JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

AI and Forensic Accounting for Financial Fraud Prevention in MSMEs: A Systematic Literature Review

Mrs. Surbhi Akash Jain
Research Scholar
Place of Higher Learning and Research in Commerce, Hislop College, Nagpur, India
Affiliated to Rashtrasant Tukadoji Maharaj Nagpur University

Dr. Dini Menon
Associate Professor and Research Supervisor
Hislop College, Nagpur, India
Affiliated to Rashtrasant Tukadoji Maharaj Nagpur University

Corresponding author: Mrs. Surbhi Akash Jain, surbhijain2505@gmail.com

Abstract

Purpose. A single fraud can threaten the survival of a small firm, and small firms are the least likely to run the controls that catch it. Artificial intelligence (AI) and forensic accounting are both offered as remedies. This review asks what the published evidence shows about their effect on fraud in micro, small and medium enterprises (MSMEs), and where it stops.

Design. We followed PRISMA 2020. Two AI-assisted discovery tools (Consensus and SciSpace), citation chaining and Google Scholar returned 1,478 records published up to December 2025. After de-duplication, screening and eligibility assessment against a seven-criterion relevance rubric, 24 studies published between 2020 and 2025 were included. Each included study was then re-read at the fullest level of text we could obtain, which was the full text for 15 and the abstract or database record for nine. Findings were grouped into five themes and read against the Technology Acceptance Model and the Technology–Organization–Environment framework.

Findings. Random forest was the best classifier in both studies that compared algorithms on firm-level financial data, but at very different levels: 93.7% cross-validated accuracy on a panel of 341 Turkish SMEs, against 0.67 test accuracy on 593 Indonesian listed firms. Evidence that forensic accounting improves detection is perception data. The one regression of detection and prevention on an AI-based forensic system, among staff of Nigerian banks, found no significant effect. No included study measured fraud incidence or asset recovery in MSMEs, and only one SME study concerned fraud at all. Implementation cost was tested once, for general AI adoption among Saudi SMEs, and was not significant. Prior fraud experience was not tested in any study we could read in full. Knowledge and skills are the most consistently supported driver of adoption.

Originality. The review separates what has been measured from what is assumed, reports where the source studies are internally inconsistent, and sets out six research gaps and five propositions for work on Indian MSMEs.

Keywords: financial fraud; forensic accounting; artificial intelligence; MSMEs; technology adoption; systematic literature review

JEL classification: M41, M42, G32, O33

1. Introduction

The Association of Certified Fraud Examiners analysed 1,921 occupational fraud cases investigated between January 2022 and September 2023 and found that organisations with fewer than 100 employees accounted for 21% of them, with a median loss of USD 141,000 (ACFE, 2024). The same report shows how thin small-firm defences are. Only 27% of small organisations ran a hotline, against 82% of larger ones, and just 14% carried out formal fraud risk assessments. A loss of that size can end a firm that has no balance sheet to absorb it. Weak controls are the opportunity side of Cressey's (1953) fraud triangle, and they leave the other two sides, pressure and rationalisation, to do the rest.

MSMEs in India face this exposure while digitising quickly. Payments move through gateways, bookkeeping sits in cloud software and invoices are generated automatically. Each step removes a manual check and adds a point where a record can be altered, a payment redirected or a login phished. Large institutions can staff analytics teams to watch for such events. A manufacturer with twenty staff and a part-time accountant usually cannot.

Two bodies of work claim to help. The first applies machine learning, natural language processing and related techniques to transaction and financial-statement data, with the aim of flagging anomalies before losses mount. The second is forensic accounting, which brings investigative and evidentiary methods to an incident that has already occurred, and which increasingly relies on digital tools. Reviews exist for each. Existing systematic reviews of AI-based detection focus on algorithms for card and e-commerce fraud (Hafez et al., 2025; Mutemi & Bação, 2024) or on model taxonomies for financial networks (Sarna et al., 2025). Their settings are card systems, online retail and financial networks. Whether any of it applies to a small firm that cannot buy a platform, or hire an analyst to read the alerts, is left open.

This review was designed to close that distance. It asks four questions.

- **RQ1.** How effective are AI-based tools and forensic-accounting practices at reducing fraud and improving recovery?
- **RQ2.** Which factors, including cost, awareness and prior fraud experience, and which theoretical lenses explain adoption among small and medium firms?
- **RQ3.** What integrated AI–forensic frameworks exist, and how far have they been deployed and explained in practice?

- **RQ4.** Where are the regional, methodological and conceptual gaps, particularly for Tier-2 Indian MSMEs?

The review was prepared to position a doctoral study of MSMEs in the Nagpur region, a Tier-2 industrial centre in Maharashtra. Its conclusions are not limited to that setting. The central finding is a mismatch between what the literature claims and what it has measured, and that mismatch holds wherever small firms are the subject.

The paper proceeds as follows. Section 2 defines the constructs and the two adoption frameworks. Section 3 sets out the review method, including what was read for each study. Section 4 reports the selection outcome and the characteristics of the 24 included studies, and Section 5 synthesises them by theme. Section 6 answers the research questions and audits five propositions against the evidence. Section 7 states the research gaps with the evidence behind each, and Sections 8 and 9 cover limitations and conclusions.

2. Conceptual background

2.1 Constructs

We use *financial fraud* for intentional misstatement, misappropriation or manipulation of financial records or assets, including embezzlement, invoice fraud, financial-statement fraud and cyber-enabled theft. *AI-based fraud detection* covers supervised and unsupervised machine learning, deep learning, natural language processing and anomaly detection applied to financial or transactional data. *Forensic accounting* means the application of accounting, auditing and investigative skills to disputes and suspected fraud, including digital forensics for tracing funds and preserving evidence. The two are complementary in time. AI works mainly before and during an event by raising alerts; forensic accounting works mainly after one, by establishing what happened and building evidence that can be used in court. A firm could in principle use either, both or neither, and the empirical literature should therefore compare these cases. We return to this point in Section 6.

MSME follows the Indian statutory category of micro, small and medium enterprises. Most included studies refer instead to SMEs, and the two terms are used as the source studies use them.

2.2 Adoption frameworks

The Technology Acceptance Model explains individual uptake through perceived usefulness and perceived ease of use (Davis, 1989). It is a model of a single user's intention, and in an owner-managed firm the owner is often the only user who matters. The Technology–Organization–Environment (TOE) framework widens the view to the technology's attributes, the firm's resources and structure, and external pressure from competitors, customers and regulators (Tornatzky & Fleischer, 1990). Both appear repeatedly in the adoption studies we reviewed, and several combine them (Bin-Nashwan et al., 2024). We treat TAM as the lens for the owner's or manager's decision and TOE as the lens for the conditions that make adoption feasible, which is also how the doctoral study will use them. The unified model of Venkatesh et al. (2003) extends TAM with social influence and facilitating conditions, but none of the included studies relies on it, so it is not carried further.

3. Method

3.1 Design and reporting

The review follows the PRISMA 2020 statement (Page et al., 2021) and the general logic of evidence-informed management reviews (Tranfield et al., 2003). The research questions, eligibility criteria and scoring rubric are stated in full below. Where our practice fell short of the protocol we had set, we say so in the section concerned and again in Section 8.

3.2 Information sources

Four sources were used. Consensus and SciSpace are AI-assisted discovery tools that retrieve and rank papers from large bibliographic indexes; Consensus reports drawing on Semantic Scholar, IEEE Xplore, Scopus and Web of Science among others, and SciSpace on a corpus of more than 270 million papers. SciSpace also supports backward and forward citation chaining, which we used to find additional works from the core set. Google Scholar and grey literature (reports of the Reserve Bank of India, the Ministry of MSME and NITI Aayog) completed the search. Searches targeted 2015 to 2025, with 2020 to 2025 prioritised to reflect current practice. Searches were run in September 2026. The publication window closed at the end of December 2025 so that it covers whole calendar years, and work published in 2026 is therefore not covered (Section 8). Because the discovery tools rank results with proprietary methods, the retrieved sets cannot be regenerated exactly.

3.3 Search strategy

The synopsis of the doctoral study fixed a master Boolean string joining four concept groups: AI or machine learning; digital forensics, forensic accounting or cyber forensics; financial fraud or fraud detection; and small business terms (small businesses, SMEs, MSMEs). A fifth group listing qualitative methods (case study, thematic analysis, grounded theory, interpretive analysis) was part of the original string. Consensus and SciSpace accept natural-language questions rather than Boolean syntax, so the string was converted into one question: "integration of Artificial Intelligence and Digital Forensics/Forensic Accounting in financial fraud detection and prevention".

SciSpace turned the question into five search statements, the first being the question itself, and applied two filters of its own: publication in 2020 to 2025 with foundational work from 2015, and peer-reviewed journal articles. Three of the five statements lean towards graph neural networks, federated learning and anti-money-laundering, which helps explain why so many SciSpace records were later excluded as technical. SciSpace reported 664 records from these statements and 69 more from citation chaining.

Consensus ran 21 searches and one citation-graph query under six strategies: TAM and TOE foundations; SME and MSME settings with emphasis on India; terminology variants such as machine learning and digital forensics; adoption barriers (cost, awareness, prior experience); adjacent constructs such as risk management and regulatory compliance; and evidence on effectiveness. It identified 730 records. The tool reports the six strategies but not the wording of each of the 21 searches. Appendix C therefore reproduces the master string, the five SciSpace statements as the tool worded them, and the six Consensus strategies. It cannot list the 21 individual Consensus searches.

3.4 Eligibility criteria

Table 1 gives the criteria. A study could enter by one of two routes. The main route required that it apply AI, forensic accounting or both to the detection, prevention or investigation of financial fraud. The second route served RQ2 only. It admitted empirical TAM or TOE studies of AI adoption in SMEs, or in accounting and audit work, even where fraud was not the application, because the adoption question cannot be answered from fraud studies alone. The rubric scored this route under criterion C5, but an earlier statement of our criteria listed only the main route. We now state both, and we label the three studies admitted by the second route as adoption-context studies wherever they appear (Badghish & Soomro, 2024; Bin-Nashwan et al., 2024; Polisetty et al., 2023). They are not counted as evidence about fraud.

In addition, a study had to concern small firms or be transferable to them; be empirical, a review or a framework with a clear contribution; and be peer-reviewed, in English and published in 2015 or later. We excluded work on general cybersecurity, anti-money-laundering alone or credit scoring; studies that were large-firm or bank-only with no plausible transfer to small firms; opinion pieces; and papers whose main contribution was an algorithm or network architecture with no forensic or adoption angle.

Table 1. Eligibility criteria

Dimension	Inclusion	Exclusion
Topic, main route	AI and/or forensic accounting applied to financial-fraud detection, prevention or investigation	Generic cybersecurity, anti-money-laundering only, credit scoring or network security not centred on financial fraud
Topic, adoption-context route (RQ2 only)	Empirical TAM or TOE study of AI adoption in SMEs or in accounting and audit work	Adoption studies outside SMEs and outside accounting, audit or fraud work
Context	SME or MSME, or transferable to small-firm settings; emerging-economy or India relevance preferred	Large-corporation or bank-only studies with no small-firm transferability; developed-economy-only studies with no relevance
Method	Empirical, review or framework studies with a clear methodological or conceptual contribution	Opinion or editorial pieces; purely technical algorithm papers (for example graph neural network or deep-learning architectures) with no forensic or adoption lens
Quality and recency	Peer-reviewed or indexed; 2015–2025 (2020–2025 prioritised); English	Non-indexed venues; pre-2015 unless foundational; non-English; neither full text nor abstract retrievable

3.5 Selection process

Records were de-duplicated, including overlap between the two discovery tools, and screened on title and abstract. Reports that passed were sought in full text. Where the full text could be opened it was used. Where it could not, but an abstract or the discovery tool's structured summary was available, the report was assessed on that. Only reports with neither were counted as not retrieved. This is weaker than the full-text rule usual in systematic reviews, and it means that the eligibility stage in Figure 1 is not a full-text stage for every report.

Each assessed report was scored on a seven-criterion relevance rubric (Table 2). The scores operationalise the scope of the doctoral study, so that inclusion depends on explicit weights and not on impression. A report

was retained if it scored at least 55 of 100 and passed every eligibility criterion. Scores of 75 or more are labelled highly relevant, 68 to 74 relevant and 55 to 67 supporting.

The first author screened all records and scored all reports, in two passes. Cases she was unsure of were discussed with the second author, who also checked the final set. No second reviewer coded independently, so we cannot report an inter-rater statistic, and the scores should be read as one reviewer's structured judgement (Section 8).

Table 2. Relevance rubric (maximum 100)

Code	Criterion	Weight
C1	Relevance to AI in fraud detection or prevention	20
C2	Relevance to forensic accounting or digital forensics	20
C3	SME or MSME organisational context	15
C4	Emerging-economy, India or Tier-2 relevance	15
C5	Adoption factors and TAM/TOE alignment	10
C6	Empirical rigour and methodology	10
C7	Recency (2020–2025) and strength of evidence	10

The rubric deliberately rewards fit with the review's population and setting, so a methodologically excellent study of card fraud at a large bank can score below a modest survey of small firms. It is a relevance instrument and not a risk-of-bias tool. For that reason Section 4.3 adds a separate appraisal of study design, venue and internal consistency.

3.6 Re-reading of the included studies, extraction and synthesis

Before the synthesis was finalised, each of the 24 included studies was read again at source. The full text was read for 15. For eight, only the abstract or publisher record could be accessed (Akinbowale et al., 2023; Dako et al., 2020; Mehta et al., 2022; Narula & Bhagat, 2024; Naz & Khan, 2025; Odeyemi et al., 2024; Polisetty et al., 2023; Sathe, 2025). For Bin-Nashwan et al. (2024) we had only the title and the discovery tool's one-sentence summary. Table 3 records the basis for each study. Statements in Sections 5 to 7 that something is absent from the literature say whether they rest on the 15 studies read in full or on all 24.

For each study we extracted year, venue, design, setting, sample and respondents, the technology or practice examined, the outcome variable, and effect sizes or performance metrics where reported. We also recorded whether implementation cost and prior fraud experience were measured as variables, and any point at which a paper's text, tables and abstract disagreed. Journal quartiles were taken from the SCImago Journal Rank (SJR) field in the Consensus export where it was reported.

The studies are too heterogeneous in design and outcome for meta-analysis, and no pooled effect is reported. We used thematic synthesis. Each study was allocated to the theme that best matched its main contribution, and cross-theme evidence is flagged in the text. Finally, we audited five propositions derived from the doctoral study's hypotheses (Section 6.2) by recording, for each, whether any included study had tested it directly, indirectly or not at all.

4. Results

4.1 Study selection

The searches returned 1,478 records (Figure 1): 730 from Consensus, 664 from SciSpace, 69 from citation chaining and 15 from Google Scholar and grey literature. Removing duplicates, including overlap between Consensus and SciSpace, left 1,122 records for title and abstract screening, and 878 of them were excluded as off-topic for AI, forensic accounting or fraud. Of the 244 reports sought, 43 could not be retrieved in any form, so 201 were assessed against Table 1. Of these, 177 were excluded: 78 were technical or outside the scope (mostly network architectures and anti-money-laundering models), 34 concerned large firms only, 25 were not centred on the target kinds of fraud, 22 were opinion pieces without empirical or conceptual contribution, 10 related only to developed economies with no transferable lesson, and 8 were redundant or in unsuitable venues. Twenty-four studies were included. That is 1.6% of the records identified and 11.9% of the reports assessed. Appendix B lists sixteen illustrative exclusions with reasons.

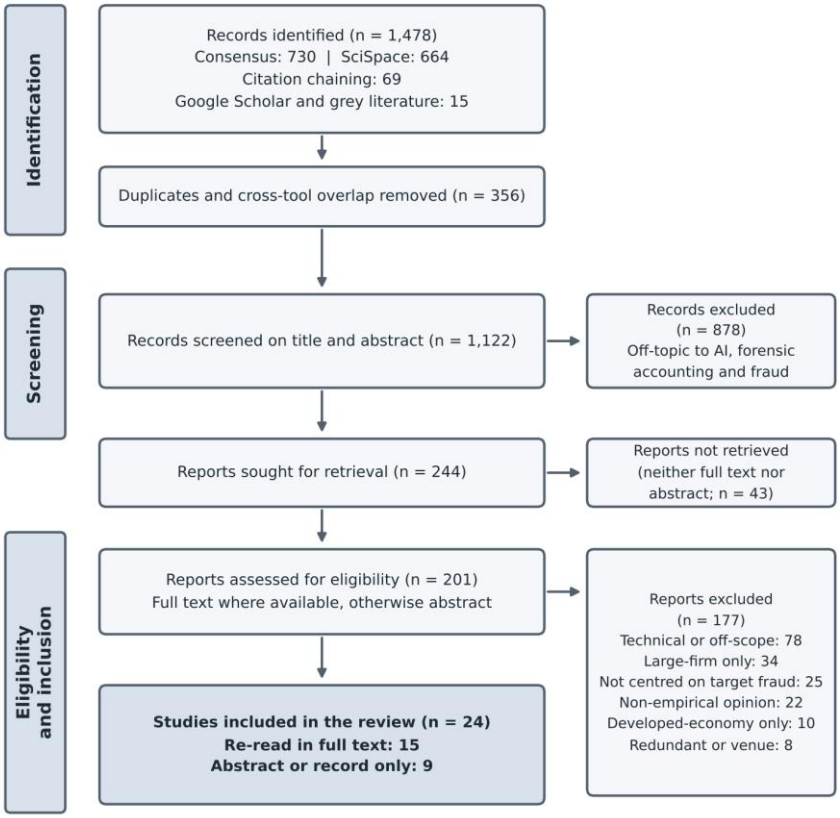

Figure 1. PRISMA 2020 flow diagram of study selection. Eligibility was assessed on the full text where it was available and otherwise on the abstract.

4.2 Characteristics of the included studies

Table 3 lists the 24 studies. Eighteen were published in 2024 or 2025, one in 2020, two in 2021, one in 2022 and two in 2023 (Figure 2). The recent weighting is partly a product of the 2020–2025 prioritisation in the search, but it also reflects a literature that has grown quickly while remaining thin on small firms.

By design, nine studies were quantitative surveys or regression studies, two used mixed methods (Okonta & Nnamdi, 2025; Polisetty et al., 2023), four evaluated classification models on data, one was a qualitative comparison built on secondary sources, five were literature reviews and three were conceptual. Ten studies

collected primary data from respondents. One analysed company data whose source is not stated (Ali et al., 2024), four tested models on data, and nine rested on secondary literature or argument alone.

Only three studies were set in SMEs: Hamal and Senvar (2021) in Turkey, Badghish and Soomro (2024) in Saudi Arabia and Polisetty et al. (2023) in India. Of those, only Hamal and Senvar concern fraud. The other two are adoption-context studies of AI in general, and Badghish and Soomro do not mention fraud anywhere in the paper.

By theme, the largest group of studies concerned integrated frameworks (T3, seven studies) and forensic accounting practice (T2, six), followed by adoption determinants (T4, five), detection performance (T1, three) and explainability, ethics and governance (T5, three).

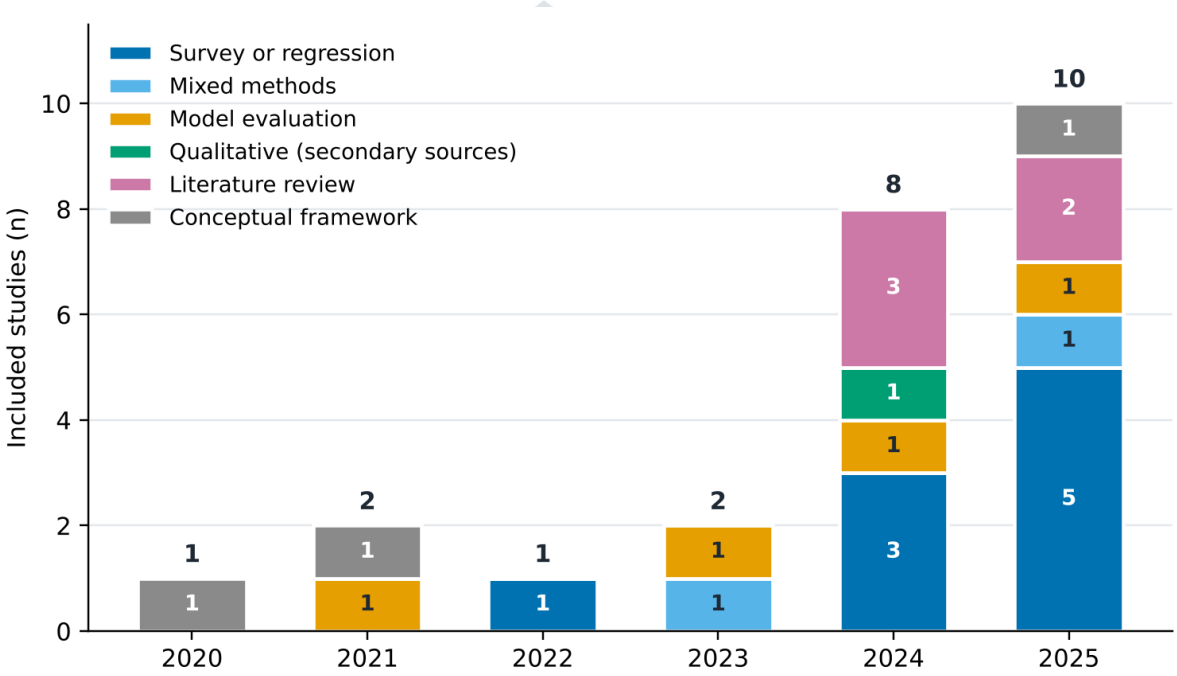

Figure 2. Included studies by publication year and study design.

Figure 3 cross-tabulates theme against design. The evidence is lopsided. The five adoption-determinant studies (T4) are four surveys and one mixed-methods study, and all three explainability studies (T5) are reviews, so no study of explainability collected data from firms or users. The three detection-performance studies (T1) are all model evaluations, and only one integrated-framework study (Akinbowale et al., 2023) evaluates a model. The empty cells matter as much as the filled ones. There is no qualitative field study of adoption, and the single qualitative study in the set works from documents, not interviews.



Figure 3. Evidence map of the 24 included studies: theme by study design (number of studies per cell; hollow circles mark empty cells).

Figure 4 shows where the studies were set. Twelve of the 24 are not tied to one country, usually because they are conceptual papers, reviews, or studies that do not say where their data came from. Of the twelve that name a setting, ten are tied to one country (or, for Yaseen and Al-Amarneh, two Gulf states) and two compare countries: Devisa and Purwanto (2025) cover the United States, the United Kingdom, China, Indonesia and Nigeria, and Simbolon et al. (2024) contrast developed and developing countries. India appears twice. Polisetty et al. (2023) surveyed managers of B2B SMEs about AI in general, and Goyal et al. (2025) surveyed bank employees about AI tools for credit risk and fraud. Neither studies fraud technology in small firms.

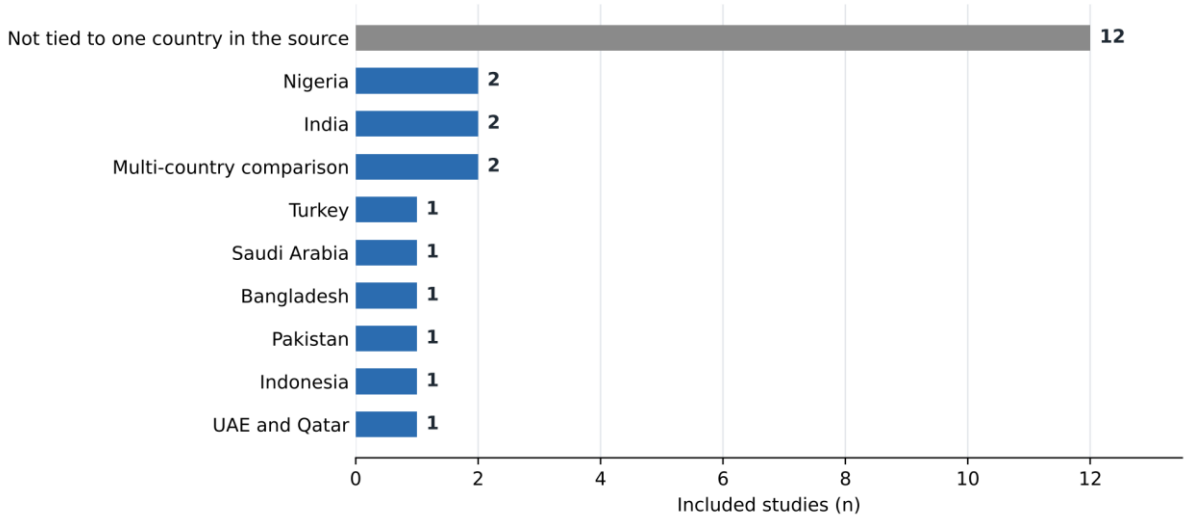

Figure 4. Settings of the 24 included studies, by country named in the source (studies not tied to one country are shown in grey).

Table 3. Characteristics of the 24 included studies

Study	Year	Design	Setting and sample (as reported)	Theme	Text read	SJR	Score
Akinbowale et al.	2023	Model evaluation	Simulated banking data; two-layer neural network	T3	Abstract	n/r	77
Ali et al.	2024	Regression on company data	100 companies (analytics, AI); 50 (blockchain); country not stated	T3	Full	Q2	67
Badghish & Soomro	2024	Survey, PLS-SEM	220 Saudi SMEs (115 medium, 105 small); AI adoption in general	T4*	Full	Q1	82
Bhardwaj	2025	Systematic review	162 Scopus publications on explainable AI, 2010–2025	T5	Full	n/r	60
Bin-Nashwan et al.	2024	Survey	Accounting and auditing firms; AI in reporting and audit	T4*	Record	Q1	72
Dako et al.	2020	Conceptual	Forensic accounting frameworks in emerging markets	T2	Abstract	n/r	73
Devisa & Purwanto	2025	Literature review	40 articles; US, UK, China, Indonesia, Nigeria	T2	Full	n/r	74
Elemide & Araoye	2025	Survey, regression	232 staff of Nigerian banks	T3	Full	n/r	85
Goyal et al.	2025	Survey, SEM–ANN	271 Indian bank employees who use AI tools	T4	Full	n/r	71
Hamal & Senvar	2021	Model evaluation	341 Turkish SMEs, 2013–2017; 321 fraud and 1,384 non-fraud observations	T1	Full	Q2	76
Kalluri	2024	Model evaluation	No data set described; self-reported 98.7% accuracy	T1	Full	n/r	65
Lee et al.	2025	Model evaluation	593 Indonesian listed firms; 2,373 firm-years, 2015–2023	T1	Full	Q2	69
Mehta et al.	2022	Survey, SEM	AI as mediator between forensic accounting and detection; sample not given in the abstract	T3	Abstract	n/r	79
Narula & Bhagat	2024	Review (chapter)	AI challenges in detecting financial statement fraud	T5	Abstract	n/r	66
Naz & Khan	2025	Survey	Pakistan; professionals in investigating firms and professional bodies	T2	Abstract	Q1	76
Odeyemi et al.	2024	Review	Forensic accounting techniques	T2	Abstract	n/r	64
Odonkor et al.	2021	Conceptual	AI–forensic framework; no data	T3	Full	n/r	65
Okonta & Nnamdi	2025	Mixed methods	Nigeria (Asaba); 150 professionals targeted; the abstract says documentary	T3	Full	n/r	78
Polisetty et al.	2023	Mixed methods	866 managers of Indian B2B SMEs; AI adoption in general	T4*	Abstract	Q1	73
Saha et al.	2025	Survey	Bangladesh; 400 finance, audit and regulatory professionals	T2	Full	n/r	81
Sathe	2025	Conceptual	AI anomaly detection in forensic accounting	T3	Abstract	n/r	57
Simbolon et al.	2024	Qualitative, secondary sources	Developed and developing countries compared	T2	Full	n/r	70

Study	Year	Design	Setting and sample (as reported)	Theme	Text read	SJR	Score
Syahfir et al.	2024	Systematic review	20 Scopus articles, 2020–2024; public sector	T5	Full	n/r	66
Yaseen & Al-Amarneh	2025	Survey, SEM	409 bank professionals, UAE and Qatar	T4	Full	Q2	74

Notes. Themes: T1 detection performance; T2 forensic accounting practice; T3 integrated AI–forensic frameworks; T4 adoption determinants; T5 explainability, ethics and governance. An asterisk marks an adoption-context study admitted for RQ2 that does not concern fraud. Text read: Full, full text; Abstract, abstract or publisher record only; Record, title and discovery-tool summary only. SJR is the SCImago quartile reported in the Consensus export; n/r means not reported there. Score is the relevance total from Table 2 (maximum 100).

4.3 Quality appraisal and sensitivity analysis

The relevance totals ranged from 57 to 85, with a mean of 71.7. Eight studies scored 75 or more (highly relevant), eight scored 68 to 74 (relevant) and eight scored 55 to 67 (supporting). Empirical studies had a higher mean score on the rigour criterion (C6) than reviews and conceptual papers, 9.0 against 6.6 out of 10. Only six studies reached 10 or more of 15 on the SME-context criterion (C3): Badghish and Soomro (2024), Hamal and Senvar (2021), Saha et al. (2025), Naz and Khan (2025), Dako et al. (2020) and Polisetty et al. (2023). The remaining eighteen are relevant to small firms by transfer, not by setting. Figure 5 shows the full pattern.

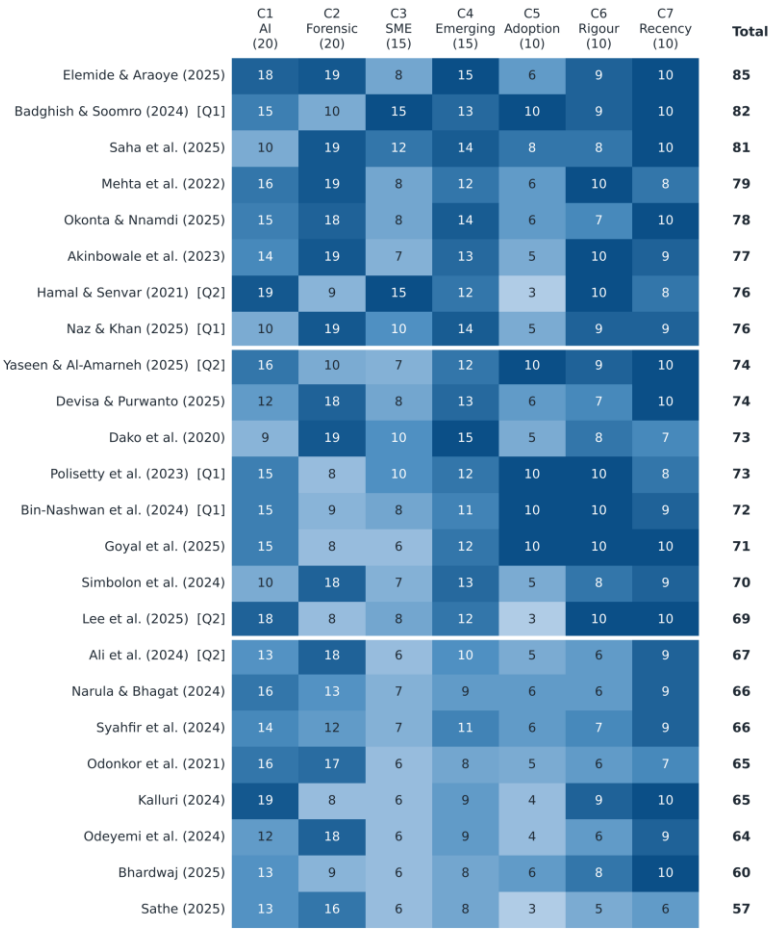

Figure 5. Relevance rubric scores of the 24 included studies by criterion. Shading shows the score as a share of the criterion maximum. Q1 and Q2 mark the SJR quartile where reported, and white lines separate the highly relevant, relevant and supporting tiers.

Scoring of the adoption-context studies. The three adoption-context studies each received 15 of 20 on C1 and 8 to 10 of 20 on C2, although none applies AI to fraud. Those scores credited general relevance to AI in

accounting and are too generous for criteria worded around fraud and forensic work. We therefore re-ran the threshold with stricter scoring. With C1 set to zero, the three totals fall to 67, 58 and 57, and all three stay above 55. With C1 and C2 both set to zero, Badghish and Soomro stay in at 57, while Polisetty et al. and Bin-Nashwan et al. fall to 50 and 48 and would be excluded. Dropping those two would leave T4 with three studies. It would not change any conclusion below, because neither is used as evidence about fraud, and it would remove the only study for which we had no abstract.

Venue. The Consensus export reported an SJR quartile for eight of the 24 studies: Q1 for Badghish and Soomro (2024), Polisetty et al. (2023), Bin-Nashwan et al. (2024) and Naz and Khan (2025), and Q2 for Hamal and Senvar (2021), Lee et al. (2025), Ali et al. (2024) and Yaseen and Al-Amarneh (2025). No quartile was reported for the other sixteen. Absence from that field does not prove that a journal is unindexed, and we did not verify indexing journal by journal. We treat the sixteen as unrated.

Internal consistency. Reading the papers in full showed problems that the abstracts hide, including in rated venues. Ali et al. (2024) describe their sample in three different ways (100 companies, 200 forensic accounting firms, 300 surveyed accountants), list companies under placeholder names, and report regression coefficients without p-values or R^2. Kalluri (2024) names no data set and no algorithm, and the methods section lists laboratory techniques that have nothing to do with fraud detection. In Lee et al. (2025) the confusion matrices for the five algorithms imply different numbers of fraud cases in what is described as the same test set. Badghish and Soomro (2024) describe one multi-group difference as significant while reporting $p = 0.250$ for it. The abstract of Okonta and Nnamdi (2025) describes a documentary study, but the body reports a questionnaire and interviews without stating how many people responded. Saha et al. (2025) attach the same percentages to different survey items in different sections. Devisa and Purwanto (2025) state that 40 articles were reviewed but do not say how they were chosen, and Simbolon et al. (2024) mention expert interviews in the abstract that the method section does not describe. We kept these studies, because excluding them would hide how weak parts of this literature are, but we weight them accordingly in Section 5.

Sensitivity. We checked the main conclusions against two narrower sets. The first is the eight studies in rated venues. It covers four of the five themes and has no study on explainability. The second is the 15 studies read in full, which cover all five themes. In both sets the conclusions of Section 6 hold in the same direction: no study measures fraud incidence or asset recovery in MSMEs, and none of the studies read in full tests prior fraud experience as a driver of adoption. The gaps reported below are therefore not an artefact of weak venues or of abstract-only reading. The positive claims are more exposed. The 98.7% accuracy in Kalluri (2024) cannot be verified from the paper and is reported as a claim.

5. Thematic synthesis

5.1 T1: Detection performance of AI models

Three studies evaluate classifiers directly. Hamal and Senvar (2021) offer the closest match to the review's population. They worked with 341 Turkish SMEs observed between 2013 and 2017. The data set holds 1,384 non-fraud and 321 fraud observations, the latter from 122 firms, so fraud makes up 18.8% of cases. They described each observation with 32 financial ratios and compared seven classifiers (support vector machine,

naive Bayes, artificial neural network, k-nearest neighbours, random forest, logistic regression and bagging) under two resampling schemes, SMOTE oversampling and undersampling, and two feature-selection methods. Models were assessed by ten-fold cross-validation. Random forest with SMOTE and no feature selection came out best, with accuracy of 93.7%, sensitivity of 94.7%, specificity of 92.8%, precision of 92.5% and an area under the ROC curve of 0.984. The same classifier with undersampling reached 87.1% accuracy, bagging was second, naive Bayes was among the weakest, and feature selection did not improve any model. Two limits apply. The paper does not report a held-out test set, so the figures are cross-validated estimates on resampled data. And the comparison was among machine-learning algorithms. It did not set them against manual review or conventional audit procedures, so the study shows which classifier to prefer, not that classifiers beat existing practice.

Lee et al. (2025) worked with non-financial companies listed on the Indonesia Stock Exchange from 2015 to 2023, 2,373 firm-years from 593 companies, split into 1,898 training and 475 test observations. They compared logistic regression, k-nearest neighbours, a support vector machine, a decision tree and random forest. Random forest was again best on the test set, but the level is modest: precision 0.80, recall 0.67, accuracy 0.67 and F1 0.71. Logistic regression was close behind at 0.65 accuracy. The decision tree scored 1.00 on every training metric and 0.61 on the test set, and k-nearest neighbours fell from 0.80 to 0.61, which is the overfitting the authors report. The paper does not state how many firm-years were labelled as fraud. The label appears to be derived from the Beneish M-score and not from detected cases, so the models are partly learning to reproduce a screening formula.

Kalluri (2024) reports 98.7% accuracy against 75–85% for conventional systems and processing of more than five million transactions per second. The paper names no data set, no algorithm and no validation design, so we include the figures as a claim and give them no weight.

Across the two usable studies, random forest wins both comparisons, and class imbalance and overfitting are the recurring problems. The gap between 93.7% and 0.67 is the more useful finding. Performance depends heavily on how fraud is labelled and how the model is validated, and a model tuned on a Turkish SME panel has not been shown to work on Indian MSME books, where accounting standards, document types and informal practices differ.

5.2 T2: Forensic accounting practice

Six studies describe forensic accounting rather than test it. Saha et al. (2025) surveyed 400 professionals in Bangladesh, drawn by stratified sampling from bankers (90), accountants (80), auditors (70), regulators (60), corporate executives (50), forensic accountants (30) and academics (20). Of these, 82% regarded financial crime as prevalent and 89% believed forensic accounting improves detection. The barriers they named were a shortage of trained forensic accountants (68%), weak regulatory enforcement (63%), inadequate technological infrastructure (56%) and the expense of forensic services (47%). The paper reports percentages only, gives no breakdown by firm size, and concedes that smaller organisations may not be represented.

Naz and Khan (2025) sent structured questionnaires to professionals in investigating firms and professional bodies in Pakistan. On the abstract, which is all we could read, fraud investigation, litigation support and

dispute resolution each had a positive effect on fraud detection and prevention. The abstract gives no sample size or coefficients. Simbolon et al. (2024) compare developed and developing countries through a thematic analysis of published articles and reports. They conclude that countries with stronger regulation integrate forensic accounting more deeply, and they name cost and limited funding among the barriers in developing countries. Their statement that strong frameworks go with lower fraud is not backed by figures. Devisa and Purwanto (2025) review technology in forensic work across five countries and place Indonesia and Nigeria well behind the United States, the United Kingdom and China, citing limited infrastructure, immature regulation and low digital literacy. Dako et al. (2020) and Odeyemi et al. (2024), read at abstract level, argue the case conceptually or by review.

The shared conclusion is that forensic accounting is valued, under-used and limited by skills and enforcement. What the six do not provide is a comparison of firms that use it with firms that do not, or any figure for recovered losses. The claim that forensic accounting improves detection therefore rests on what practitioners and respondents believe.

5.3 T3: Integrated AI–forensic frameworks

Seven studies bring the two tools together, and they differ widely in what they test.

Two report statistical tests on survey data, and they point in different directions. Mehta et al. (2022) used structural equation modelling and found that AI tools mediate the effect of forensic accounting on the detection of financial fraud. Only the abstract was available to us, and it does not give the sample or the respondents, so we cannot say whose perceptions were modelled. Elemide and Araoye (2025) distributed 250 questionnaires to forensic accounting, internal audit, compliance and IT security staff in Nigerian commercial, merchant and microfinance banks and received 232 (92.8%). They regressed perceived fraud detection and perceived fraud prevention on a scale describing the banks' AI-based forensic system. Neither relationship was significant (detection: $\beta = 0.077$, $p = 0.172$, $R^2 = 0.008$; prevention: $\beta = 0.058$, $p = 0.352$, $R^2 = 0.004$). Respondents agreed strongly that the AI systems have flaws, such as false positives and slow adaptation to new fraud patterns (mean 4.18 on a five-point scale). The authors conclude that AI cannot manage fraud on its own and recommend combining it with human forensic judgement, explainable models and training.

The remaining five carry less. Akinbowale et al. (2023) trained a two-layer neural network on simulated banking data and report that it classified about 95% of cases correctly; we read the abstract only. Okonta and Nnamdi (2025) targeted 150 forensic accountants, IT experts and law-enforcement officers in Asaba, Nigeria, with a questionnaire and interviews. They report that 69% of respondents saw higher detection accuracy with AI, that 60% cited financial constraints and that 50% cited the need for specialised training, but they give no response count and no test statistics. Ali et al. (2024) regress "fraud detection efficacy" on data analytics ($\beta = 0.35$), AI ($\beta = 0.30$), blockchain ($\beta = 0.25$) and professional training ($\beta = 0.20$). The paper gives no p-values or R^2, does not say which country the companies are in, and describes the sample inconsistently (Section 4.3), so the coefficients cannot be interpreted. Odonkor et al. (2021) and Sathe (2025) propose

frameworks without testing them. Odonkor et al. do note that implementation cost can be prohibitive for smaller organisations and suggest subscription services and pilot projects.

Three points follow. First, the one study that reports full regression output found no effect, and it was done in banks, whose compliance staff and data volumes are far from those of a small manufacturer. Second, every outcome in this theme is a perception or a model score on test data. Third, nobody has compared AI alone, forensic accounting alone and the two together in the same sample.

5.4 T4: Adoption determinants

This theme has the strongest designs and is closest to the doctoral study. It also needs the most care, because three of its five studies are adoption-context studies that do not concern fraud.

Badghish and Soomro (2024) surveyed 220 SMEs in the Jeddah industrial area, 115 medium and 105 small, and modelled AI adoption with TOE and partial least squares. Market and customer demand had the largest effect ($\beta = 0.432$), followed by sustainable human capital ($\beta = 0.223$), relative advantage ($\beta = 0.127$), compatibility ($\beta = 0.090$) and government support ($\beta = 0.080$). Implementation cost, measured with items such as "Implementing an AI-based system is very expensive", was not significant ($\beta = 0.042$, $p = 0.204$), and neither were complexity and organisational support. Adoption was associated with operational ($\beta = 0.557$) and economic performance ($\beta = 0.672$), and the link to economic performance was stronger in medium than in small firms ($p = 0.047$). This is the only included study that tests cost as a variable. The technology is AI in general, and the paper never mentions fraud.

Polisetty et al. (2023) studied 866 managers of Indian B2B SMEs with a mixed-methods design built on TOE. From the abstract, most AI enablers raised AI readiness, perceived benefits and role clarity did not, and AI ethics moderated several relationships. We could not check from the abstract whether cost was among the enablers. For Bin-Nashwan et al. (2024) we had no abstract. The discovery tool's summary states that AI adoption in accounting and auditing firms improves reporting accuracy and audit efficiency and that data governance strengthens the effect. We cite it for the combination of TOE, TAM and resource dependence theory in its title and draw nothing further from it.

Two studies concern the adoption of fraud technology itself, both in banks. Yaseen and Al-Amarneh (2025) surveyed 409 auditors, compliance officers, risk managers and AI specialists in banks in the UAE and Qatar. Trust was the strongest predictor of adoption ($f^2 = 0.31$). Algorithmic bias lowered adoption ($\beta = -0.21$), and most of that effect ran through perceived fairness (indirect effect $\beta = -0.14$). Cost was not in the model. Goyal et al. (2025) surveyed 271 employees of Indian banks who already used AI tools for credit risk and fraud detection, and analysed the data with PLS-SEM and a neural-network stage. Knowledge of the technology had the largest direct effect on continued use ($\beta = 0.325$), followed by perceived usefulness ($\beta = 0.197$). Perceived ease of use shaped attitude ($\beta = 0.504$) but had no direct effect on continued use. Awareness ($\beta = 0.047$, $p = 0.417$) and perceived risk ($\beta = -0.049$, $p = 0.325$) were not significant. Neither cost nor prior fraud experience was measured; cost appears only in the authors' discussion of public-sector bank funding.

Taken with the other themes, the pattern on the three factors the doctoral study cares about is as follows. Knowledge and skills are supported wherever they are tested or asked about: Goyal et al. and Badghish and

Soomro in models, Saha et al. and Okonta and Nnamdi in descriptive percentages. Awareness, as distinct from knowledge, was tested once and was not significant. Cost is named as a barrier by respondents in two descriptive surveys (60% in Okonta and Nnamdi, 47% in Saha et al.) and by several conceptual papers, but the single statistical test found no effect, and that test was not about fraud technology. Prior fraud experience was not a variable in any of the 15 studies read in full. Yaseen and Al-Amarneh used experience in fraud or compliance work as a condition for taking part, which is a different thing. Nothing in the abstracts of the other nine suggests that they measured it, but we cannot rule that out.

5.5 T5: Explainability, ethics and governance

Three reviews address whether a flagged transaction can be explained to the person who must act on it. Bhardwaj (2025) reviewed 162 Scopus-indexed publications on explainable AI in finance from 2010 to 2025. Post-hoc methods, mainly SHAP and LIME, are used in about three-quarters of the empirical studies, tests of whether people actually understand the explanations remain rare, and small firms are not treated as a category. Syahfir et al. (2024) reviewed 20 Scopus articles from 2020 to 2024 on AI and fraudulent financial reporting in the public sector, and name weak infrastructure and a shortage of skilled staff as the main obstacles, alongside privacy and bias. Narula and Bhagat (2024), read at abstract level, discuss the challenges of AI in detecting financial statement fraud.

The relevance for small firms is practical. An owner who receives an unexplained alert has to decide alone whether to freeze a payment or confront an employee. None of the three reviews examines explainability in a small-firm workflow. Two empirical studies in other themes suggest it matters: transparency and fairness shaped adoption among bank professionals (Yaseen & Al-Amarneh, 2025), and the bank staff surveyed by Elemide and Araoye (2025) rated opacity among the flaws of their AI systems. Both are bank studies, so the connection to small firms is plausible and unmeasured.

6. Discussion

6.1 Answers to the research questions

RQ1: effectiveness. The included studies show that classifiers can separate fraud-labelled from other records in a test sample, at levels that range from 67% to 93.7% accuracy depending on the data and the validation, and that practitioners expect forensic accounting to help. They do not show that either reduces fraud or recovers assets in MSMEs. Accuracy is a property of a model on a data set, and perception is a property of a respondent. Neither is a loss figure. Where perceived effectiveness was regressed on an AI-based forensic system with full reporting, in Nigerian banks, the effect was not significant (Elemide & Araoye, 2025).

RQ2: adoption. Knowledge and skills are the best-supported factor, followed by trust, perceived fairness and perceived usefulness in the two bank studies of fraud technology, and by market demand and human capital in the one SME study of AI in general. Cost was tested once, in that SME study, and was not significant. It has not been tested for fraud technology, and prior fraud experience has not been tested in any study we read in full. TAM and TOE are the working frameworks.

RQ3: integrated frameworks. Seven studies propose or test combinations of AI and forensic methods. Two report statistical tests, one positive through mediation (Mehta et al., 2022) and one null (Elemide & Araoye, 2025). The others are a model trained on simulated data, a mixed-methods study with incomplete reporting, a regression that cannot be interpreted and two conceptual proposals. Explainability is treated in a separate stream (T5) and has not been connected to the integrated frameworks in any study.

RQ4: gaps. Section 7 sets them out with the evidence behind each. The common thread is that evidence from small firms, from India and from outcomes is missing, while evidence from banks, other countries and algorithm performance is plentiful.

6.2 Audit of the doctoral study's propositions

The doctoral study states five hypotheses. We restated each as a proposition about the literature and recorded what the included studies say (Table 4).

Table 4. Propositions audited against the included studies

Proposition (hypothesis)	Evidence in the 24 studies	Status
P1 (H1). AI adoption reduces fraud incidence in MSMEs	Classification metrics on firm data (Hamal & Senvar, 2021; Lee et al., 2025). Perceived detection, positive through mediation in Mehta et al. (2022) and not significant in Elemide and Araoye (2025). No study measures incidence in MSMEs	Indirect and mixed
P2 (H2). Forensic accounting improves detection and recovery	Perceived improvement in detection (Saha et al., 2025; Naz & Khan, 2025). Recovery is not examined	Detection supported by perception only; recovery untested
P3 (H3). High cost lowers adoption	Tested once, for AI in general among Saudi SMEs, and not significant (Badghish & Soomro, 2024). Named as a barrier by 60% of respondents in Okonta and Nnamdi (2025) and 47% in Saha et al. (2025)	One null test outside fraud; untested for fraud technology
P4 (H4). Awareness and skills raise adoption	Knowledge of the technology (Goyal et al., 2025) and human capital (Badghish & Soomro, 2024) are significant. Training shortages are reported by Saha et al. (2025) and Okonta and Nnamdi (2025). Awareness was not significant in Goyal et al. (2025)	Skills supported; awareness not supported in its one test
P5 (H5). Prior fraud experience raises adoption	Not a variable in any of the 15 studies read in full; no sign of it in the other nine abstracts	No evidence

The pattern is useful to a doctoral design. P4 can be tested with an established toolkit, and the split between knowledge and awareness in Goyal et al. suggests the two should be measured separately. P3 is more open than it looks: respondents say cost matters, and the one model that included it found that it did not. P1, P2 and P5 are open, and an empirical study of Nagpur MSMEs would add something new if it measured them directly. A null result on P5 would be informative as well, because the literature has simply assumed the opposite.

6.3 Implications

For theory, the adoption studies treat TAM and TOE as general technology frameworks. None of those we read in full conditions adoption on a firm's own experience of fraud, although fraud is the reason the technology exists. Adding fraud exposure and perceived risk as antecedents would connect the adoption frameworks to the fraud triangle (Cressey, 1953) and make P5 testable. The evidence also shows a break between adoption and outcome. Some studies explain who adopts and others measure whether classifiers work, but none follows the same firms from adoption to losses.

For practice, the evidence supports modest conclusions. Knowledge and skills are the factor documented most consistently, so training and advisory support for owners and accountants are better supported than platform subsidies. The case for subsidies is weaker than is often assumed, since the only test of cost found no effect, although that test did not concern fraud tools and should not be over-read. The studies that point to hybrid designs (Elemide & Araoye, 2025; Mehta et al., 2022) suggest that tools which keep a human reviewer in the loop suit firms without analysts. For policymakers, the most useful step may be to collect incident and recovery data from MSMEs in a consistent way, since the literature cannot currently say how much fraud small firms lose or recover.

7. Research gaps and agenda

7.1 Six gaps and the evidence behind them

The gaps below are derived from the 24 included studies. Table 5 states, for each, what the studies contain and which of them were checked in full text. The scoping matrices in the doctoral synopsis and in the Consensus report, which earlier framed some of these gaps, are reproduced in Section 7.3 for transparency. They were generated at the scoping stage, were not re-coded by us, and are not used as evidence here.

Table 5. Research gaps and the evidence for each in the included studies

Gap	What the included studies contain	Basis
G1. No Tier-2 India MSME study of fraud prevention	Two Indian studies: SME managers on AI in general (Polisetty et al., 2023) and bank employees on AI credit-risk and fraud tools (Goyal et al., 2025). One SME study of fraud anywhere (Hamal & Senvar, 2021, Turkey)	All 24; settings checked in full text for 15
G2. Outcomes are unmeasured	Outcomes are classifier metrics or perceptions in every empirical study. None reports incidence, loss, time to detection or recovery for MSMEs	All 24
G3. Cost and prior fraud experience are untested for fraud technology	Cost: one null test for AI in general (Badghish & Soomro, 2024); descriptive percentages in two surveys. Prior fraud experience: not a variable in any study	15 read in full; nine not verifiable
G4. No comparison of AI alone, forensic accounting alone and both	Seven integrated-framework studies; none has comparison groups	All 24
G5. Explainable AI in small-firm workflows	Three reviews, none with small-firm or user data; Bhardwaj (2025) finds tests of human understanding rare in finance generally	All three T5 studies; two read in full
G6. Field evidence on blockchain	One descriptive table for 50 companies without a statistical test (Ali et al., 2024); blockchain is otherwise a proposed component (Odonkor et al., 2021) or a feature of developed-country practice (Devisa & Purwanto, 2025)	Read in full

G1. No Tier-2 India MSME study of fraud prevention. Studies of Tier-2 cities, where MSMEs cluster and advisory services are thinner, would show whether findings from Saudi Arabia, Turkey or Nigeria transfer.

G2. Outcomes are unmeasured. Fraud incidence, loss size, time to detection and asset recovery should be measured before and after adoption, or compared between adopters and non-adopters. Even a small panel of firms with accounting records would be a step beyond perception surveys.

G3. Cost and prior experience are untested for fraud technology. Survey instruments with direct items for the cost of fraud tools and for past fraud events would resolve this. Because we could not read nine studies in full, this gap is stated for the 15 we could.

G4. No comparison of AI alone, forensic accounting alone and both. The question of whether the combination adds anything is open. A quasi-experimental or matched design could answer it.

G5. Explainable AI in small-firm workflows. The question is whether an owner-manager can act on an alert. Usability studies with owners and accountants, rather than model-level explainability metrics, are needed.

G6. Field evidence on blockchain and other emerging technologies. The evidence on small-firm use is thin and mostly conceptual.

Methodologically, the field would benefit from shared fraud-case data sets for SMEs, fraud labels based on detected cases and not on screening scores, pre-registration of hypotheses in survey work, and complete reporting of test statistics. Several of the included papers would not have passed a check of their own tables

against their own text. Theoretically, TAM and TOE could be complemented by a fraud-specific lens. The fraud triangle (Cressey, 1953) is used in the forensic accounting stream but never linked to the adoption frameworks, although a firm's perceived exposure to fraud should plausibly shape its readiness to invest.

7.2 Term map of the wider Consensus set

To show how the surrounding literature is organised, Figures 6 and 7 map the terms in a wider set of 50 records. The set needs explaining, because it is neither the 730 records Consensus identified nor our 24 included studies. Consensus filtered its 730 records to 240 screened and 164 eligible, ranked them, and exported the 50 it ranked most relevant with title, abstract, a one-sentence summary and study type. Those 50 exported records are the input to the maps. SciSpace exported 22 records in a different layout without comparable summary fields, so they were left out. The maps therefore describe one tool's top-ranked set and are context for the review, not part of its evidence.

Figure 6 was produced in Python in the manner of a VOSviewer term map, and not with the VOSviewer program. We counted a fixed list of 31 domain terms in each record's title, abstract and summary and kept the 29 that appeared in at least five records. The two search concepts that appear in most records (fraud detection and artificial intelligence, 43 records each) were then removed, as is usual in term maps, because they would sit at the centre of every cluster, which left 27 terms. Links between terms were weighted by association strength and the terms were grouped with Louvain clustering. Node size shows the number of records mentioning the term. Digital forensics, one of the review's two core concepts, appears in only two of the 50 records, which is why it is not on the map.



Figure 6. Term co-occurrence map of the 50 records exported by Consensus as its top-ranked set (node size is the number of records mentioning a term; colour shows cluster; links are drawn for the strongest associations only).

The map has five clusters. The largest (blue) joins forensic accounting with emerging economies, prevention, skills and training, investigation and evidence, big data and blockchain, which is the forensic-practice stream.

A second (yellow) links financial fraud, explainability, trust, auditing, human oversight and card and transaction fraud. A third (green) holds SMEs, banking, risk management and technology adoption, with real-time monitoring, and it sits at the edge of the map. A fourth (orange) groups regulatory compliance, ethics and bias, deep learning, data privacy and cost barriers, and a fifth (pink) groups machine learning, anomaly detection, internal control and cybersecurity. The SME and adoption terms are peripheral and thinly linked, which is consistent with the finding from the 24 studies that small-firm evidence is the weakest part of the literature.

Figure 7 shades the same map by the mean publication year of the records that mention each term. The span is narrow, from 2023.2 to 2024.4, because the set is recent. Within it, technology adoption (2024.3), regulatory compliance (2024.1), skills and training (2024.1) and cybersecurity (2024.4, five records) are the newest terms, and SMEs (2023.3), risk management (2023.4), cost barriers (2023.4) and human oversight (2023.2, six records) are the oldest. Terms with few records move easily, so the overlay should be read with care.

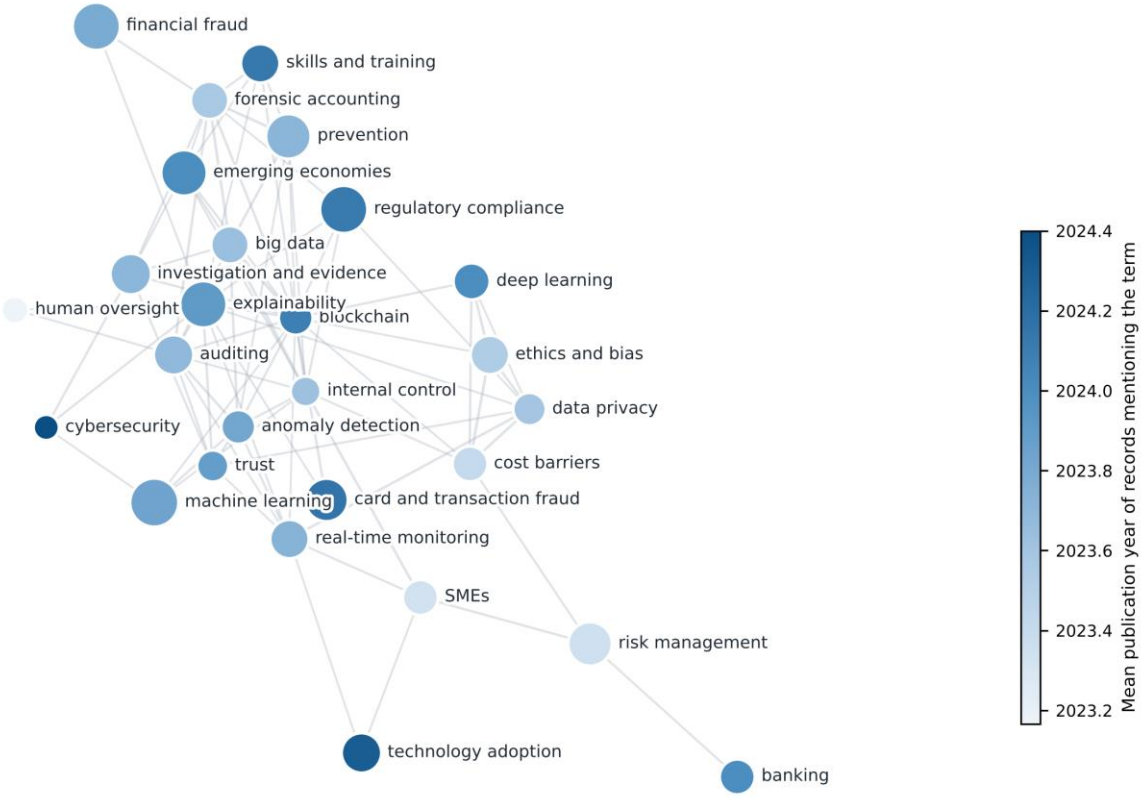

Figure 7. Overlay map: the same terms shaded by the mean publication year of the records that mention them (darker is more recent).

7.3 Scoping matrices (descriptive context, not used as evidence)

Two matrices were produced at the scoping stage, before the 24 studies were selected, and Figure 8 shows them as heat maps. Table 6 comes from the doctoral research outline, where it summarised the preliminary reading by topic and study method; the outline does not state how many papers it covers. Table 7 was generated by Consensus for its 50 exported records. Neither was re-coded by us. A paper is counted in every cell that applies, so the cells overlap and the totals exceed the number of papers. The row labels "in SMEs" in Table 6 overstate small-firm evidence, since only three of our 24 included studies were set in SMEs. The

matrices are reproduced because they shaped the early framing of the doctoral study. They are descriptive context only, and the gaps in Section 7.1 do not depend on them.

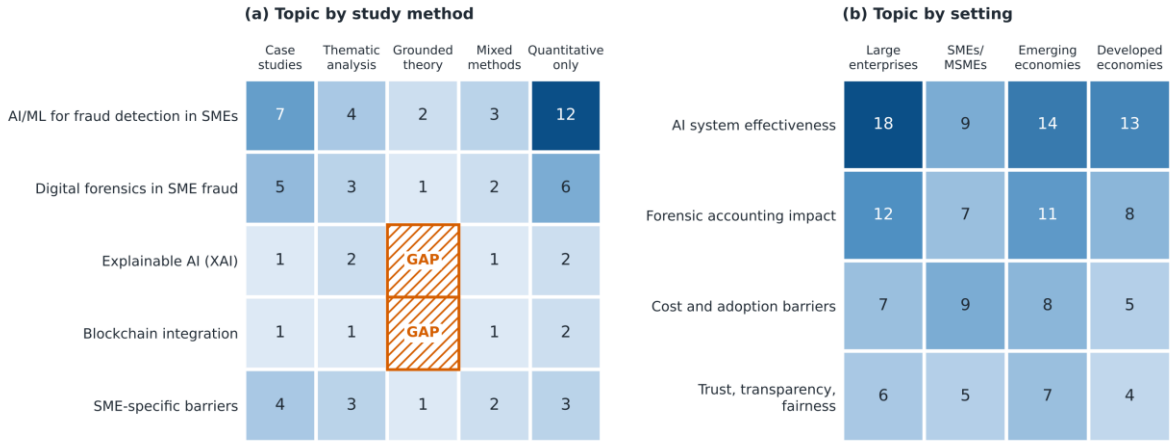

Figure 8. The two scoping matrices as heat maps: (a) topic by study method and (b) topic by setting (darker cells hold more papers; hatched cells are empty).

Table 6. Scoping matrix: topic by study method (papers per cell)

Topic	Case studies	Thematic analysis	Grounded theory	Mixed methods	Quantitative only	Row total
AI/ML for fraud detection in SMEs	7	4	2	3	12	28
Digital forensics in SME fraud	5	3	1	2	6	17
Explainable AI (XAI)	1	2	0	1	2	6
Blockchain integration	1	1	0	1	2	5
SME-specific barriers and challenges	4	3	1	2	3	13
Column total	**18**	**13**	**4**	**9**	**25**	**69**

Notes. Cell counts are taken from the doctoral research outline. Row and column totals are ours.

Table 7. Scoping matrix: topic by setting (papers per cell)

Topic or outcome	Large enterprises	SMEs/MSMEs	Emerging economies	Developed economies
AI system effectiveness	18	9	14	13
Forensic accounting impact	12	7	11	8
Cost and adoption barriers	7	9	8	5
Trust, transparency and fairness factors	6	5	7	4

Notes. Counts are taken from the Consensus report. Settings overlap, since a study of a large firm in an emerging economy appears in two columns.

8. Limitations

Discovery relied on two AI-assisted tools, citation chaining and Google Scholar. Direct searches of Scopus and Web of Science were not run. The tools rank with proprietary methods, so the retrieved sets cannot be regenerated exactly, the wording of the 21 Consensus searches is not available, and relevant studies may have been missed.

Not every study was read in full. Eligibility was assessed on the abstract where the full text could not be opened, and when the 24 included studies were re-read, nine could be read only at abstract or record level. Details of design and effect size are missing for those nine, and statements that a variable is untested are limited to the 15 read in full.

Screening and scoring were done by one reviewer. The second author advised on uncertain cases and checked the final set, but did not code independently, so there is no inter-rater statistic. The rubric measures fit with the doctoral study's population and setting, not risk of bias, and its weights were set by the authors. The scores given to the three adoption-context studies on the fraud criteria were too generous, as Section 4.3 shows. A different weighting would change the ranking, though not the identity of the gaps.

Journal quality was available for only eight of the 24 studies, and several included papers are internally inconsistent. We kept them and say where, but a stricter review would exclude some, and the positive findings would then rest on fewer studies still.

The search window closed at the end of 2025 although the searches were run in September 2026, so studies published in 2026 are not included. The review protocol was not registered. Finally, the heterogeneity of designs ruled out meta-analysis, and the thematic allocation of each study reflects our judgement of its main contribution.

9. Conclusion

The literature on AI and forensic accounting in fraud prevention is larger than the evidence about whether either helps small firms. Classifiers do well on some data sets and modestly on others, and practitioners think forensic accounting is useful. No included study shows that an MSME lost less to fraud or recovered more after adopting either, and the one fully reported test of an AI-based forensic system, in banks, found no effect on perceived detection or prevention.

For adoption, knowledge and skills are the most consistent factor. Cost, which is usually assumed to be decisive for small firms, was not significant in the only model that included it, and it has never been tested for fraud tools. Prior fraud experience has not been tested in anything we could read in full.

For researchers, the most direct contribution would be a study that measures outcomes in Indian MSMEs and tests cost and prior experience alongside skills. For practitioners and policymakers, the evidence so far supports investment in training before investment in platforms. It does not yet support claims about loss reduction.

Declarations

Author contributions. Surbhi Akash Jain: conceptualisation, methodology, literature search, screening, data curation, formal analysis, visualisation, writing of the original draft. Dini Menon: supervision, validation, review and editing. Both authors approved the final manuscript.

Use of AI tools. Two AI-assisted discovery tools (Consensus and SciSpace) were used to identify literature, as described in Section 3.2. The authors checked the output against the sources, take full responsibility for the content and verified every citation.

Protocol registration. The review protocol was not registered.

Funding. The authors received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Conflicts of interest. The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Ethics approval. Not applicable. The review used published literature only.

Data availability. The study-level scores and extraction used in this review are reproduced in Table 3 and Appendix A. The screening log and the discovery-tool outputs are available from the corresponding author on reasonable request.

References

ACFE. (2024). *Occupational fraud 2024: A report to the nations*. Association of Certified Fraud Examiners.

Akinbowale, O. E., Mashigo, P., & Zerihun, M. F. (2023). The integration of forensic accounting and big data technology frameworks for internal fraud mitigation in the banking industry. *Cogent Business & Management, 10*(1), Article 2163560. https://doi.org/10.1080/23311975.2022.2163560

Ali, A. M., Futahi, R. F., Shukur, M., & Al-Orfali, A. K. (2024). Forensic accounting and fraud detection: Emerging trends and techniques. *Journal of Ecohumanism, 3*(5), 525–542. https://doi.org/10.62754/joe.v3i5.3921

Badghish, S., & Soomro, Y. (2024). Artificial intelligence adoption by SMEs to achieve sustainable business performance: Application of Technology–Organization–Environment framework. *Sustainability, 16*(5), Article 1864. https://doi.org/10.3390/su16051864

Bhardwaj, A. (2025). Systematic literature review on explainable AI in finance: Methods, applications, and research gaps. *International Journal for Multidisciplinary Research, 7*(5). https://doi.org/10.36948/ijfmr.2025.v07i05.58963

Bin-Nashwan, S. A., Li, J. Z., Jiang, H., Bajary, A. R., & Ma'aji, M. M. (2024). Does AI adoption redefine financial reporting accuracy, auditing efficiency, and information asymmetry? An integrated model of TOE-TAM-RDT and big data governance. *Computers in Human Behavior Reports*, Article 100572. https://doi.org/10.1016/j.chbr.2024.100572

Cressey, D. R. (1953). *Other people's money: A study in the social psychology of embezzlement*. Free Press.

Dako, O. F., Onalaja, T. A., Nwachukwu, P. S., Bankole, F. A., & Lateefat, T. (2020). Forensic accounting frameworks addressing fraud prevention in emerging markets through advanced investigative auditing techniques. *Journal of Frontiers in Multidisciplinary Research, 1*(2), 46–63. https://doi.org/10.54660/.jfmr.2020.1.2.46-63

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340.

Devisa, F., & Purwanto, A. (2025). The development of technology in supporting the implementation of forensic accounting and investigative auditing in fraud detection (A study of the United States, the United Kingdom, China, Indonesia, and Nigeria). *Dinasti International Journal of Economics, Finance & Accounting, 6*(5), 4247–4257. https://doi.org/10.38035/dijefa.v6i5.5252

Elemide, A. O., & Araoye, F. E. (2025). Impact of artificial based forensic accounting on cyber fraud in deposit money banks in Nigeria. *International Journal of Business and Management Review, 13*(8), 139–159. https://eajournals.org/ijbmr/vol13-issue-8-2025/impact-of-artificial-based-forensic-accounting-on-cyber-fraud-in-deposit-money-banks-in-nigeria/

Goyal, K., Garg, M., & Malik, S. (2025). Adoption of artificial intelligence-based credit risk assessment and fraud detection in the banking services: A hybrid approach (SEM-ANN). *Future Business Journal*, *11*, Article 44. https://doi.org/10.1186/s43093-025-00464-3

Hafez, I. Y., Hafez, A. Y., Saleh, A., Abd El-Mageed, A. A., & Abohany, A. A. (2025). A systematic review of AI-enhanced techniques in credit card fraud detection. *Journal of Big Data, 12*(1). https://doi.org/10.1186/s40537-024-01048-8

Hamal, S., & Senvar, Ö. (2021). Comparing performances and effectiveness of machine learning classifiers in detecting financial accounting fraud for Turkish SMEs. *International Journal of Computational Intelligence Systems, 14*, 769–782. https://doi.org/10.2991/ijcis.d.210203.007

Kalluri, K. (2024). AI-driven risk assessment model for financial fraud detection: A data science perspective. *International Journal of Scientific Research and Management, 12*(12), 1764–1774. https://doi.org/10.18535/ijsrm/v12i12.ec01

Lee, C.-W., Fu, M.-W., Wang, C.-C., & Azis, M. I. (2025). Evaluating machine learning algorithms for financial fraud detection: Insights from Indonesia. *Mathematics, 13*(4), Article 600. https://doi.org/10.3390/math13040600

Mehta, K., Mittal, P., Gupta, P., & Tandon, J. K. (2022). Analyzing the impact of forensic accounting in the detection of financial fraud: The mediating role of artificial intelligence. In *Proceedings of the International Conference on Innovative Computing and Communications* (Advances in Intelligent Systems and Computing, Vol. 1388, pp. 585–592). Springer. https://doi.org/10.1007/978-981-16-2597-8_50

Mutemi, A., & Bação, F. (2024). E-commerce fraud detection based on machine learning techniques: Systematic literature review. *Big Data Mining and Analytics, 7*, 419–444. https://doi.org/10.26599/bdma.2023.9020023

Narula, A., & Bhagat, N. (2024). Artificial intelligence challenges and its impact on detection and prevention of financial statement fraud. In *Advances in human resources management and organizational development book series* (pp. 60–80). IGI Global. https://doi.org/10.4018/979-8-3693-0724-3.ch004

Naz, I., & Khan, S. N. (2025). Impact of forensic accounting on fraud detection and prevention: A case of firms in Pakistan. *Journal of Financial Crime, 32*(1), 192–206. https://doi.org/10.1108/jfc-01-2024-0010

Odeyemi, O., Ibeh, C. V., Mhlongo, N. Z., Asuzu, O. F., Awonuga, K. F., & Olatoye, F. O. (2024). Forensic accounting and fraud detection: A review of techniques in the digital age. *Finance & Accounting Research Journal, 6*(2). https://doi.org/10.51594/farj.v6i2.788

Odonkor, T. N., Adewale, T. T., & Olorunyomi, T. D. (2021). AI-powered financial forensic systems: A conceptual framework for fraud detection and prevention. *Magna Scientia Advanced Research and Reviews, 2*(2), 119–136. https://doi.org/10.30574/msarr.2021.2.2.0055

Okonta, C. N., & Nnamdi, C. A. (2025). Forensic investigation and fraud detection in Nigeria: Leveraging on artificial intelligence. *Journal of International Accounting, Taxation and Information Systems, 2*(1), 19–27. https://doi.org/10.70865/jiatis.v2i1.80

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, Article n71. https://doi.org/10.1136/bmj.n71

Polisetty, A., Chakraborty, D., G, S., Kar, A. K., & Pahari, S. (2023). What determines AI adoption in companies? Mixed-method evidence. *Journal of Computer Information Systems, 64*, 370–387. https://doi.org/10.1080/08874417.2023.2219668

Saha, P., Dey, K. N., Chowdhury, M. S. I., Das, R. C., Khan, M. M. H., Tanvir, M. T., & Halimuzzaman, M. (2025). Applications of forensic accounting in detecting and preventing financial crimes. *International Journal of Finance Research*, *6*(2), 71–86. https://doi.org/10.47747/ijfr.v6i2.2692

Sarna, N. J., Rithen, F. A., Jui, U. S., Belal, S., Amin, A., Oishee, T. K., & Islam, A. M. (2025). AI-driven fraud detection models in financial networks: A review. *IEEE Access, 13*, 141204–141233. https://doi.org/10.1109/access.2025.3596060

Sathe, A. K. (2025). Forensic accounting in the age of AI: Detecting fraud with data anomalies. *International Journal of Social Impact*, *10*(3). https://doi.org/10.25215/2455/1003053

Simbolon, R., Adriana, N., Rustam, A., Sulistyowati, N. W., & Rewa, K. A. (2024). The impact of forensic accounting on financial fraud prevention: A comparative analysis across countries. *The Journal of Academic Science*, *1*(8), 1074–1084. https://doi.org/10.59613/m6jrt421

Syahfir, H. A., Nirwana, N., & Haliah, H. (2024). Building public trust: The role of AI in preventing and exposing fraudulent financial reporting in the public sector – A systematic literature review. *West Science Accounting and Finance*, *2*(3), 519–535. https://doi.org/10.58812/wsaf.v2i03.1322

Tornatzky, L. G., & Fleischer, M. (1990). *The processes of technological innovation*. Lexington Books.

Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management, 14*(3), 207–222.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, *27*(3), 425–478.

Yaseen, H., & Al-Amarneh, A. (2025). Adoption of artificial intelligence-driven fraud detection in banking: The role of trust, transparency, and fairness perception in financial institutions in the United Arab Emirates and Qatar. *Journal of Risk and Financial Management, 18*(4), Article 217. https://doi.org/10.3390/jrfm18040217

Appendix A. Relevance rubric scores by criterion

Criteria and weights are given in Table 2. Studies are ordered by total score.

Study	C1	C2	C3	C4	C5	C6	C7	Total
Elemide & Araoye (2025)	18	19	8	15	6	9	10	85
Badghish & Soomro (2024)	15	10	15	13	10	9	10	82
Saha et al. (2025)	10	19	12	14	8	8	10	81
Mehta et al. (2022)	16	19	8	12	6	10	8	79
Okonta & Nnamdi (2025)	15	18	8	14	6	7	10	78
Akinbowale et al. (2023)	14	19	7	13	5	10	9	77
Hamal & Senvar (2021)	19	9	15	12	3	10	8	76
Naz & Khan (2025)	10	19	10	14	5	9	9	76
Yaseen & Al-Amarneh (2025)	16	10	7	12	10	9	10	74
Devisa & Purwanto (2025)	12	18	8	13	6	7	10	74
Dako et al. (2020)	9	19	10	15	5	8	7	73
Polisetty et al. (2023)	15	8	10	12	10	10	8	73
Bin-Nashwan et al. (2024)	15	9	8	11	10	10	9	72
Goyal et al. (2025)	15	8	6	12	10	10	10	71
Simbolon et al. (2024)	10	18	7	13	5	8	9	70
Lee et al. (2025)	18	8	8	12	3	10	10	69
Ali et al. (2024)	13	18	6	10	5	6	9	67
Narula & Bhagat (2024)	16	13	7	9	6	6	9	66
Syahfir et al. (2024)	14	12	7	11	6	7	9	66
Odonkor et al. (2021)	16	17	6	8	5	6	7	65
Kalluri (2024)	19	8	6	9	4	9	10	65
Odeyemi et al. (2024)	12	18	6	9	4	6	9	64
Bhardwaj (2025)	13	9	6	8	6	8	10	60
Sathe (2025)	13	16	6	8	3	5	6	57

Appendix B. Illustrative exclusions at eligibility assessment

Sixteen of the 177 exclusions are listed with the score assigned and the reason for exclusion. Studies are identified by author and year only.

ID	Study	Reason for exclusion	Score	Category
E-01	Rahmati (2025)	Purely technical GNN/FL architecture; no forensic-accounting, adoption or MSME dimension	34	Technical / off-scope
E-02	Vallarino (2025)	Algorithm-centric GNN study; no organizational or forensic-accounting lens	33	Technical / off-scope
E-03	Olaniyan et al. (2025)	Deep-learning model engineering; outside a Commerce-faculty adoption scope	30	Technical / off-scope
E-04	Gurrala et al. (2025)	Technical detection pipeline; no adoption/behavioural or SME context	31	Technical / off-scope
E-05	Masunda & Barot (2025)	AML-specific and technical; not MSME fraud or forensic-accounting practice	36	AML-only / off-topic
E-06	Thakur et al. (2025)	Privacy-engineering focus; no MSME adoption or forensic-accounting evidence	35	Technical / off-scope
E-07	Nicholls et al. (2023)	Technical AML/LLM pipeline; not aligned with forensic accounting in MSMEs	32	AML-only / off-topic
E-08	Aljunaid et al. (2025)	Large-institution banking + technical; no SME/MSME transferability	40	Large-firm only
E-09	Vanini et al. (2023)	Large-scale payment-fraud engineering; no forensic-accounting or MSME focus	39	Large-firm only
E-10	Mutemi & Bação (2024)	E-commerce technical ML review; not MSME/forensic-accounting centred	37	Not centred on target fraud
E-11	Ashta & Herrmann (2021)	Broad fintech overview; not specific to fraud/forensic accounting	38	Not centred on target fraud
E-12	Islam et al. (2024)	Conceptual future-directions piece; no empirical or methodological contribution	41	Non-empirical opinion
E-13	Chotrani (2024)	Conceptual and redundant with stronger empirical selections	44	Non-empirical / redundant
E-14	Dowuona (2024)	Conceptual; non-indexed venue; content covered by included studies	42	Redundant / venue
E-15	Weber et al. (2018)	Pre-2020 anchor and technical AML; superseded by later included work	28	Developed/older, off-scope
E-16	Elumilade et al. (2021)	Overlaps/duplicates digital-age forensic reviews already included	46	Redundant / duplicate

Appendix C. Search strings, search statements and strategies

Master Boolean string (doctoral synopsis). ("Artificial Intelligence" OR "AI" OR "Machine Learning") AND ("Digital Forensics" OR "Forensic Accounting" OR "Cyber Forensics") AND ("Financial Fraud" OR "Fraud Detection") AND ("Small Businesses" OR "SMEs" OR "MSMEs") AND ("Qualitative Study" OR "Case Study" OR "Thematic Analysis" OR "Grounded Theory" OR "Interpretive Analysis").

SciSpace: the five search statements, as worded by the tool.

- Integration of Artificial Intelligence (AI) and Digital Forensics/Forensic Accounting in financial fraud detection and prevention

- Real-time AI-driven forensic analytics for financial fraud detection and prevention: integrating forensic accounting, digital forensics, blockchain analytics, and explainable AI in banking and corporate finance
- Exploring AI-assisted digital forensics, forensic accounting, and auditing methods for financial fraud prevention: integrating graph neural networks, federated learning, and explainable AI across banking, fintech, and corporate finance transactions
- Fusion of Federated Graph Learning, Explainable AI, and Digital Forensics for Cross-Institutional Financial Fraud and AML Detection: graph-based anomaly detection, privacy-preserving collaboration, and explainable models across banking, fintech, and corporate finance networks
- Graph-based, privacy-preserving and explainable AML/financial fraud detection using cross-institution collaboration: federated and vertical FL, graph neural networks on temporal transaction networks, and knowledge-graph–driven explainability for cross-border financial crime

SciSpace filters. Publication in 2020 to 2025, with foundational studies from 2015 to 2019; peer-reviewed journal articles. Yield: 664 records from the statements and 69 from backward and forward citation chaining.

Consensus: the six strategies (21 searches and one citation-graph query; the wording of individual searches is not reported by the tool).

- Foundational frameworks: TAM and TOE
- SME and MSME contexts, with emphasis on India
- Terminology variants, for example "machine learning" and "digital forensics"
- Adoption barriers: cost, awareness and prior experience
- Adjacent constructs: risk management and regulatory compliance
- Empirical evidence on effectiveness

Consensus yield. 730 records identified; 240 screened by the tool; 164 eligible; 50 top-ranked records exported (the set used in Section 7.2).