, 2023).

Procurement creates a related risk when institutions acquire models, data, or cloud services from vendors. Outsourcing implementation does not outsource accountability. Contracts and due diligence should address data provenance, intellectual property, security, audit access, service continuity, update practices, subcontractors, performance claims, incident notification, and exit arrangements. Institutions also need enough internal competence to challenge vendor evidence and reproduce essential monitoring independently (NIST, 2023).

Finally, organizational learning requires preserving negative evidence. Failed pilots, rejected features, threshold incidents, model degradation, and customer complaints can reveal more about system boundaries than a successful benchmark. If only positive performance is circulated, later teams may repeat earlier mistakes. Structured post-incident reviews and searchable decision records convert individual

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 09/08/2026 - **Accepted:** 10/09/2026
Vol: 01.04
DOI: 10.66104/t68phf90
Pages: 1-37

experience into institutional knowledge while supporting accountability (Sculley et al., 2015).

5.6 Limitations of This Review

This integrative narrative review is not exhaustive and provides no pooled effect. Methodological limitations include a compact purposively assembled corpus, single-author screening and coding, no inter-rater reliability, and unavailable retrospective retrieval counts. Evidentiary limitations include restricted production disclosure, aging public datasets, artificial prevalence, weak or inconsistent comparators, scarce prospective or external validation, and limited reporting of latency, calibration, customer consequences, or negative results. Operational limitations include delayed labels, selective investigation, missing counterfactuals for blocked transactions, policy feedback, adversarial adaptation, and institution-specific capacity. Temporal limitations include changing fraud tactics, architectures, regulation, Pix rules, and payment behavior. LOAF-RTFD is a proposed synthesis and communication tool, not an empirically validated operational standard.

6. Conclusion

Deep learning expands the representational tools available for financial fraud detection, but no architecture is intrinsically reliable. Operational value emerges when the representation matches the fraud mechanism, evaluation reflects rare-event costs, features are available within the deadline, and monitoring detects changing behavior. Benchmark accuracy is evidence about a model under specified conditions, not proof of production safety (Hilal et al., 2022; Sculley et al., 2015).

Reliable real-time detection requires a layered sociotechnical system: well-governed data, point-in-time-correct features, calibrated scores, rules and authentication, human escalation, customer communication, privacy controls, drift monitoring, incident response, and independent validation. Pix illustrates that speed increases the importance of coordination and recovery as well as prediction (BCB, 2024, 2025; NIST, 2023).

The most defensible research and governance question is therefore not which neural network wins in the abstract, but which complete system produces the best risk-adjusted outcome for a defined payment environment and affected population. Progress should be measured through prevented harm, controlled customer burden, reproducibility, resilience, and accountability alongside predictive performance (EBA & ECB, 2024; NIST, 2023).

Models and thresholds should not be transferred across payment products, institutions, countries, or fraud typologies without temporal validation, calibration,

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 09/08/2026 - **Accepted:** 10/09/2026
Vol: 01.04
DOI: 10.66104/t68phf90
Pages: 1-37

customer-impact assessment, and operational stress testing. The same architecture may produce different utility when prevalence, labels, intervention options, customer population, infrastructure, or legal duties change.

Appendix A Search and Selection Record

The search supported a structured narrative review rather than systematic-review claims. Exact historical retrieval and duplicate counts were not preserved; the table therefore reports only documented facts and marks unavailable retrospective counts explicitly.

Platform	Fields and filters	Representative query	Historical count
IEEE Xplore	Metadata/abstract; 2020-2026 priority	fraud detection AND (real time OR streaming OR graph OR drift)	Not preserved
ACM Digital Library	Title/abstract; 2020-2026 priority	financial fraud AND (deep learning OR graph neural network OR explainability)	Not preserved
Google Scholar	Broad fields; relevance-ranked focused searches	Pix fraud; delayed labels; calibration; customer burden	Not preserved
Scopus-indexed/Crossref-linked platforms	Title/abstract/keywords where available	payment fraud AND (latency OR governance OR privacy)	Not preserved
Official repositories	Issuing body; current and superseded rules checked	Pix MED; cautious blocking; fraud marking; AI/model risk	Not applicable

Stage	Procedure	Recorded result

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 09/08/2026 - **Accepted:** 10/09/2026
Vol: 01.04
DOI: 10.66104/t68phf90
Pages: 1-37

Platform	Fields and filters	Representative query	Historical count
Identification	Database searching, official repositories and citation tracing	Exact total unavailable	
Deduplication	Duplicate publications and superseded versions excluded	Exact count unavailable	
Title/abstract relevance	Direct relevance to at least one review domain	Exact count unavailable	
Full-text eligibility	Transparent empirical, methodological, benchmark or authoritative contextual contribution	Exclusions: promotional, unverifiable, duplicate, aggregate-accuracy-only	
Final inclusion	Sources assigned a principal evidence role	40 sources	

Appendix B Source Characteristics

Source	Source type	Primary context	Evidence role
Abdallah, A., Maarof, M. A., & Zainal, A. (2016)	Review/survey	Evaluation/governance	Analytical
Baesens, B., Van Vlasselaer, V., & Verbeke, W. (2015)	Book/chapter	Evaluation/governance	Methodological
Banco Central do Brasil. (2024)	Institutional/regulatory	Pix/instant payments	Contextual
Banco	Institutional/regulatory	Pix/instant payments	Contextual

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 09/08/2026 - **Accepted:** 10/09/2026
Vol: 01.04
DOI: 10.66104/t68phf90
Pages: 1-37

Source	Source type	Primary context	Evidence role
Central do Brasil. (2025)			
Bolton, R. J., & Hand, D. J. (2002)	Review/survey	Evaluation/governance	Analytical
Borisov, V., Leemann, T., Seßler, K., Haug, J., Pawelczyk, M., & Kasneci, G. (2022)	Review/survey	Architecture/ML	Analytical
Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002)	Empirical/methodological	Architecture/ML	Analytical
Dal Pozzolo, A., Boracchi, G., Caelen, O., Alippi, C., & Bontempi, G. (2018)	Empirical/methodological	Architecture/ML	Analytical
Dou, Y., Liu, Z., Sun, L., Deng, Y., Peng, H., & Yu, P. S. (2020)	Conference/method	Architecture/ML	Analytical
Dwork, C., & Roth, A. (2014)	Empirical/methodological	Evaluation/governance	Analytical

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 09/08/2026 - **Accepted:** 10/09/2026
Vol: 01.04
DOI: 10.66104/t68phf90
Pages: 1-37

Source	Source type	Primary context	Evidence role
European Banking Authority, & European Central Bank. (2024)	Institutional/regulatory	Evaluation/governance	Contextual
Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014)	Review/survey	Evaluation/governance	Analytical
Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017)	Conference/method	Architecture/ML	Analytical
Hand, D. J. (2006)	Empirical/methodological	Evaluation/governance	Analytical
Hilal, W., Gadsden, S. A., & Yawney, J. (2022)	Review/survey	Architecture/ML	Analytical
Hochreiter, S., & Schmidhuber, J. (1997)	Empirical/methodological	Architecture/ML	Analytical
Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier,	Empirical/methodological	Evaluation/governance	Analytical

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 09/08/2026 - **Accepted:** 10/09/2026
Vol: 01.04
DOI: 10.66104/t68phf90
Pages: 1-37

Source	Source type	Primary context	Evidence role
P.-E., He-Guelton, L., & Caelen, O. (2018)			
Kairouz, P., et al. (2021)	Empirical/methodological	Evaluation/governance	Analytical
Kreps, J., Narkhede, N., & Rao, J. (2011)	Conference/method	Evaluation/governance	Analytical
Lopez-Rojas, E. A., Elmir, A., & Axelsson, S. (2016)	Conference/method	Evaluation/governance	Analytical
Lundberg, S. M., & Lee, S.-I. (2017)	Conference/method	Architecture/ML	Analytical
Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021)	Review/survey	Evaluation/governance	Analytical
National Institute of Standards and Technology. (2023)	Institutional/regulatory	Evaluation/governance	Contextual
Pang, G., Shen, C., Cao, L., &	Review/survey	Architecture/ML	Analytical

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 09/08/2026 - **Accepted:** 10/09/2026
Vol: 01.04
DOI: 10.66104/t68phf90
Pages: 1-37

Source	Source type	Primary context	Evidence role
van den Hengel, A. (2021)			
Saito, T., & Rehmsmeier, M. (2015)	Empirical/methodological	Evaluation/governance	Analytical
Sculley, D., et al. (2015)	Conference/method	Architecture/ML	Analytical
Vaswani, A., et al. (2017)	Conference/method	Architecture/ML	Analytical
Weber, M., Domeniconi, G., Chen, J., Weidele, D. K. I., Bellei, C., Robinson, T., & Leiserson, C. E. (2019)	Empirical/methodological	Architecture/ML	Analytical
Žliobaitė, I., Pechenizkiy, M., & Gama, J. (2016)	Book/chapter	Evaluation/governance	Methodological
Ali, A., Abd Razak, S., Othman, S. H., Eisa, T. A. E., Al-Dhaqm, A., Nasser, M., Elhassan, T., Elshafie, H., & Saif, A. (2022)	Review/survey	Evaluation/governance	Analytical

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 09/08/2026 - **Accepted:** 10/09/2026
Vol: 01.04
DOI: 10.66104/t68phf90
Pages: 1-37

Source	Source type	Primary context	Evidence role
Banco Central do Brasil. (2021)	Institutional/regulatory	Pix/instant payments	Contextual
Banco Central do Brasil. (2024)	Institutional/regulatory	Pix/instant payments	Contextual
Banco Central do Brasil. (2025)	Institutional/regulatory	Pix/instant payments	Contextual
Banco Central do Brasil. (2025)	Institutional/regulatory	Pix/instant payments	Contextual
Banco Central do Brasil. (2025)	Institutional/regulatory	Pix/instant payments	Contextual
Banco Central do Brasil. (2026)	Institutional/regulatory	Pix/instant payments	Contextual
Banco Central do Brasil. (2026)	Institutional/regulatory	Pix/instant payments	Contextual
Chen, Y., Liu, X., & Li, H. (2025)	Review/survey	Evaluation/governance	Analytical
Corso, G., Stark, H., Jegelka, S., Jaakkola, T., & Barzilay, R. (2024)	Review/survey	Architecture/ML	Analytical
Tong, G., & Shen, J.	Empirical/methodological	Architecture/ML	Analytical

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 09/08/2026 - **Accepted:** 10/09/2026
Vol: 01.04
DOI: 10.66104/t68phf90
Pages: 1-37

Source	Source type	Primary context	Evidence role
(2023)			

Appendix C Critical Appraisal Matrix

Source	Data/split	Metrics/calibration	Latency/drift	Explainability	Operational relevance	Principal limitation
Abdallah, A., Maarof, M. A., & Zainal, A. (2016)	Varied	Synthesized	Varied	Varied/not reported	Direct or methodological	Heterogeneous included evidence
Baesens, B., Van Vlasselaer, V., & Verbeke, W. (2015)	As reported	Often incomplete	Often not reported	Varied/not reported	Direct or methodological	Dataset/context limits transfer
Banco Central do Brasil. (2024)	N/A	N/A	N/A	N/A	High contextual	No comparative effect evidence
Banco Central do Brasil. (2025)	N/A	N/A	N/A	N/A	High contextual	No comparative effect evidence
Bolton, R. J., & Hand, D. J. (2002)	Varied	Synthesized	Varied	Varied/not reported	Direct or methodological	Heterogeneous included evidence
Borisov, V., Leemann, T., Seßler, K., Haug, J., Pawelczyk, M., & Kasneci, G. (2022)	Varied	Synthesized	Varied	Varied/not reported	Direct or methodological	Heterogeneous included evidence
Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002)	As reported	Often incomplete	Often not reported	Varied/not reported	Direct or methodological	Dataset/context limits transfer
Dal Pozzolo, A., Boracchi, G., Caelen, O., Alippi, C.,	As reported	Often incomplete	Often not reported	Varied/not reported	Direct or methodological	Dataset/context limits transfer

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 09/08/2026 - **Accepted:** 10/09/2026
Vol: 01.04
DOI: 10.66104/t68phf90
Pages: 1-37

Source	Data/split	Metrics/calibration	Latency/drift	Explainability	Operational relevance	Principal limitation
& Bontempi, G. (2018)						
Dou, Y., Liu, Z., Sun, L., Deng, Y., Peng, H., & Yu, P. S. (2020)	As reported	Often incomplete	Often not reported	Varied/not reported	Direct or methodological	Dataset/context limits transfer
Dwork, C., & Roth, A. (2014)	As reported	Often incomplete	Often not reported	Varied/not reported	Direct or methodological	Dataset/context limits transfer
European Banking Authority, & European Central Bank. (2024)	N/A	N/A	N/A	N/A	High contextual	No comparative effect evidence
Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014)	Varied	Synthesized	Varied	Varied/not reported	Direct or methodological	Heterogeneous included evidence
Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017)	As reported	Often incomplete	Often not reported	Varied/not reported	Direct or methodological	Dataset/context limits transfer
Hand, D. J. (2006)	As reported	Often incomplete	Often not reported	Varied/not reported	Direct or methodological	Dataset/context limits transfer
Hilal, W., Gadsden, S. A., & Yawney, J. (2022)	Varied	Synthesized	Varied	Varied/not reported	Direct or methodological	Heterogeneous included evidence
Hochreiter, S., & Schmidhuber, J. (1997)	As reported	Often incomplete	Often not reported	Varied/not reported	Direct or methodological	Dataset/context limits transfer
Jurgovsky, J., Granitzer, M.,	As reported	Often incomplete	Often not reported	Varied/not reported	Direct or methodological	Dataset/context limits transfer

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 09/08/2026 - **Accepted:** 10/09/2026
Vol: 01.04
DOI: 10.66104/t68phf90
Pages: 1-37

Source	Data/split	Metrics/calibration	Latency/drift	Explainability	Operational relevance	Principal limitation
Ziegler, K., Calabretto, S., Portier, P.-E., He-Guelton, L., & Caelen, O. (2018)						
Kairouz, P., et al. (2021)	As reported	Often incomplete	Often not reported	Varied/not reported	Direct or methodological	Dataset/context limits transfer
Kreps, J., Narkhede, N., & Rao, J. (2011)	As reported	Often incomplete	Often not reported	Varied/not reported	Direct or methodological	Dataset/context limits transfer
Lopez-Rojas, E. A., Elmir, A., & Axelsson, S. (2016)	As reported	Often incomplete	Often not reported	Varied/not reported	Direct or methodological	Dataset/context limits transfer
Lundberg, S. M., & Lee, S.-I. (2017)	As reported	Often incomplete	Often not reported	Varied/not reported	Direct or methodological	Dataset/context limits transfer
Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021)	Varied	Synthesized	Varied	Varied/not reported	Direct or methodological	Heterogeneous included evidence
National Institute of Standards and Technology. (2023)	N/A	N/A	N/A	N/A	High contextual	No comparative effect evidence
Pang, G., Shen, C., Cao, L., & van den Hengel, A. (2021)	Varied	Synthesized	Varied	Varied/not reported	Direct or methodological	Heterogeneous included evidence
Saito, T., & Rehmsmeier, M. (2015)	As reported	Often incomplete	Often not reported	Varied/not reported	Direct or methodological	Dataset/context limits transfer

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 09/08/2026 - **Accepted:** 10/09/2026
Vol: 01.04
DOI: 10.66104/t68phf90
Pages: 1-37

Source	Data/split	Metrics/calibration	Latency/drift	Explainability	Operational relevance	Principal limitation
Sculley, D., et al. (2015)	As reported	Often incomplete	Often not reported	Varied/not reported	Direct or methodological	Dataset/context limits transfer
Vaswani, A., et al. (2017)	As reported	Often incomplete	Often not reported	Varied/not reported	Direct or methodological	Dataset/context limits transfer
Weber, M., Domeniconi, G., Chen, J., Weidele, D. K. I., Bellei, C., Robinson, T., & Leiserson, C. E. (2019)	As reported	Often incomplete	Often not reported	Varied/not reported	Direct or methodological	Dataset/context limits transfer
Žliobaitė, I., Pechenizkiy, M., & Gama, J. (2016)	As reported	Often incomplete	Often not reported	Varied/not reported	Direct or methodological	Dataset/context limits transfer
Ali, A., Abd Razak, S., Othman, S. H., Eisa, T. A. E., Al-Dhaqm, A., Nasser, M., Elhassan, T., Elshafie, H., & Saif, A. (2022)	Varied	Synthesized	Varied	Varied/not reported	Direct or methodological	Heterogeneous included evidence
Banco Central do Brasil. (2021)	N/A	N/A	N/A	N/A	High contextual	No comparative effect evidence
Banco Central do Brasil. (2024)	N/A	N/A	N/A	N/A	High contextual	No comparative effect evidence
Banco Central do Brasil. (2025)	N/A	N/A	N/A	N/A	High contextual	No comparative effect evidence
Banco Central do Brasil. (2025)	N/A	N/A	N/A	N/A	High contextual	No comparative effect evidence
Banco Central do	N/A	N/A	N/A	N/A	High contextual	No comparative

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 09/08/2026 - **Accepted:** 10/09/2026
Vol: 01.04
DOI: 10.66104/t68phf90
Pages: 1-37

Source	Data/split	Metrics/calibration	Latency/drift	Explainability	Operational relevance	Principal limitation
Brasil. (2025)						effect evidence
Banco Central do Brasil. (2026)	N/A	N/A	N/A	N/A	High contextual	No comparative effect evidence
Banco Central do Brasil. (2026)	N/A	N/A	N/A	N/A	High contextual	No comparative effect evidence
Chen, Y., Liu, X., & Li, H. (2025)	Varied	Synthesized	Varied	Varied/not reported	Direct or methodological	Heterogeneous included evidence
Corso, G., Stark, H., Jegelka, S., Jaakkola, T., & Barzilay, R. (2024)	Varied	Synthesized	Varied	Varied/not reported	Direct or methodological	Heterogeneous included evidence
Tong, G., & Shen, J. (2023)	As reported	Often incomplete	Often not reported	Varied/not reported	Direct or methodological	Dataset/context limits transfer

Ethical Approval

Ethical approval was not required because this study analyzed publicly available literature and institutional documents and involved no human participants, identifiable personal data, animals, or intervention.

Funding

The author received no financial support for the research, authorship, or publication of this article.

Conflicts of Interest

The author declares no conflict of interest.

References

Abdallah, A., Maarof, M. A., & Zainal, A. (2016). Fraud detection system: A survey. Journal of Network and Computer Applications, 68, 90-113. https://doi.org/10.1016/j.jnca.2016.04.007

Ali, A., Abd Razak, S., Othman, S. H., Eisa, T. A. E., Al-Dhaqm, A., Nasser, M., Elhassan, T., Elshafie, H., & Saif, A. (2022). Financial fraud detection based on

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 09/08/2026 - **Accepted:** 10/09/2026
Vol: 01.04
DOI: 10.66104/t68phf90
Pages: 1-37

machine learning: A systematic literature review. Applied Sciences, 12(19), 9637. https://doi.org/10.3390/app12199637

Baesens, B., Van Vlasselaer, V., & Verbeke, W. (2015). Fraud analytics using descriptive, predictive, and social network techniques. Wiley.

Banco Central do Brasil. (2021). Resolução BCB nº 147, de 28 de setembro de 2021. https://www.bcb.gov.br/estabilidadefinanceira/exibenormativo?numero=147&tipo=Resolução%20BCB

Banco Central do Brasil. (2024). Annual report 2023. https://www.bcb.gov.br/content/publications/Annual_Report_docs/AR-2023.pdf

Banco Central do Brasil. (2024). Resolução BCB nº 403, de 22 de julho de 2024. https://www.bcb.gov.br/estabilidadefinanceira/exibenormativo?numero=403&tipo=Resolução%20BCB

Banco Central do Brasil. (2025). Annual report 2024. https://www.bcb.gov.br/content/publications/Annual_Report_docs/AR-2024.pdf

Banco Central do Brasil. (2025a). Instrução Normativa BCB nº 655, de 28 de agosto de 2025. https://www.bcb.gov.br/estabilidadefinanceira/exibenormativo?numero=655&tipo=Instrução%20Normativa%20BCB

Banco Central do Brasil. (2025b). BC aprimora o Mecanismo Especial de Devolução do Pix. https://www.bcb.gov.br/detalhenoticia/20817/nota

Banco Central do Brasil. (2025c). Pix terá botão de contestação. https://www.bcb.gov.br/detalhenoticia/20865/noticia

Banco Central do Brasil. (2026a). Guia de implementação dos procedimentos de devolução no Pix, com ênfase no Mecanismo Especial de Devolução (Versão 4.2). https://www.bcb.gov.br/content/estabilidadefinanceira/pix/Regulamento_Pix/versoes_futuras/Guia_MED-versao4-2.pdf

Banco Central do Brasil. (2026b). Relatório de Gestão do Pix 2023-2025. https://www.bcb.gov.br/content/estabilidadefinanceira/pix/relatorio_de_gestao_pix/relatorio_gestao_pix_2026.pdf

Bolton, R. J., & Hand, D. J. (2002). Statistical fraud detection: A review. Statistical Science, 17(3), 235-255. https://doi.org/10.1214/ss/1042727940

Borisov, V., Leemann, T., Seßler, K., Haug, J., Pawelczyk, M., & Kasneci, G. (2022). Deep neural networks and tabular data: A survey. IEEE Transactions on

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 09/08/2026 - **Accepted:** 10/09/2026
Vol: 01.04
DOI: 10.66104/t68phf90
Pages: 1-37

Neural Networks and Learning Systems. https://doi.org/10.1109/TNNLS.2022.3229161

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, 321-357. https://doi.org/10.1613/jair.953

Chen, Y., Liu, X., & Li, H. (2025). Deep learning in financial fraud detection: A systematic literature review. Journal of Innovation & Knowledge, 10, 100682.

Corso, G., Stark, H., Jegelka, S., Jaakkola, T., & Barzilay, R. (2024). Graph neural networks. Nature Reviews Methods Primers, 4, Article 17. https://doi.org/10.1038/s43586-024-00294-7

Dal Pozzolo, A., Boracchi, G., Caelen, O., Alippi, C., & Bontempi, G. (2018). Credit card fraud detection: A realistic modeling and a novel learning strategy. IEEE Transactions on Neural Networks and Learning Systems, 29(8), 3784-3797. https://doi.org/10.1109/TNNLS.2017.2736643

Dou, Y., Liu, Z., Sun, L., Deng, Y., Peng, H., & Yu, P. S. (2020). Enhancing graph neural network-based fraud detectors against camouflaged fraudsters. Proceedings of CIKM, 315-324. https://doi.org/10.1145/3340531.3411903

Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. Foundations and Trends in Theoretical Computer Science, 9(3-4), 211-407. https://doi.org/10.1561/0400000042

European Banking Authority, & European Central Bank. (2024). 2024 report on payment fraud. https://www.eba.europa.eu/publications-and-media/press-releases/eba-and-ecb-release-joint-report-payment-fraud

Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. ACM Computing Surveys, 46(4), Article 44. https://doi.org/10.1145/2523813

Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. Proceedings of ICML, 1321-1330.

Hand, D. J. (2006). Classifier technology and the illusion of progress. Statistical Science, 21(1), 1-14. https://doi.org/10.1214/088342306000000060

Hilal, W., Gadsden, S. A., & Yawney, J. (2022). Financial fraud: A review of anomaly detection techniques and recent advances. Expert Systems with Applications, 193, 116429. https://doi.org/10.1016/j.eswa.2021.116429

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 09/08/2026 - **Accepted:** 10/09/2026
Vol: 01.04
DOI: 10.66104/t68phf90
Pages: 1-37

Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P.-E., He-Guelton, L., & Caelen, O. (2018). Sequence classification for credit-card fraud detection. Expert Systems with Applications, 100, 234-245. https://doi.org/10.1016/j.eswa.2018.01.037

Kairouz, P., et al. (2021). Advances and open problems in federated learning. Foundations and Trends in Machine Learning, 14(1-2), 1-210. https://doi.org/10.1561/2200000083

Kreps, J., Narkhede, N., & Rao, J. (2011). Kafka: A distributed messaging system for log processing. Proceedings of the NetDB.

Lopez-Rojas, E. A., Elmir, A., & Axelsson, S. (2016). PaySim: A financial mobile money simulator for fraud detection. Proceedings of the European Modeling and Simulation Symposium, 249-255.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. ACM Computing Surveys, 54(6), Article 115. https://doi.org/10.1145/3457607

National Institute of Standards and Technology. (2023). Artificial intelligence risk management framework (AI RMF 1.0). https://doi.org/10.6028/NIST.AI.100-1

Pang, G., Shen, C., Cao, L., & van den Hengel, A. (2021). Deep learning for anomaly detection: A review. ACM Computing Surveys, 54(2), Article 38. https://doi.org/10.1145/3439950

Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLOS ONE, 10(3), e0118432. https://doi.org/10.1371/journal.pone.0118432

Sculley, D., et al. (2015). Hidden technical debt in machine learning systems. Advances in Neural Information Processing Systems, 28.

Tong, G., & Shen, J. (2023). Financial transaction fraud detector based on imbalance learning and graph neural network. Applied Soft Computing, 149, 110984. https://doi.org/10.1016/j.asoc.2023.110984

Vaswani, A., et al. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30.

Weber, M., Domeniconi, G., Chen, J., Weidele, D. K. I., Bellei, C., Robinson, T., & Leiserson, C. E. (2019). Anti-money laundering in Bitcoin: Experimenting with

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 09/08/2026 - **Accepted:** 10/09/2026
Vol: 01.04
DOI: 10.66104/t68phf90
Pages: 1-37

graph convolutional networks for financial forensics. arXiv. https://doi.org/10.48550/arXiv.1908.02591

Žliobaitė, I., Pechenizkiy, M., & Gama, J. (2016). An overview of concept drift applications. In N. Japkowicz & J. Stefanowski (Eds.), Big data analysis: New algorithms for a new society (pp. 91-114). Springer. https://doi.org/10.1007/978-3-319-26989-4_4

CC BY