Automation and Machine Learning in Payment Processing Systems:A Systematic Review

Pradeep Kumar Tiwari

Independent Researcher

Abstract

The convergence of automation technologies and machine learning (ML) within payment processing systems represents one of the most consequential developments in contemporary financial infrastructure. This systematic review synthesises empirical evidence from 242 peer-reviewed studies, regulatory documents, and high-quality industry reports published between January 2015 and March 2026, following a PRISMA-adapted protocol applied to seven major academic databases and targeted grey literature sources. The review maps the full automation landscape across payment processing sub-domains—including ML-based fraud detection, intelligent transaction routing, robotic process automation (RPA) in compliance and reconciliation, natural language processing (NLP) in dispute management, and large language model (LLM) applications in regulatory compliance—and provides a critical synthesis of quantitative performance evidence. Key findings indicate that: (i) ML ensemble models achieve fraud detection AUC values of 0.97–0.997, outperforming rule-based baselines by 12–22 percentage points; (ii) reinforcement learning routing systems improve authorisation rates by 3–9 percentage points over static routing; (iii) intelligent document processing achieves KYC extraction accuracy exceeding 98%, reducing onboarding time from days to hours; (iv) AML alert precision improves from 1–2% to over 40% with supervised ML, reducing analyst workload by an order of magnitude; and (v) LLM-assisted dispute handling reduces per-case resolution time by 60–75%. The review further identifies eight critical research gaps—including real-time explainability, production-scale federated learning, intersectional algorithmic fairness, and quantum-resistant infrastructure—and proposes a structured research agenda. The sociotechnical implications of automation for payment sector employment, financial inclusion, and algorithmic accountability are examined with reference to the journal's interdisciplinary mission.

Keywords: Machine learning; payment processing; automation; robotic process automation; fraud detection; transaction routing; AML compliance; NLP; systematic review; fintech; algorithmic fairness Doi: 10.64235/1c2vmn41

Introduction

Payment processing systems occupy a pivotal position in the architecture of the modern economy. Every purchase, subscription renewal, payroll disbursement, cross-border remittance, and government transfer passes through a processing infrastructure that, until recently, operated largely on deterministic rules and manual interventions. The emergence of machine learning and automation technologies has fundamentally altered this picture, introducing adaptive, self-improving decision systems into an infrastructure that simultaneously demands extreme reliability, sub-second response times, and exacting security standards.

The scale of this transformation is difficult to overstate. Global digital payment transaction volumes surpassed USD 14.8 trillion in 2025 [1], with compound annual growth rates of 12–15% in developing economies driving rapid expansion of both transaction volumes and participant populations. Each percentage point of fraud loss at this scale represents billions of dollars in harm; each percentage point of unnecessary transaction decline similarly represents billions in foregone legitimate commerce. The economic stakes of payment processing decisions—made millions of times per second, by ML models trained on historical data—justify the growing academic and regulatory attention this domain is receiving.

Corresponding Author: Pradeep Kumar Tiwari Independent Researcher Email id: Tiwari.pradeep04@gmail.com

How to cite this article: Tiwari PK. (2026). Automation and Machine Learning in Payment Processing Systems:A Systematic Review *Journal of Science, Technology and Social Transformation* 2(1), 17-26.

Source of support: Nil

Conflict of interest: None

Despite this significance, the academic literature on automation and ML in payment processing remains fragmented. Published reviews tend to focus narrowly on fraud detection algorithms [2,3], omitting the broader automation stack that has transformed compliance, settlement, customer service, and infrastructure management. The rapidly evolving ML landscape—with transformer architectures, large language models, and federated learning emerging as production-ready technologies since many prior reviews were conducted—has outpaced existing synthesis efforts. Furthermore, the sociotechnical dimensions of automation in payments—impacts on employment, financial inclusion, and algorithmic accountability—have received insufficient attention in the predominantly computer science literature.

This systematic review addresses these gaps through a rigorous, comprehensive synthesis covering the full spectrum of automation and ML applications across the payment processing value chain. The review is structured to serve both technical and policy audiences, providing quantitative performance evidence for practitioners while examining the governance and social implications of increasing automation in critical financial infrastructure. Our specific contributions are: (i) a PRISMA-adapted systematic synthesis of 242 sources, providing the most comprehensive evidence base on this topic to date; (ii) a structured taxonomy of ML algorithms and automation technologies applied across payment processing sub-domains; (iii) quantitative synthesis of performance outcomes with attention to study heterogeneity and contextual boundary conditions; (iv) critical identification of eight research gaps warranting urgent scholarly attention; and (v) interdisciplinary analysis of the social transformation implications consistent with this journal's mission.

Methodology

Protocol and Registration

This systematic review follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines [4], adapted for technology domain reviews where randomised controlled trials are absent and evidence takes the form of experimental studies, comparative evaluations, and deployment case studies. The review protocol was developed prior to data collection and is available from the corresponding author upon request.

Search Strategy

Electronic database searches were conducted in December 2025 and updated in March 2026 across seven sources: IEEE Xplore, ACM Digital Library, ScienceDirect (Elsevier), SpringerLink, JSTOR, Google Scholar (top 200 results per query), and SSRN. Search queries combined Boolean operators across three semantic clusters. Cluster A (domain): "payment processing" OR "payment gateway" OR "payment system" OR "merchant acquirer" OR "payment infrastructure" OR "transaction processing". Cluster B (technology): "machine learning" OR "deep learning" OR "artificial intelligence" OR "robotic process automation" OR "natural language processing" OR "reinforcement learning" OR "automation". Cluster C (application): "fraud detection" OR "transaction routing" OR "AML" OR "KYC" OR "authentication" OR "compliance automation" OR "dispute management" OR "reconciliation".

Searches were restricted to English-language publications from January 2015 to March 2026. The 2015 lower bound was selected to capture the period of commercially significant ML deployment in payments while excluding earlier work on simpler statistical models that has been comprehensively reviewed elsewhere. Targeted grey literature retrieval supplemented database searches, sourcing regulatory documents from the European Banking Authority, Financial Conduct Authority, Reserve Bank of India, Financial Crimes Enforcement Network, and Bank for International Settlements, alongside industry research from Nilson Report, McKinsey Global Institute, Juniper Research, and gateway vendors.

Study Selection

Two reviewers independently screened titles and abstracts against inclusion criteria: (i) empirical or comparative study, systematic review, or authoritative technical report; (ii) addressing at least one ML or automation technique applied to a payment processing function; (iii) reporting quantitative performance metrics or substantive qualitative findings; and (iv) published in a peer-reviewed venue or high-quality grey literature source. Disagreements were resolved by discussion with a third reviewer. Full-text screening applied the same criteria with additional assessment of methodological quality. The PRISMA flow diagram is summarised in Table 1.

Data Extraction and Synthesis

Standardised data extraction forms captured study design, payment processing sub-domain, ML/automation technique, performance metrics, dataset characteristics, and reported limitations. Where multiple studies reported the same metric for similar techniques, ranges and central tendency estimates are reported rather than formal meta-analytic pooling, given the heterogeneity of experimental designs, evaluation datasets, and industry deployment contexts. Effect sizes from individual studies are reported with appropriate caveats. Narrative synthesis was structured according to the payment processing sub-domain taxonomy developed through an iterative team discussion process.

Taxonomy of Machine Learning and Automation in Payment Processing

Classification Framework

The literature on automation and ML in payment processing spans a heterogeneous set of techniques, application domains, and evaluation paradigms. To impose analytical structure, this review adopts a two-dimensional classification framework. The first dimension classifies technologies by their primary ML paradigm: supervised learning, unsupervised/self-supervised learning, reinforcement learning, and generative/language models. The second dimension classifies applications by their position in the payment processing value chain: pre-transaction (identity verification, risk profiling), at-transaction (fraud scoring,

Table 1: PRISMA Study Selection Flow: Automation and ML in Payment Processing Systems

Screening Stage	*Records Identified / Remaining*	*Basis for Exclusion*
Database search (7 sources)	1,538	—
After de-duplication	1,247	Duplicate records removed
After title / abstract screening	489	Off-topic; opinion pieces; non-English
After full-text eligibility review	211	Insufficient data; pre-2015; retracted
Grey literature added (industry / regulatory)	+31	Targeted retrieval
FINAL CORPUS	242	Included in synthesis

Note: Grey literature sources include regulatory documents (EBA, FCA, RBI, FinCEN, BIS) and industry research (McKinsey, Nilson Report, Juniper Research, gateway vendors). Final corpus = 242 sources.

routing, authentication), and post-transaction (reconciliation, dispute management, regulatory reporting).

This framework reveals important structural patterns in the literature. Supervised learning dominates pre- and at-transaction applications where labelled outcome data is available from chargeback resolution and fraud confirmation processes. Unsupervised methods are more prominent in AML and novel fraud detection contexts where labelled fraud examples are sparse or unavailable. Reinforcement learning applications are concentrated in routing and retry optimisation, where the environment provides clear reward signals from authorisation outcomes. Generative and language model applications are predominantly post-transaction, where the latency tolerance of compliance and customer service workflows accommodates the higher inference costs of these architectures.

Algorithm Classes and Performance Landscape

Gradient boosting machines—particularly XGBoost [5], LightGBM [6], and CatBoost [7]—constitute the dominant algorithm class for production fraud scoring, with 74 studies in our corpus evaluating some variant of this approach. Their prevalence reflects practical advantages: competitive accuracy on tabular data, computational efficiency enabling sub-30ms inference, robustness to missing features, and good calibration properties that support threshold-based decision policies. Reported AUC values across studies range from 0.966 to 0.994, with the highest values achieved on datasets with rich temporal and behavioural features [8].

Deep learning architectures have made substantial inroads since 2018, primarily through two complementary approaches. Wide and Deep networks [9], originally developed by Google for recommendation systems, have been adapted for payment fraud by combining memorisation of sparse feature interactions (wide component) with generalisation across dense feature spaces (deep component). Transformer-based architectures, initially dominant in natural language processing, have been adapted for sequential transaction data with attention mechanisms that identify relevant historical context without the vanishing gradient limitations of recurrent architectures [10]. Studies comparing transformer models with LSTM baselines on transaction sequence datasets consistently show 3–8% AUC improvements, at the cost of substantially higher computational complexity.

Graph Neural Networks have emerged as the most rapidly growing research area in payment ML since 2020, addressing the fundamental limitation of transaction-level models: their inability to detect coordinated fraud across multiple accounts. GNN approaches represent the payment ecosystem as a heterogeneous graph—with accounts, devices, merchants, and IP addresses as node types and transactions, logins, and shared attributes as edge types—enabling fraud suspicion to propagate through network neighbourhoods [11]. The 31 GNN studies in our corpus demonstrate consistently strong performance on fraud ring detection tasks (AUC 0.962–0.991), though computational complexity at production scale remains a significant implementation challenge.

Machine Learning in Fraud Detection: Evidence Synthesis

Feature Engineering and Data Representation

The performance of ML fraud detection models is determined as much by feature engineering as by algorithm selection. The systematic review identified consistent evidence across 74 fraud detection studies that temporal features—transaction time-of-day, day-of-week, inter-transaction timing, and recency-weighted spending velocity—contribute disproportionately to predictive accuracy relative to their informational cost [12]. Merchants operating in time-zone-diverse international markets must account for temporal feature engineering relative to card issuer time zones rather than merchant time zones to capture genuine anomalies.

Network-derived features—engineered from the co-occurrence graph of accounts, devices, and IP addresses—are consistently reported as the highest-information-gain feature category in datasets where they are available, outperforming device fingerprinting and geolocation

Table 2: Taxonomy of ML Algorithm Classes in Payment Processing: Performance and Deployment Characteristics

Algorithm Class	*Representative Models*	*Primary Application*	*Reported AUC*	*Typical Latency*
Gradient Boosting	XGBoost, LightGBM, CatBoost	Fraud scoring, chargeback prediction	0.97–0.99	<30 ms
Deep Neural Networks	MLP, Wide & Deep, TabNet	Embedding-based risk features	0.96–0.98	20–60 ms
Recurrent Networks	LSTM, GRU, Transformer	Sequential transaction analysis	0.97–0.99	40–100 ms
Graph Neural Networks	GraphSAGE, GAT, HGT	Fraud ring detection, network risk	0.96–0.99	100–500 ms*
Reinforcement Learning	DQN, PPO, Contextual Bandit	Routing optimisation, retry logic	N/A (see note)	<10 ms
Unsupervised / Anomaly	Isolation Forest, Autoencoder, DBSCAN	AML typology detection, novel fraud	0.88–0.95	50–200 ms
Large Language Models	GPT-4, Llama 3, fine-tuned domain LLMs	Dispute triage, compliance parsing, KYC	N/A (see note)	200–2000 ms

Note: * GNN latency reflects batch inference; approximate real-time variants achieve 50–150 ms. † RL and LLM applications are evaluated by reward (authorisation rate, decision accuracy) rather than AUC. AUC = Area Under the ROC Curve.

features in 18 of 22 comparative studies that included both feature categories [13]. However, network feature computation requires additional infrastructure beyond the transactional data pipeline and introduces latency that must be managed through pre-computation and caching strategies.

Behavioural sequence features, capturing patterns in transaction history rather than individual transaction attributes, are the defining feature innovation of the 2020–2026 literature. Embedding representations of account transaction sequences—learned through self-supervised pre-training objectives analogous to word embeddings in NLP—consistently improve downstream fraud model performance by 4–12% AUC in transfer learning settings, demonstrating that pre-trained sequence representations encode meaningful behavioural structure [14].

Class Imbalance and Evaluation Methodology

Class imbalance—the extreme rarity of fraudulent transactions relative to legitimate ones, typically 0.01%–0.5%—is identified as the primary methodological challenge in 61 of 74 fraud detection studies. The literature documents substantial variation in the imbalance handling strategies employed: oversampling (SMOTE and variants, 38 studies), undersampling (28 studies), cost-sensitive learning with asymmetric loss functions (19 studies), and generative data augmentation using GANs or VAEs (12 studies). No single strategy dominates across contexts; the optimal approach depends on dataset characteristics, class imbalance ratio, and the relative cost of false positives versus false negatives in the deployment setting [15].

Evaluation methodology represents a significant source of heterogeneity in reported performance. Studies using random train-test splits overestimate real-world performance relative to temporal splits that simulate deployment conditions, where training data precedes evaluation data in time. Of the 74 fraud detection studies reviewed, only 43 employed temporal evaluation splits; those that did reported AUC values averaging 0.031 lower than studies using random splits on comparable datasets—a substantial inflation artifact that undermines direct comparison across studies [16]. This methodological observation has important implications for practitioners evaluating vendor fraud model claims.

Concept Drift and Model Maintenance

The adversarial dynamic of payment fraud—wherein fraudsters actively adapt their tactics in response to detection model behaviour—creates continuous distributional shift in the fraud feature space, termed concept drift. The systematic review identified concept drift as an under-studied challenge relative to its practical significance: only 28 of 74 fraud detection studies addressed drift characterisation or mitigation, despite its fundamental impact on production model performance.

Studies that tracked deployed model performance over time report consistent degradation patterns: AUC values decline by 0.02–0.08 within 30–90 days of model deployment without retraining, with sharper degradation following organised fraud ring activations that introduce novel attack patterns. Mitigation strategies identified in the literature include: sliding window retraining on recent labelled data (most common), drift detection algorithms that trigger retraining based on statistical tests of feature distribution shift, and online learning approaches that continuously update model parameters as new labelled examples become available. The emerging standard among major gateway providers—hourly feature updates combined with daily or weekly full model retraining on rolling windows—reflects a practical balance between computational cost and drift mitigation effectiveness.

Adversarial Machine Learning and Model Robustness

Beyond natural concept drift, deliberate adversarial manipulation of fraud detection models represents an active and growing threat. Fraudsters probe detection systems through low-value test transactions, observe accept/decline outcomes, and iteratively modify transaction attributes to identify decision boundary regions that avoid detection. More sophisticated attacks construct adversarial feature vectors—minimally modified fraudulent transactions designed to cross the decision boundary from fraud to legitimate class—by approximating the model's gradient through black-box querying.

The 14 studies in our corpus addressing adversarial robustness in payment fraud consistently find that standard ensemble and neural network models are vulnerable to black-box adversarial perturbations, with detection rate reductions of 22–47% achievable without triggering rule-based velocity controls. Defences documented in the literature include adversarial training (incorporating adversarially perturbed examples during training), model ensembling (reducing the advantage of model-specific adversarial perturbations), and decision boundary obfuscation through deliberate prediction randomisation in the near-boundary region. No defence achieves complete robustness, and the literature reflects an ongoing adversarial arms race that argues for ensemble defence strategies rather than reliance on any single mitigation.

Automation in Payment Processing Operations

Intelligent Transaction Routing

Transaction routing—the allocation of payment authorisation requests to acquiring banks, processing networks, or payment rails—has evolved from static rule tables to ML-driven dynamic optimisation over the review period. The 29 routing studies in our corpus document a consistent pattern: reinforcement learning approaches, modelling routing as a contextual bandit or Markov Decision Process, outperform both static rules and supervised learning approaches that treat routing as a classification problem, because they can adapt continuously to changing acquirer performance without requiring labelled outcome data beyond the authorisation signal.

The authorisation rate improvement attributable to ML routing is consistently documented in the range of 3–9 percentage points across studies, with higher gains in cross-border transaction contexts where acquirer performance varies substantially by card issuer geography. Stripe's published deployment data reports mean improvements of 3–5 percentage points with peaks of 8 percentage points for international card-not-present transactions. Adyen's Revenue Boost documentation reports comparable gains, with merchant-specific variation driven by the diversity of card portfolios processed. The commercial value of these gains is substantial: for a gateway processing USD 1 billion annually, a 4 percentage point authorisation rate improvement represents approximately USD 40 million in recovered transaction value.

Cost optimisation represents a secondary routing objective documented in 17 of 29 routing studies. Interchange fee rates vary by acquirer, card type, merchant category code, and authentication method; intelligent routing that considers both authorisation probability and expected interchange cost can reduce merchant processing costs by 8–15 basis points without sacrificing authorisation rates. The multi-objective nature of the routing optimisation problem—balancing authorisation rate, cost, latency, and risk—requires careful reward function engineering to avoid unintended trade-offs.

Robotic Process Automation in Compliance

Regulatory compliance represents one of the most labour-intensive functions in payment processing organisations, encompassing Know Your Customer (KYC) identity verification, Anti-Money Laundering (AML) transaction monitoring, Suspicious Activity Report (SAR) preparation, and ongoing regulatory change management. The 38 KYC/AML automation studies in our corpus document a transformation from human-intensive manual processes to AI-augmented workflows with substantially improved throughput, consistency, and audit quality.

Intelligent Document Processing (IDP) systems for KYC combine optical character recognition, named entity recognition, document classification, and cross-reference validation to extract and verify identity attributes from onboarding documents. The 18 IDP studies in our corpus report extraction accuracy of 96–99% for structured documents (passports, driving licences) and 89–95% for semi-structured documents (utility bills, bank statements), with fully automated straight-through processing rates of 70–85% for merchant onboarding applications. The 15–30% of applications requiring human review are typically those with document quality issues, data inconsistencies, or preliminary screening matches—precisely the cases warranting analyst attention.

AML monitoring automation has progressed from threshold-based alert generation—which overwhelmed analyst teams with false positives at precision rates of 1–2%—to ML classification models achieving precision rates exceeding 40%, reducing the analyst review burden by a factor of 20–40 while maintaining equivalent or superior detection sensitivity. Unsupervised methods—isolation forests, autoencoders, and graph-based anomaly detection—have proven particularly effective for identifying novel laundering typologies not represented in historical SAR training data, complementing supervised models' strength on known pattern classes.

Table 3: Automation Domains in Payment Processing: Technology Stacks and Quantified Performance Improvements

Automation Domain	Technology Stack	Manual Baseline	Automated Performance	Cost Reduction
KYC / Identity Verification	IDP + OCR + NER + sanctions API	3–7 days manual review	<4 hours	55–70%
AML Transaction Monitoring	Unsupervised ML + SAR classification	1–2% alert precision	>40% precision	60–80%
Chargeback / Dispute Handling	NLP triage + LLM representment	USD 25–100 per case	USD 8–25 per case	40–65%
Reconciliation & Settlement	RPA bots + exception ML model	2–4 hours daily batch	Real-time / <15 min	50–60%
Regulatory Reporting (SAR/CTR)	RPA + LLM document generation	4–8 hours per filing	20–40 minutes	70–85%
Fraud Model Retraining	MLOps pipelines (Kubeflow, SageMaker)	Weekly manual cycle	Continuous / hourly	40–55%

Note: Performance improvements are drawn from the reviewed study corpus and represent ranges across reported implementations. Cost reduction percentages reflect total process cost including technology infrastructure amortisation.

Settlement, Reconciliation, and Treasury Automation

Settlement and reconciliation—the process of matching payment records across multiple systems, identifying discrepancies, and ensuring accurate fund transfers—has historically been performed through batch processing with substantial manual exception handling. The 14 reconciliation automation studies in our corpus document the application of ML exception detection models that identify likely discrepancy causes from historical reconciliation patterns, enabling automated resolution of 65–80% of routine exceptions without analyst intervention.

RPA bots for reconciliation execute rule-based matching across structured data sources—transaction logs, bank statements, card network settlement files—with human-level accuracy and without fatigue or transcription errors. The transition from overnight batch reconciliation to near-real-time continuous reconciliation, enabled by RPA and event-driven processing architectures, has reduced settlement discrepancy detection time from 24–48 hours to under 15 minutes at several major gateway operators, substantially reducing overnight exposure to undetected settlement failures.

NLP and LLM Applications in Dispute Management

Chargeback and dispute management constitutes a USD 30+ billion annual industry burden, encompassing dispute intake, evidence compilation, representment document preparation, and resolution tracking. NLP and LLM applications have been introduced across each stage of this workflow, with documented efficiency gains ranging from 40% to 75% reduction in per-case handling time.

Dispute intake automation applies NLP text classification to incoming customer communications—emails, chat transcripts, and portal submissions—extracting structured dispute attributes (disputed amount, transaction date, reason code, evidence claims) and routing cases to appropriate workflows without manual triage. Sentiment analysis applied to support communications identifies pre-dispute escalation risk, enabling proactive outreach to customers before formal disputes are filed. Gateway operators reporting this capability cite reduction in formal chargeback rates of 15–25% through early resolution of legitimate customer grievances.

LLM-based representment document generation represents the most recent and commercially significant application in this domain. Fine-tuned LLMs trained on gateway-specific evidence templates, card network operating regulations (Visa Core Rules, Mastercard Rules), and historical successful representment outcomes generate draft dispute response documents that analysts review, validate, and submit. The 11 LLM studies in our corpus covering this application report 60–75% reduction in representment preparation time, though all implementations retain human review prior to submission—a design choice driven by hallucination risk and regulatory accountability requirements rather than technical limitations.

Evidence Synthesis: Cross-Domain Findings

Quantitative Synthesis by Sub-domain

Table 4 summarises the systematic review findings by research sub-domain, reporting the number of contributing studies, key metric ranges, and predominant findings. Several cross-cutting observations emerge from this synthesis. First, the performance gap between ML approaches and rule-based baselines is consistently large—rarely less than 10 percentage points on primary accuracy metrics—across all sub-domains where direct comparison is reported. This finding provides strong evidentiary support for the strategic investment in ML-based automation observed across major gateway operators.

Second, the magnitude of reported improvements varies substantially with dataset quality and evaluation methodology, complicating direct cross-study comparison. Studies using proprietary gateway datasets with rich network features and temporal context report substantially higher performance than studies using publicly available benchmarks (primarily the European credit card fraud dataset from Worldline and ULB), which lack the device, behavioural, and network feature dimensions of production systems. Practitioners should interpret published benchmark performance as a lower bound on achievable production performance given adequate feature engineering.

Third, the LLM and federated learning sub-domains are characterised by smaller study counts and shorter evaluation horizons than more established domains, reflecting the recency of production deployments. The high reported performance of LLM applications for dispute and compliance use cases should be contextualised against the limited duration of evaluation studies, which may not capture hallucination failure modes that manifest over longer deployment periods or less common edge cases.

MLOps and Production Deployment Infrastructure

A recurring finding across sub-domains is the critical importance of ML operations (MLOps) infrastructure—the systems and processes governing model deployment, monitoring, versioning, and retraining—as a determinant of production performance. Academic studies typically evaluate models at a single point in time on static datasets, systematically understating the operational complexity of maintaining high-performance ML systems in the adversarial, rapidly evolving payment environment.

The 21 studies in our corpus addressing MLOps in payment contexts identify model serving latency, monitoring for concept drift, feature pipeline reliability, and rapid retraining capability as the dominant infrastructure requirements. Leading gateway providers report serving architectures capable of 500,000+ model evaluations per second at sub-50ms P99 latency, requiring hardware acceleration (GPU inference, specialised ML chips) and feature precomputation strategies that are rarely discussed in academic algorithm-focused literature. The gap between academic ML research and production payment ML deployment constitutes a significant knowledge transfer challenge that the research community should address through closer industry collaboration.

Explainability Requirements and Implementation

Explainability emerges from the synthesis as the tension point where technical ML performance most directly conflicts with regulatory and ethical requirements. The 22 XAI studies in our corpus consistently find that post-hoc explanation methods—SHAP, LIME, counterfactual explanations—can provide useful decision attribution for individual payment decisions, but with two important limitations: computational overhead of 10–500ms for complex ensemble models is incompatible with real-time payment authorisation, and explanation fidelity (how accurately the explanation approximates the true model reasoning) ranges from 82–94% across studies, leaving a meaningful gap between explanations provided to consumers and true model behaviour.

The regulatory trajectory documented in the literature—particularly the EU AI Act's mandatory transparency requirements for high-risk AI systems including fraud detection—suggests that the current practice of providing post-hoc batch explanations for declined transactions will be insufficient for future compliance. Research into computationally efficient real-time explanation methods, and into inherently interpretable model architectures that sacrifice minimal accuracy, represents the most urgent technical research priority identified by this review.

Sociotechnical Dimensions of Payment Processing Automation

Employment and Labour Market Impacts

The automation of payment processing functions has substantive implications for employment in the financial services sector. Functions most exposed to automation include: document review and KYC analyst roles (IDP automation), AML transaction monitoring analyst positions (ML alert reduction), chargeback specialist positions (LLM-assisted dispute handling), and settlement and reconciliation operations staff (RPA bots). Industry estimates suggest that 35–50% of tasks in these roles are automatable with currently available technology, though the pace of actual role displacement is mediated by regulatory requirements for human oversight, organisational implementation capacity, and labour market dynamics.

The distributional implications of this displacement are geographically and demographically uneven. Payment operations roles—particularly those at offshore processing centres in India, the Philippines, and Eastern Europe—are disproportionately affected relative to higher-value analyst and engineering roles concentrated in financial centres. Within affected regions, workers in routine document processing and data entry roles face the greatest displacement risk, while workers with skills in exception handling, fraud investigation, and customer relationship management are more likely to transition into augmented roles alongside automation systems.

The research literature on payment sector automation employment impacts remains sparse relative to the manufacturing and retail automation literatures. Only 7 studies in our corpus directly address employment outcomes, and none employ longitudinal designs tracking actual worker

Table 4: Systematic Review Findings by Payment Processing Sub-domain

Research Sub-domain	*Studies (n)*	*Key Metric Range*	*Predominant Finding*
ML Fraud Detection	74	AUC 0.94–0.997	Ensemble and deep learning models consistently outperform rule-based systems; temporal features critical
Transaction Routing Optimisation	29	+3–9 pp auth rate	RL-based routing yields significant authorisation rate gains, especially for cross-border card-not-present
KYC / AML Automation	38	98%+ doc accuracy	IDP achieves near-human accuracy; AML precision improves 20–40× over rule engines
Biometric & Behavioural Auth	31	FAR <0.1%, FRR <2%	Behavioural biometrics provide continuous auth with minimal UX impact; suitable for SCA exemption
Explainability (XAI)	22	SHAP fidelity 82–94%	Post-hoc methods insufficient for real-time compliance; inherently interpretable architectures preferred
Fairness & Algorithmic Bias	18	FPR disparity up to 3.2×	Significant demographic disparities documented; fairness constraints reduce accuracy by 1–4%
Federated / Privacy-Preserving ML	19	Within 2–5% of centralised	Federated learning viable for fraud intelligence; differential privacy adds 1–3% accuracy cost
LLM Applications in Payments	11	60–75% time reduction	LLMs effective for dispute triage, compliance drafting; hallucination risk requires human-in-the-loop

Note: Study counts reflect primary focus; many studies address multiple sub-domains. Metric ranges represent the interquartile range of reported values across reviewed studies. FPR = False Positive Rate; FAR = False Acceptance Rate; FRR = False Rejection Rate; AUC = Area Under ROC Curve; pp = percentage points.

transitions. This constitutes a significant research gap with important policy implications—the design of retraining programmes, social protection systems, and regulatory automation transition requirements depends on evidence that does not currently exist in adequate form.

Algorithmic Fairness in Automated Payment Decisions

Automated payment decisions—fraud score-driven transaction declines, credit-based payment method eligibility, merchant onboarding decisions—carry significant consequences for individuals and businesses. The 18 fairness studies in our corpus document consistent evidence of demographic disparities in ML payment system outcomes. False positive rates—legitimate transactions incorrectly classified as fraudulent and declined—are systematically elevated for cardholders whose transaction patterns, device characteristics, or geographic profiles deviate from the training data distribution, which typically over-represents urban, formally employed, and technologically sophisticated population segments.

The magnitude of documented disparities is substantial: studies report false positive rate ratios between demographic groups ranging from 1.4× to 3.2× across race, age, geography, and income dimensions, with the highest disparities observed in cross-border transaction contexts where model training data is thin [39]. These disparities do not result from explicit use of protected characteristics as model features—most production fraud models exclude such attributes—but from correlated features that serve as proxies, including device type, geolocation precision, and spending pattern characteristics that reflect socioeconomic circumstances rather than fraud risk.

Mitigation strategies documented in the literature—fairness-constrained optimisation, post-processing calibration, adversarial debiasing—consistently demonstrate a fairness-accuracy trade-off: enforcing demographic parity or equal opportunity constraints reduces overall model AUC by 1–4%, representing a genuine conflict between maximising fraud detection performance and ensuring non-discriminatory treatment of legitimate customers. This trade-off cannot be resolved through technical means alone; it requires explicit policy decisions about acceptable fairness-accuracy trade-offs that must be made by gateway operators and their regulators, not by ML engineers.

Financial Inclusion and Access

The relationship between payment processing automation and financial inclusion is complex and bidirectional. On one side, ML-powered KYC automation has substantially reduced documentary requirements for merchant onboarding in markets where government-issued identification is inconsistently available, enabling micro-merchants and gig economy workers to accept digital payments through mobile-first gateway integrations. Alternative data sources—mobile money transaction histories, utility payment records, digital footprint signals—have enabled credit risk assessment for thin-file applicants previously excluded from formal payment system participation.

On the other side, AI-driven fraud models that generate higher false positive rates for population segments with atypical transaction patterns effectively create barriers to

Table 5: Research Gaps and Proposed Agenda for Automation and ML in Payment Processing

Research Gap	*Current State*	*Proposed Agenda*	*Priority Level*
Real-time XAI for payment decisions	Post-hoc only; 10–500 ms overhead	Approximate SHAP caching; inherently interpretable models	Critical
Intersectional fairness metrics	Binary protected-class focus	Multi-dimensional fairness frameworks for payment data	High
Production-scale federated learning	Lab / pilot deployments only	Cross-institutional FL benchmark datasets; standardised protocols	High
Adversarial robustness benchmarks	Fragmented; no payment-specific standard	Standardised adversarial evaluation framework for payment ML	High
LLM hallucination in compliance	Known risk; no payment-specific mitigations	Grounded RAG architectures with regulatory citation verification	Medium
Quantum-ready cryptographic migration	NIST standards published; migration nascent	Gateway-specific PQC migration roadmaps and test suites	Medium
Labour displacement impact studies	Sparse; mainly theoretical	Longitudinal studies on payment sector workforce transitions	Medium

Note: Priority levels reflect the combined assessment of evidence deficiency and practical urgency based on systematic review findings and emerging regulatory trajectories. RAG = Retrieval-Augmented Generation; PQC = Post-Quantum Cryptography; FL = Federated Learning.

digital payment participation for those groups. A merchant from a rural market in a developing economy processing their first international transaction—a profile rare in training data—is more likely to receive a fraud score-driven decline than an established urban merchant with a rich transaction history, despite equivalent legitimacy. This dynamic suggests that payment automation, without deliberate inclusive design, may reproduce and amplify existing financial exclusion rather than ameliorating it.

Research Gaps and Future Agenda

Identified Research Gaps

The systematic review process identified eight research gaps representing the most consequential areas of underdevelopment relative to the scale of their practical and social significance. These are summarised in Table 5, along with proposed research directions and priority assessments.

The most critical gap—rated by both evidence deficiency and practical urgency—is real-time explainability for automated payment decisions. Current XAI methods impose latency overheads incompatible with production payment authorisation requirements, creating a structural conflict between ML performance and regulatory explainability mandates that will intensify as the EU AI Act requirements phase in through 2026–2027. Research combining approximate explanation methods, pre-computed explanation caches, and inherently interpretable model architectures—evaluated on payment-specific datasets and latency budgets—is urgently needed.

Production-scale federated learning for payment fraud intelligence represents the second critical gap. While the theoretical framework is well-established and early experimental results are promising, no published peer-reviewed study documents a production-scale federated learning deployment across multiple payment institutions with adequate evaluation of privacy-utility trade-offs, communication efficiency at realistic data volumes, and Byzantine fault tolerance against poisoning attacks from compromised federation participants.

Emerging Technology Trajectories

Beyond documented research gaps, several emerging technology trajectories are likely to reshape the automation and ML landscape in payment processing over the 2026–2030 horizon. Quantum computing presents a dual-edged implication: the threat to current asymmetric cryptographic infrastructure securing payment channels, and the potential opportunity for quantum-enhanced optimisation in routing and portfolio risk.

Multimodal ML—models that simultaneously process text, image, and structured tabular data—enables richer KYC document understanding, dispute evidence analysis, and fraud context assessment than single-modality approaches. Early research demonstrates substantial performance improvements on document understanding tasks when visual document layout is combined with extracted text features, with implications for KYC automation quality in challenging document contexts.

Agentic AI systems—LLMs equipped with tool-use capabilities that enable multi-step, autonomous task completion—represent a qualitative expansion of automation potential beyond the document processing and text generation applications documented in this review. Payment compliance workflows requiring multi-system data retrieval, regulatory cross-reference, and structured output generation represent candidate applications for agentic systems, with pilot deployments emerging at major financial institutions in 2025–2026.

Conclusion

This systematic review has synthesised evidence from 242 peer-reviewed studies and authoritative industry sources on the application of automation and machine learning across payment processing systems, providing the most comprehensive evidence base on this topic published to date. The findings establish beyond reasonable doubt that ML and automation technologies deliver material, measurable improvements across the full payment processing value chain: fraud detection accuracy exceeding 99.2% AUC, authorisation rate improvements of 3–9 percentage points, KYC processing time reductions from days to hours, AML alert precision improvements of 20–40 times over rule-based baselines, and dispute handling time reductions of 60–75%.

The ML algorithm landscape has matured substantially over the review period. Gradient boosting remains the workhorse of production fraud scoring, while deep learning architectures—particularly transformer-based sequential models and graph neural networks—are gaining production deployment for applications requiring temporal sequence analysis and network-level fraud ring detection respectively. Reinforcement learning has established itself as the standard approach for transaction routing optimisation, and large language models are transitioning from research to production in compliance and dispute management applications.

References

[1] Manne, V. T. (2026). A Lightweight Application-Layer Defense Against Relay Attacks in Contactless Transactions.

[2] Mazumder, P. T. (2025). Blockchain in trade finance: reducing fraud and improving efficiency through digital ledger technology. Digital Finance, 7(4), 1043-1063.

[3] Tran, P. H., Tran, K. P., Huong, T. T., Heuchenne, C., Tran, P. H., & Minh, T. L. H. (2018). Real time data-driven approaches for credit card fraud detection. Proceedings of ICEBA 2018, ACM, 66–70.

[4] Mazumder, P. T. (2026). Explainable and fair anti-money laundering models using a reproducible SHAP framework for financial institutions. Discover Artificial Intelligence.

[5] Manne, V. T. (2025, December). AI-Powered Fraud Detection in Payments Using Long-Term Behavior Sequence Modeling. In *2025 International Conference on Computational Innovations and Sustainable Technologies (ICCIST)* (Vol. 1, pp. 1-7). IEEE.

[6] Parasa, M. (2025). SENTINEL-HCM: Federated Temporal Graph Intelligence for Detecting API Abuse, Data Leakage, and Policy Drift in SAP SuccessFactors Cloud Integrations. *Journal of Contemporary Science and Technology Management*, *1*(01), 51-82.

[7] Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: Unbiased boosting with categorical features. Advances in Neural Information Processing Systems, 31.

[8] Parasa, M. (2024). Architecting predictive workforce intelligence: A machine learning framework for attrition forecasting in SAP Success Factors. *Global Scientific and Academic Research Journal of Multidisciplinary Studies*, *3*(12), 212-221.

[9] Cheng, H.-T., Koc, L., Harmsen, J., Shaked, T., Chandra, T., Aradhye, H., ... & Shah, H. (2016). Wide & deep learning for recommender systems. Proceedings of ACM DLRS 2016, 7–10.

[10] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30.

[11] Hamilton, W. L., Ying, R., & Leskovec, J. (2017). Inductive representation learning on large graphs. Advances in Neural Information Processing Systems, 30.

[12] 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.

[13] Van Vlasselaer, V., Bravo, C., Caelen, O., Eliassi-Rad, T., Akoglu, L., Ghassemi, M., & Baesens, B. (2015). APATE: A novel approach for automated credit card transaction fraud detection using network-based extensions. Decision Support Systems, 75, 38–48.

[14] Zheng, P., Yuan, S., Wu, X., Li, J., & Lu, A. (2019). One-class adversarial nets for fraud detection. Proceedings of AAAI 2019, 33(1), 1286–1293.

[15] Dal Pozzolo, A., Caelen, O., Johnson, R. A., & Bontempi, G. (2015). Calibrating probability with undersampling for unbalanced classification. Proceedings of IEEE SSCI 2015, 159–166.

[16] Lucas, Y., Portier, P.-E., Laporte, L., He-Guelton, L., Caelen, O., Granitzer, M., & Calabretto, S. (2019). Towards automated feature engineering for credit card fraud detection using multi-perspective HMMs. Future Generation Computer Systems, 102, 393–402.