

Evidence Preserving Forensic Accounting Advisory A Cost Sensitive Model for Financial Misstatement Triage

Kehinde Oyediji¹, Uchechi Ngonadi², Esther Nwozaku³

¹ KPMG, United Kingdom

² J.L. Richards & Associates

³ Central Michigan University, Michigan, United States

Corresponding author: Kehinde Oyediji; kennyoyediji@gmail.com

ABSTRACT

Forensic accounting screens can generate substantial review burdens when the target condition is uncommon, even when assumed detection properties appear favourable. This paper develops an evidence-preserving and cost-sensitive advisory model linking alerts, corroboration, alternative explanations, case disposition, and review capacity. A purposive literature synthesis supports conceptual development, and deterministic sensitivity analysis evaluates explicitly hypothetical screening and economic assumptions. With 10,000 items, one percent prevalence, 80 percent sensitivity, and a five percent false-positive rate, the example produces 80 true alerts and 495 false alerts, giving 13.91 percent precision despite 94.85 percent overall accuracy. An illustrative expected-value calculation yields a 4.76 percent review threshold under its stated loss, recovery, cost, and harm assumptions. Changes in recovery, timing, or supported loss alter that threshold. These results concern an assumed target condition and do not establish fraudulent intent. The proposed method requires population reconciliation, preserved source records, reproducible transformations, separate loss schedules, and documented treatment of unresolved cases. Its contribution is a transparent link between analytical screening and authorised evidence review. Operational detection rates, investigative returns, and causal performance effects remain questions for empirical validation.

DOI: 10.21590/ijtmh20230904449

INTRODUCTION

The alert and the allegation

Forensic accounting advisers face a practical allocation problem. A client presents thousands of transactions, incomplete explanations, and limited funds for investigation. Screening every unusual entry as suspected fraud wastes resources and risks unsupported accusations. Ignoring anomalies also leaves exposure unresolved. This paper develops a

model for prioritising financial misstatement concerns while preserving the distinction between an analytical alert, a corroborated irregularity, and a conclusion supported by evidence.

Financial misstatement, accounting error, and intentional deception are different categories. A misstatement concerns an incorrect financial representation. Fraud requires consideration of intent and the circumstances of the act. An anomaly describes departure from a statistical or business expectation. An unusual transaction might reflect a new product, a system migration, a reversal, or a legitimate exception. The proposed advisory model therefore treats analytical results as prompts for investigation and gives documentary corroboration a separate role.

The literature establishes several screening approaches. Beneish develops an earnings manipulation model from financial information [1]. Dechow and colleagues develop a model for material accounting misstatements [2]. These contributions support examining relationships among reported balances and performance indicators. Their predictors and classifications do not replace inspection of records or explanations. A ratio-based alert at enterprise level also differs from an allegation about the employee who entered a specific transaction.

Comparative research further complicates model selection. Perols compares statistical and machine learning approaches under different misclassification costs and fraud-to-nonfraud ratios [3]. Ngai and colleagues organise data mining applications in financial fraud detection [4]. West and Bhattacharya review intelligent financial fraud detection [5]. These studies motivate evaluating the decision setting, rather than choosing an algorithm solely because a publication reports high accuracy. The present paper translates this concern into an advisory queue with explicit investigation costs and outcome categories.

Rare events require particular care. A low false-positive rate still produces many false alerts when the screened population contains few relevant cases. Saito and Rehmsmeier examine precision-recall evaluation for imbalanced classification [6]. Bolton and Hand review statistical fraud detection [7]. The resulting lesson for advisory design is specific: management needs the expected composition of the alert queue, the share of relevant cases missed, and the cost of follow-up. A single overall accuracy percentage does not supply these quantities.

Anomaly detection adds a second boundary. Chandola and colleagues organise anomaly detection techniques around underlying assumptions [8]. An unsupervised score measures unusualness relative to a representation of normal behaviour. Such scores require attention to business context and data preparation. A change in the accounting system's coding practices might create many anomalies without a change in underlying misconduct. The adviser should therefore record how the population was assembled and which business processes the score represents.

Country and population differences also affect interpretation. Marais and colleagues compare Beneish and Dechow models in South Africa [9]. This comparison motivates scrutiny of model transfer across settings. The present conceptual study does not adopt a published threshold as a proven probability of fraud in SMEs. Instead, the study asks how an adviser should order follow-up after screening, when prevalence, costs, evidence completeness, and reviewer capacity constrain the engagement.

The research question is: how should forensic advisers translate screening alerts into a defensible investigation queue without confusing analytical suspicion with proof? The contribution has three parts. First, the model separates screening, corroboration, and findings. Second, a transparent numerical exercise shows the effect of prevalence and review capacity. Third, an evidence-preservation process links each finding to source records and a documented explanation. The numerical results illustrate assumed screening properties and do not measure detection performance in a real client population.

Disclosure quality and investigative judgement

The advisory setting creates incentive conflicts before analysis begins. A manager might seek confirmation of an allegation, a supplier might challenge the meaning of a payment record, and employees might fear an adverse interpretation of an unexplained exception. Agency theory supplies a framework for analysing monitoring, information, and incentives [10] [11]. The model responds by specifying the question and outcome categories before reviewing cases. A reviewer should investigate the defined condition rather than search selectively for facts supporting a preferred narrative.

Corporate disclosure and earnings quality research further caution against treating every accounting signal as equivalent. Disclosure addresses the information available to outsiders, while accrual quality concerns how accounting estimates relate to underlying cash flows [12]

[13] [14]. A receivable increase, an estimate revision, and an unsupported journal entry therefore need different follow-up. The proposed model records the mechanism implied by each alert and the evidence required to investigate that mechanism. This avoids grouping unlike risks under a single fraud label.

Financial and nonfinancial evidence also play complementary roles. Research using nonfinancial measures to assess fraud risk and research on trust after fraud supply context for checking reporting consistency and governance [15] [16]. The paper does not assume a particular discrepancy establishes deception. A sales increase without corresponding operational activity might warrant questions about billing and revenue recognition. The interpretation still depends on delivery records, contract terms, and explanations which the reviewer has examined.

Machine learning requires the same discipline. Random forests and research on accounting fraud detection demonstrate different approaches to classification [17] [18]. Algorithmic flexibility does not remove the need for an appropriate target condition, representative population, and realistic evaluation. A model trained on publicly traded firms also differs from a screen used for individual payroll transactions. The proposed model treats population transfer as an explicit limitation and keeps the original decision context visible in the case register.

Interpretability research supplies another design consideration. Rudin argues for interpretable models in high-stakes decisions, while Guidotti and colleagues review methods explaining black-box models [19] [20]. Explanations help a reviewer identify why an alert arose. They do not demonstrate that the indicated feature caused an irregularity. The advisory model uses explanations to organise questions and evidence requests, without presenting a feature contribution as a factual finding about intent or responsibility.

Data quality and information overload affect the review process as much as algorithm choice. Research on data quality assessment, business consequences of poor records, and information overload motivates inspection of the inputs and the form of the queue [21] [22] [23]. A large unfiltered export might conceal missing records, duplicates, and inconsistent identifiers. The adviser therefore checks population completeness before ranking alerts and summarises linked entries as cases when the underlying issue is shared.

Governance research adds another reason to keep screening and findings separate. Studies of board composition, audit committee activity, and corporate fraud examine oversight arrangements [24] [25]. Earnings management research considers accounting behaviour under a specific reporting incentive [26]. These sources motivate investigation of incentives and oversight but do not establish intent in a new case. The proposed register keeps the observed transaction, surrounding controls, and supported interpretation distinct so a general governance concern does not become an unsupported individual allegation.

Analytics and AI capability studies also describe resources beyond an algorithm [27] [28] [29] [30]. A forensic screen needs appropriate records, reviewers who understand the process, and an accountable response. The model therefore treats analytics as one part of an engagement. It measures relevant findings and review effort rather than equating successful software use with successful investigation. Organisational knowledge and absorptive capacity provide related reasons to test reviewer understanding and client follow-through [31] [32].

Obukadata and colleagues distinguish delivered changes from their use, financial consequences, stakeholder outcomes and persistence. Their narrative review also distinguishes measurement quality from attribution. The evaluation design supplies constructs for organising follow-up, without estimating an investigative effect. Corrective action requires follow-up evidence of use and persistence after the investigation closes. [33]

Okoli and colleagues place diagnosis and method choice within a governance and learning process for large operations. Their framework routes different operational problems to Lean, Six Sigma, or analytical approaches. For case review, this argument supports choosing an analytical technique after defining the suspected discrepancy; the source does not quantify investigative returns. [34]

Okoli and colleagues develop controllable-variable isolation for first-pass-yield analysis in distributed operations. Their analysis separates operating conditions from controllable variables and identifies the management level with authority over each change. A performance exception therefore needs a supported causal account before a reviewer assigns responsibility to a person or function. It does not establish the effectiveness of the model proposed here. [35]

Mensah and colleagues review asset-liability management and interest-rate risk in banking. The relevant connection is the interaction between cash-flow timing, repricing

assumptions, and governance. Bank balance-sheet models are not directly interchangeable with SME forecasts or investigative scores. The financing study informs interpretation of the operating context; the investigation examples do not take calibrated parameters from it. [36]

Mensah and colleagues examine capital adequacy, risk-weighted assets, and bank portfolio choices. Their review places solvency and regulatory capital within the same decision problem. This differs from an operating cash floor or an evidence threshold. A reviewer must identify what a regulated capital ratio measures before interpreting a deviation as a case concern. [37]

Mensah and colleagues develop an enterprise-risk and capital-allocation model for financial institutions. A case recommendation should show how the supported exposure alters the allocation decision rather than merely listing risks. The present manuscript retains its own decision boundary and assumed numerical inputs. It does not infer financial returns or validated screening performance from this conceptual integration. [38]

The LNG procurement framework developed by Oyediji, Adesanya and Okoli examines suspected vendor schemes through connected purchasing episodes. The episode record joins supplier selection, delivered goods, invoices and settlement so a discrepancy receives a separate assessment before a scheme finding. This supplied conceptual paper motivates transaction reconstruction and a distinct loss schedule. It reports no empirical accuracy estimate for the advisory model developed here. [39]

Identity verification, competence, independence and unresolved concerns form distinct review dimensions in the board-nominee framework of Oyediji, Adesanya and Okoli. Their proposed uncertainty gate preserves material gaps instead of averaging them away. The transfer to advisory work concerns evidence interpretation and decision responsibility. A nominee review addresses different evidence from a payment investigation and does not determine the legal character of a transaction. [40]

The correspondent-banking analysis by Agboada and colleagues brings settlement economics and risk into the interpretation of cross-border financial records. The relevant link is between settlement conditions, operational controls, and commercial objectives. Network-level settlement research does not determine the finding in an individual investigation. The study supports interpretation of the case setting and does not supply validation or estimated case probabilities. [41]

Pedro and colleagues organise third-party fund-administration governance around valuation integrity, data lineage, delegation, resilience, and reporting. Their narrative framework distinguishes service quality from isolated efficiency or revenue measures. This provides a bounded example of why responsibility and source records matter when work is delegated. It is not evidence of a measured improvement in the present model. [42]

Published work on investment-fraud exposure, API governance, fairness auditing, cost of poor quality, and financial-data auditability adds several dimensions to case prioritisation. Exposure differs from likelihood, data access differs from evidence completeness, and an efficient queue differs from a fair review. The present model therefore avoids a single score concealing these distinctions. It reports exposure, analytical priority, evidence status, and review outcome separately. Fairness concerns require attention to which populations enter the screen and whether unrelated personal characteristics influence selection. The cited material motivates these questions but does not certify a compliant or unbiased implementation. A future evaluation needs authorised data and a defined assessment of the particular decision process, with independent examination of consequential findings. [43] [44] [45] [46] [47]

Risk-based audit, vendor governance, supplier relationships, public-system risk detection, and payment automation concern control arrangements around transactions. In the proposed model, an exception in a payment process triggers inspection of the approval and counterparty records. A recommendation to automate payments also needs a reliable rule for legitimate exceptions. Automation might reproduce an existing error more quickly when the supplier register or invoice identifier is unreliable. These sources motivate connecting the screen to the business process and the corrective action. They do not establish that an automated alert proves a duplicate payment or that a new workflow prevents all recurrence. The evidence register should therefore retain the original entry, control condition, explanation, and subsequently tested correction as different components. [48] [49] [50] [51] [52]

Supply-chain prediction, audit liaison, and cross-border control reviews provide context for reconciling operating explanations and financial evidence. The supplied solar and street-lighting studies add an application setting in which technical feasibility and financial reporting require different records. A reviewer investigating an equipment project would distinguish projected energy output, approved procurement cost, actual delivery, and recorded payment. Engineering plausibility does not verify the payment, while a correct

payment record does not verify technical performance. The model uses this separation to organise evidence requests and avoid conclusions spanning beyond the inspected material. The cited energy studies supply context rather than fraud labels, transaction outcomes, or parameters for the numerical screen. This boundary permits relevant multidisciplinary discussion without overstating what each source establishes. [53] [54] [55] [56] [57]

RESEARCH METHOD

Sources and analytical scope

Case provenance defines the design of this study. A screening result enters a documented review process, where the analyst must preserve the original record, test competing explanations and state the evidence supporting a disposition. The source set covers misstatement screening, anomaly detection, imbalanced classification and investigation governance, with recorded publication years through 2023. Jaakkola informs the conceptual specification [58]. Deterministic examples then examine how prevalence, loss exposure, recoverability and reviewer capacity affect the queue. Article metadata and available DOIs were checked against the supplied records, publisher information and DOI records. Unresolved bibliographic differences are documented in the reference audit. No client transaction dataset or measured investigation outcome enters the calculations.

Discovery used title and subject searches for financial misstatement prediction, earnings manipulation, fraud detection, anomaly detection, and precision-recall evaluation. The source set is purposive. The paper does not claim a systematic review, exhaustive database coverage, independent coding, or a meta-analysis. Published model accuracy values are not combined. No client dataset, interview sample, confirmed fraud register, or labelled transaction population was provided. All numerical inputs below are declared assumptions used to inspect the proposed decision logic.

Review-method literature supports a clear distinction between a purposive conceptual source set and a systematic review [61] [62] [63]. The present design uses sources to identify decision constructs. It does not estimate a pooled fraud-detection effect or reproduce an exhaustive review protocol. This boundary matters because the deterministic examples follow assumptions chosen for transparency; their outputs do not acquire empirical validity merely through the number of references supporting the general research problem.

Population construction and record provenance

The unit of analysis is a review case. A case might comprise one transaction or a related transaction group, depending on the allegation and process. Cases need stable identifiers to avoid counting several entries from one issue as independent discoveries. The model distinguishes the screen-positive status A, relevant confirmed condition F, loss exposure L, effective remediation fraction r , review cost K, and false-escalation cost H. Relevant conditions include the precisely defined misstatement or irregularity targeted by the engagement; this category does not automatically imply criminal fraud.

An engagement begins with a written question. Examples include whether duplicate vendor disbursements occurred, whether receivable balances include unsupported invoices, or whether an employee received unauthorised payroll additions. The question identifies the process and period, the target condition, the authorised access, and the decision the report will support. A general request to find fraud is too broad for a reproducible screen. Defining the question also reduces the risk of classifying transactions using criteria introduced only after a suspicious example appears.

The population check precedes all screening. The adviser records the extraction query, date range, source system, relevant account codes, and number of cases. Totals reconcile to the applicable ledger or reporting record. A missing period, excluded subsidiary, or suppressed reversal changes the meaning of the analysis. These differences are resolved or disclosed before screening. The adviser retains the original extract separately from transformed working data so later reviewers identify which values came from the source and which were created during preparation.

Transformation records identify normalisation, joining, aggregation, and exclusion rules. Vendor names might be standardised for matching, but the original names remain available. Multiple ledger lines belonging to one payment should be grouped when the review case concerns the payment. Reversals and cancellation flags require consistent handling. A transformation which removes all unusual amounts might also remove the records the engagement intends to investigate. Each exclusion therefore needs a reason linked to scope rather than a general preference for a cleaner dataset.

Access and preservation are distinct controls. Permission to read an accounting system does not automatically authorise collecting personal communications or copying unrelated employee records. The adviser obtains the access necessary for the defined question and

limits handling to the authorised scope. Preservation records who obtained the file, where the working copy was held, and whether subsequent changes occurred. A file digest helps detect differences between copies, while an access log records handling. Neither control alone establishes the truth of the file's original contents.

Data quality has several dimensions which affect case interpretation [64] [65]. Completeness concerns whether relevant records are present, consistency concerns compatible representations, and contextual usefulness concerns fit for the engagement question. A complete export from one module might still omit the bank evidence required to resolve a payment alert. The population check consequently identifies both included records and additional evidence needed. Consulting research also motivates distinguishing a completed analytical deliverable from its demonstrated organisational consequences [66] [67].

The proposed triage model requires a defined review population before any accuracy measure has meaning. A population can consist of invoices, payments, procurement episodes, accounts, or cases. These units cannot be exchanged without changing the denominator. A single procurement episode might generate several invoices and payments, while several alerts might point to the same event. The implementation record therefore identifies the unit of screening and the unit of investigation separately. The worked example uses ten thousand screened items and assumes each item can receive one binary alert. Any client application must state whether that assumption is appropriate and retain the links needed to consolidate related alerts into a reviewable case.

Population construction also requires a coverage reconciliation. The analyst records the source system, date range, inclusion rules, excluded records, and opening record count. Cancelled transactions, reversals, internal transfers, and test records require explicit treatment. Excluding them can be reasonable for a particular rule, but an unexplained exclusion can change the apparent prevalence or remove the very pattern under investigation. A reconciliation connects the extracted population with the relevant ledger or operational total. Where the two do not match, the difference remains visible as a data limitation. Screening should not proceed on the assumption that a large file is complete simply because it contains many rows or sophisticated analytical fields.

The evidence model distinguishes an original business record from an analytical derivative. An invoice image, source-system entry, or approved purchase order supplies

evidence about a particular stage of the transaction. A joined table, calculated feature, or alert score is a derivative created by the analyst. The case file retains the source reference and the transformations producing the derivative. A reviewer should be able to reconstruct a material discrepancy from preserved inputs. This requirement does not imply that the original record is truthful. It establishes what the analysis relied on and allows contradictory evidence to be evaluated without losing the relationship between the source and the analyst's interpretation.

Rule evaluation begins with a written statement of the question being tested. A repeated invoice amount can indicate a recurring legitimate charge, a data duplication, or an improper duplicate payment. The rule specifies which fields identify the same supplier and invoice, how date differences are treated, and whether reversals are netted. A relationship rule requires its own definitions for shared addresses, contact details, or bank accounts. The analyst should retain the reason a field is relevant to the hypothesised mechanism. This makes it possible to challenge the rule before case review begins. An unexplained score is insufficient when the decision may affect payment, employment, reputation, or access to a supplier relationship.

Corroboration and case disposition

The evidence process has five stages: preserve the authorised source population, reproduce the alert, inspect related records, evaluate explanations, and classify the finding. Required case fields include identifier, source system, extraction date, query or rule version, reviewer, documents examined, alternative explanations, unresolved gaps, and disposition. A source file digest assists later integrity checks, while access logs record handling. These steps are proposed engagement controls, rather than a claim of compliance with a particular jurisdiction's evidence rules.

The review procedure also records alternative explanations prospectively. For a duplicate-payment alert, expected alternatives include reversal, instalment, correction, and duplicate disbursement. The reviewer identifies documents supporting each possibility and notes whether the evidence resolves the case. A case remaining unresolved receives an unresolved disposition. Failure to locate a document is not equivalent to proof the transaction was unauthorised. The distinction helps prevent missing records from being converted automatically into adverse findings.

The model separates factual discovery from recommendation. The factual record states which entries and documents were observed. Interpretation explains the supported meaning and remaining uncertainty. Recommendation identifies a control change, correction, recovery step, or additional review. Combining these elements into one sentence risks concealing the reasoning. A report saying a payment requires follow-up should not imply that misconduct has been established when only an unresolved document gap supports the statement.

Case review records both supporting and contradictory observations. A structured review form identifies the alerted condition, the underlying transaction, the commercial explanation offered, the records examined, unresolved gaps, and the next authorised action. It separates a corrected posting error from an unexplained discrepancy and from a conclusion supported by multiple independent records. The reviewer records why an explanation was accepted or rejected. The purpose is to preserve the reasoning, rather than to force all cases into an early binary fraud label. A case can be closed as an error with a financial correction while another remains unresolved because evidence is incomplete. These outcomes contribute differently to evaluation.

Procurement-risk research from the reviewed literature links supplier, operational, financial, and contextual exposures with data integration and decision support. Its conceptual treatment supports viewing a payment alert within the procurement episode rather than as an isolated ledger row. A reviewer should connect supplier selection, contract terms, order approval, delivery evidence, invoice processing, and settlement. These links define the records needed to examine a particular discrepancy. They do not establish a scheme through association alone. Shared supplier addresses, repeated amounts, or unusual delivery timing still require alternative explanations and corroboration. The present model uses the procurement perspective to define the investigation boundary and preserve linked records, while leaving the conclusion dependent on the evidence examined in the case. [69]

A published framework for user acceptance testing in small businesses emphasises requirement traceability, realistic scenarios, feedback, and decision checkpoints. These design features support checking whether a proposed forensic workflow operates as specified before staff rely on the workflow. The acceptance exercise should use authorised test records and known expected outcomes, including legitimate exceptions and unresolved cases. A reviewer verifies whether the workflow preserves source identifiers, reproduces the alert, retains the explanation, and assigns the correct disposition. Completing this exercise

demonstrates conformance with the defined test conditions. It does not estimate fraud-detection performance in the client population. The acceptance record therefore states the conditions tested and the requirements still unresolved before the engagement moves into operational review. [70]

The published review of IT service management adds a practical basis for defining ownership and continuity around the analytical service. Its discussion connects technology service governance with staff capability, process adoption, measurement, and organisational constraints. For forensic work, these issues arise when extract routines change, a screening rule fails, or an alert queue grows beyond reviewer capacity. The model should assign responsibility for obtaining the source extract, correcting an analytical defect, approving a rule change, and handling cases already generated by the defective rule. These responsibilities belong to distinct decisions, even where one employee performs several roles. A resolved software incident does not itself resolve an investigated accounting case. The case retains its evidence and disposition separately from the service incident. [71]

The investigation plan should distinguish the authorised scope from the possible scope suggested by an alert. A payment discrepancy might reveal a concern involving another supplier or period, but the reviewer should record the proposed extension and its justification before collecting additional records. This requirement prevents a broad analytical query from silently becoming a broader investigation. The register identifies the transaction population, source systems, relevant period, target condition, and access required for the current question. Where a new question arises, management records the decision to extend or refer the matter. The analytical result serves as the reason for considering an extension. The extension's authorisation and evidence requirements determine what work follows. This makes the investigation reproducible across reviewers and preserves the boundaries supporting the eventual report.

Evidence collection should preserve enough context to reconstruct the business event. A payment screenshot alone might omit reversal history, approval status, currency, or the identity of the relevant entity. The reviewer should therefore retain the source reference and the contextual fields needed to interpret the record. For a purchase, the context might include order number, supplier identifier, delivery quantity, contractual unit price, tax treatment, and settlement reference. The process need not copy every field available in the system. It should collect the fields necessary for the defined question and explain why they matter. A later

reviewer should be able to distinguish an observed difference from a difference produced by currency conversion, reporting filters, or inconsistent units. The analytical derivative retains the transformation steps supporting this comparison.

Reviewer disagreement should receive a recorded resolution rather than disappear in an average score. One reviewer might classify a transaction as an unresolved delivery discrepancy while another accepts the receiving document as sufficient evidence. The disagreement concerns the meaning or reliability of a record, rather than merely a numerical ranking. The case file should state the competing interpretations, the supporting documents, and the additional work required. A designated reviewer then records the basis for the disposition or preserves the uncertainty when the available evidence does not resolve the issue. This procedure is part of the proposed model and has not been evaluated here. Its intended contribution is to make consequential judgement inspectable and prevent a majority preference from concealing an evidential gap.

The method should also separate case closure from recommendation completion. A verified duplicate payment might receive a clear factual disposition while recovery remains pending. A legitimate transaction might close the suspicion while exposing a control weakness needing correction. An unresolved case might require a documented referral rather than remain indefinitely in the same queue. The register therefore retains a finding status and a follow-up status. These statuses have different owners and completion criteria. A case should not receive a stronger factual conclusion merely because management wants to finish a corrective action, and a completed investigation should not count as realised recovery without a corresponding receipt or verified settlement. Keeping these records separate supports both reporting accuracy and later evaluation of the engagement's operational consequences.

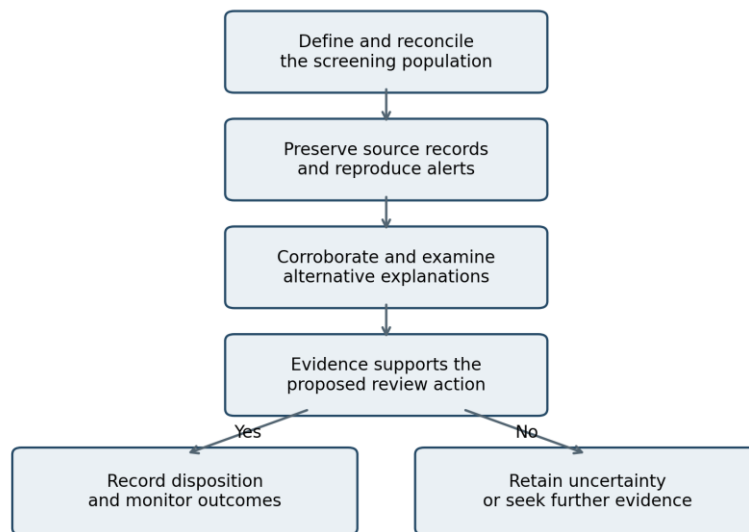


Figure 1. Proposed evidence-preserving review process. An alert does not establish misconduct.

Screening performance and calibration

For an illustrative population N , expected true positives equal $N \times p \times s$, where p is assumed prevalence and s is sensitivity. Expected false positives equal $N \times (1-p) \times f$, where f is the false-positive rate. Precision equals $ps/[ps + (1-p)f]$. Expected false negatives equal $N \times p \times (1-s)$. These equations are accounting consequences of the assumptions, not estimated model coefficients. Fractional counts in the examples represent expectations, rather than fractional observed cases.

The primary example assumes 10,000 cases, prevalence of one percent, sensitivity of 80 percent, and a false-positive rate of five percent. Prevalence is also varied to 0.5 percent and five percent, holding screening properties fixed. A second exercise compares two assumed thresholds: a broader screen with sensitivity of 90 percent and a false-positive rate of ten percent, and a narrower screen with sensitivity of 60 percent and a false-positive rate of two percent. These comparisons isolate parameter effects without training or validating a classifier.

The screen design uses process knowledge alongside financial patterns. A duplicate-payment rule might combine amount, supplier identifier, invoice number, and payment timing. A payroll rule might compare authorised employee status with disbursement records. A revenue rule might identify invoices lacking corresponding delivery evidence. These screens have different denominators and plausible explanations. The model requires a rule-

specific evaluation rather than a single overall performance statistic across unlike target conditions.

If a supervised model is considered, development and testing need separation in time and case identity. Linked records from one issue must not appear on both sides of the split. Data available only after investigation must not enter a predictor intended for earlier screening. Research on classifier evaluation and concept drift motivates these controls [59] [60]. A future evaluation should report performance by period and process because a rule which fits an earlier accounting system might lose relevance after workflow changes.

Calibration needs an explicit check. A probability estimate of ten percent has a practical meaning only when comparable cases receive relevant outcomes at approximately the stated frequency. Ranking quality and calibration are separate properties. A score might order cases well while exaggerating probability, or provide reasonable average probabilities while distinguishing cases poorly. The present numerical rule requires calibrated q . Where this condition fails, the adviser uses qualitative priority categories and avoids a monetary calculation implying unsupported probability precision.

Sensitivity and specificity require a credible outcome classification. If only alerted cases are investigated, the analyst can estimate the observed disposition of those alerts but cannot directly measure missed cases in the unalerted population. The method therefore proposes an authorised review sample from both groups for later empirical testing. Sampling weights would be needed when the reviewed groups represent different proportions of the full population. The current calculations do not use such a sample and make no measured claim about screening performance. They show the consequences of assumed rates. This distinction prevents the term accuracy from concealing the difference between a mathematical example and an evaluated operational system.

The published review of credit-risk and default prediction offers a relevant caution for forensic screening: discrimination, calibration, and economic evaluation answer different questions. A model ranking unusual cases need not provide reliable probabilities for those cases. A calibrated credit-default estimate also concerns a different outcome from a substantiated accounting irregularity. The present forensic method therefore specifies the target condition before selecting a score, and prevents transfer of a probability between unrelated decision settings. Where a screening tool supplies only a rank, the case register records the rank and its explanation. Expected-value calculations remain conditional

examples until case probabilities have an appropriate evidential basis. This distinction preserves the practical use of screening while avoiding numerical confidence unsupported by the investigated population. [68]

Review economics and capacity

The decision rule for a calibrated case probability q is $V = q \times r \times L - K - (1-q) \times H$. V is the expected net benefit of review under the declared loss and cost assumptions. Positive V supports review, subject to evidence access and engagement scope. The model does not treat an arbitrary anomaly score as q . When calibration is unavailable, a reviewer uses documented qualitative bands and records uncertainty instead of inserting a fabricated probability. Legal duties, immediate evidence loss, and critical safety issues require separate escalation rules regardless of V .

For the cost example, L is \$50,000, r is 0.50, K is \$1,000, and H is \$200. The decision threshold is $q > (K+H)/(rL+H)$, which equals $1,200/25,200$, or approximately 4.76 percent. Remediation value means the fraction of exposure expected to be prevented or recovered because of review. Historical loss already incurred is not automatically recoverable. Changing r therefore changes the value of investigation even when suspicion is unchanged.

Economic triage also requires care in defining costs. Review cost includes analyst time, management interviews, document retrieval, and external specialist work. False-escalation cost might include unnecessary investigation and disruption. Some consequences are difficult to express in money. The model's H term is therefore an analytical convenience, not a complete measure of harm from an unsupported allegation. The engagement should preserve nonmonetary constraints and mandatory actions separately from the expected-value calculation.

Loss exposure also needs a time boundary. A historical irregular payment, a possible future recurrence, and an ongoing unsupported revenue pattern create different exposures. The r parameter concerns the portion review might prevent or recover. Setting r equal to one without evidence effectively assumes full recovery and prevention. Sensitivity analysis should vary this assumption and identify the information required to narrow the range. The adviser should also avoid adding several exposure measures which refer to the same underlying transactions.

The economic model also requires separate records for loss, recovery, and review expense. A suspicious payment amount is not automatically the amount lost. The business may have received partial value, recovered a duplicate, or withheld settlement. The case schedule distinguishes actual paid exposure, attempted exposure, uncertain valuation, credits, and realised recoveries. Review expense includes the relevant staff and external effort within the stated decision boundary. Nonfinancial consequences cannot always be priced credibly, so the parameter H is an illustrative representation of avoidable harm rather than a complete monetary valuation. Where the consequences are material, an institution can apply a documented constraint in addition to the expected-value calculation.

Capacity enters the method through cases and review time. An alert queue of five hundred items can represent fewer underlying episodes, but each episode can demand different effort. A capacity estimate should therefore record case consolidation, complexity, review hours, preservation needs, and the skills required. The model can prioritize a supported payment-control action separately from a deeper inquiry into intent. A temporary hold, a document request, and a full investigation have different costs and evidential requirements. Their authorisation should be explicit. Ranking the queue by one score alone would overlook both differences in review burden and obligations to examine severe unresolved matters that do not offer an immediate financial recovery.

RESULTS AND DISCUSSION

Evidence sufficiency before escalation

A case reconstruction illustrates the evidence requirement: an alert is insufficient without a reproducible connection to records. Consider an illustrative duplicate-payment alert. Two ledger entries share a supplier, amount, and close posting date. Follow-up might identify a reversed entry, a legitimate instalment, a reissued failed payment, or a duplicate disbursement. The reviewer needs bank movements, invoice numbers, purchase authorisation, and reversal history before determining the disposition. Similarity alone does not distinguish these possibilities.

The same approach applies to financial statement screening. Unusual receivable growth might arise from extended customer terms, a seasonal sales pattern, an acquisition, errors in ageing, or unsupported revenue. The adviser should identify the explanation requiring verification and the records which would support or contradict the explanation. The model

gives screening measures a discovery role while retaining evidence evaluation as the basis for conclusions. This division follows the concern with contextual interpretation in the screening literature without importing a published model as proof for a new client.

Table 1. Proposed evidence and case-disposition controls.

Stage	Required record	Decision boundary
Preservation	Authorised extract, date, source, access log	Integrity check does not establish original truth
Reproduction	Rule version and original values	Alert is a question for review
Corroboration	Supporting and contradictory documents	Missing records are an unresolved gap
Interpretation	Explanations examined and reviewer rationale	Misstatement and intent remain distinct
Disposition	Resolution, owner and follow-up	Separate correction, recovery and further assessment

Evidence quality and risk priority should remain separate dimensions. A high-exposure case with missing records deserves a preservation and access response. Missing evidence does not warrant reducing the case's risk score merely because review is difficult. Conversely, complete records might resolve a severe alert as a legitimate transaction. The model therefore reports both priority and evidence status. This prevents a single weighted score from concealing whether the adviser has reliable documentation.

Case dispositions should distinguish resolved legitimate activity, confirmed error, confirmed irregularity requiring further assessment, and unresolved concern. A finding of intentional deception requires additional support beyond a numerical deviation. Reports should describe the observed acts, relevant records, explanations evaluated, and remaining limitations. Management actions should follow the strength and scope of the evidence. The model's purpose is to support careful investigation rather than expand allegations beyond what the record demonstrates.

Correlated alerts create a further challenge. Several suspicious entries might arise from one vendor setup error. Counting each entry as an independent case increases apparent discovery and duplicates review cost. Conversely, one entry might reveal a process weakness

affecting many transactions. A linked-case register should identify the relationship and revise exposure without counting the same amount several times. The allocation decision then compares investigation units rather than raw alert rows.

Evidence contradictions require a recorded response. A manager's explanation might be consistent with the invoice but inconsistent with the bank transfer. A supplier confirmation might conflict with the purchase order. The reviewer should state which source supports each proposition and which discrepancy remains unresolved. A conclusion should not rely solely on the most convenient document. Preserving contradictions gives another reviewer a basis for assessing the strength of the finding and identifying further work.

Behavioural biases also affect the review. Once a case enters a high-risk queue, the analyst might give greater weight to evidence supporting the alert. Research on judgement under uncertainty provides a basis for considering this possibility [72] [73]. The proposed operational response is to record alternative explanations and require review of consequential findings by someone who did not create the original alert. The paper does not estimate the frequency of bias. It includes a control intended to make reasoning more inspectable.

Model explanations require equally careful language. A feature such as unusual year-end receivable growth helps identify why a case scored highly. Explainable AI research discusses the forms and limits of such explanations [74]. A model contribution is not evidence that receivable growth was artificial. The investigator needs the underlying transactions and business explanations. Reports should therefore describe features as analytical reasons for selection and reserve factual conclusions for the subsequent evidence record.

The evidence-preservation requirement is particularly useful when a case changes direction. An apparent invoice duplication might turn out to be a legitimate split delivery, but the review can reveal an unrelated authorisation weakness. The preserved record supports a control recommendation without forcing the original fraud hypothesis to remain central. Conversely, a routine posting error can coexist with an unexplained supplier relationship. Closing one discrepancy should not erase the second question. The case register therefore permits separate issues, owners, and dispositions within a linked episode. This structure supports learning from errors and control weaknesses while retaining the distinction between a corrected account and an allegation requiring further evidence.

False escalation should also receive a specific outcome explanation. A payment alert resolved through a documented invoice reversal differs from an alert dismissed because records were unavailable. The first supplies evidence of a legitimate explanation. The second retains uncertainty. Treating both as false positives would overstate the quality of outcome labels and distort future evaluation. A reviewer should use the disposition supported by the records, even when this produces a larger unresolved category. The model's numerical example assumes binary outcomes to make the arithmetic transparent. Operational reporting needs a richer set of categories and should explain how any evaluation translates them into a target outcome. The translation should be defined before model assessment, rather than chosen afterward to improve the reported performance measure.

The schematic represents the progression from a defined population through preservation, corroboration, interpretation, and disposition. Its uncertainty branch supplies a specific outcome when available records fail to support the proposed action. Its monitoring loop connects the disposition with correction, recovery, or further authorised review. This structure makes the analytical contribution visible without representing the drawing as evidence of an investigation. The paper's principal result remains conditional: screening usefulness depends on prevalence, review burden, evidence quality, and the consequence of the follow-up decision. Expanding the source discussion strengthens the design rationale but does not change the status of the numerical examples. Operational effectiveness should be established through authentic records, credible outcomes, and an evaluation aligned with the particular engagement.

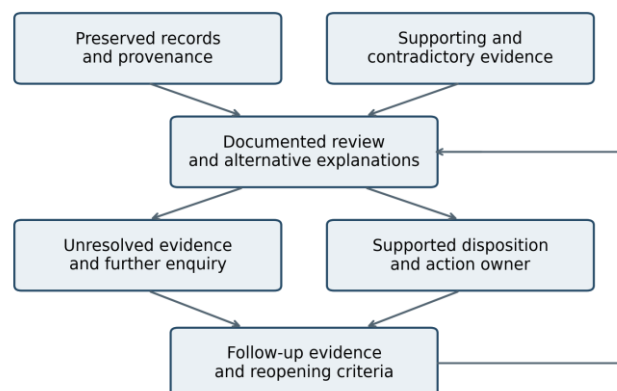


Figure 2. Evidence states and case disposition. Unresolved cases retain uncertainty; operational closure and evidential conclusions are distinct.

Base rates and the unresolved alert queue

At one-percent prevalence, the assumed screen identifies 80 of 100 relevant cases and misses 20. Five percent of the 9,900 other cases generate 495 false positives. The expected alert queue therefore contains 575 cases, with precision of 80/575, or 13.91 percent. Overall accuracy is 94.85 percent because 9,405 true negatives and 80 true positives constitute 9,485 correct classifications. The high accuracy coexists with an alert queue in which approximately 86 percent of cases are not relevant positives.

Table 2. Expected screening outcomes by assumed prevalence, N = 10,000, sensitivity = 80%, false-positive rate = 5%.

Prevalence	True positives	False positives	False negatives	Precision
0.5%	40	497.5	10	7.44%
1.0%	80	495	20	13.91%
5.0%	400	475	100	45.71%

At 0.5-percent prevalence, the same screening properties produce 40 expected true positives and 497.5 false positives. Precision falls to 7.44 percent. At five-percent prevalence, true positives increase to 400 and false positives fall to 475; precision becomes 45.71 percent. Holding sensitivity and the false-positive rate fixed does not hold investigative usefulness fixed. The calculation explains why a model assessed in a selected high-risk sample needs separate evaluation before use in routine transaction screening.

Prevalence also concerns the definition of the case population. A queue restricted to manual entries posted near period end differs from a complete sales population. Selecting a riskier population might increase precision, while leaving risks outside the selected process unexamined. The adviser should report which records entered the screen and which records were excluded. Changes in scope belong in the report because they change the meaning of the performance figures.

Reviewer capacity limits the value of additional alerts. If capacity is 100 reviews, randomly selecting 100 cases from the baseline queue of 575 gives an expected 13.91 relevant discoveries. This result does not establish the performance of a ranked queue. A ranking approach improves this yield only if higher-ranked cases contain a greater concentration of relevant conditions. The assumption needs testing through outcomes, rather than inferred from the size of an anomaly score.

The threshold comparison shows a related tradeoff. At one-percent prevalence, the broader screen produces 90 true positives and 990 false positives, for 1,080 alerts and 8.33-percent precision. The narrower screen produces 60 true positives and 198 false positives, for 258 alerts and 23.26-percent precision. Narrowing the screen reduces expected alert volume by 76.11 percent and raises precision, while missing 40 relevant cases instead of ten. Management cannot choose between thresholds responsibly without considering the consequences of missed cases and available review resources.

Table 3. Illustrative threshold tradeoff at 1% prevalence.

Screen	Sensitivity	False-positive rate	Alert cases	Precision	Missed cases
Broader	90%	10%	1,080	8.33%	10
Baseline	80%	5%	575	13.91%	20
Narrower	60%	2%	258	23.26%	40

The baseline results show how a technically appealing screen becomes operationally difficult. Five hundred seventy-five alerts exceed a capacity of one hundred reviews. The client consequently needs prioritisation, additional resources, narrower scope, or a longer engagement period. Each response changes the expected findings and costs. The model makes this allocation decision explicit rather than leaving reviewers to choose cases informally. Informal selection would make later performance difficult to interpret because the selection rule is unknown.

A threshold decision should include the unreviewed queue. Under the broader screen, 1,080 cases require follow-up, while the narrower screen produces 258. If management approves only one hundred reviews in either scenario, most alerts remain unresolved. Reporting the screen's expected precision without disclosing incomplete follow-up overstates practical coverage. The client needs the number investigated, number resolved, number pending, and basis for choosing the reviewed cases. These measures describe the engagement's actual work rather than the screen's theoretical potential.

The screening example demonstrates a base-rate problem that remains relevant even when the rule has apparently strong sensitivity. At one percent assumed prevalence, ninety-nine percent of the population does not have the target condition. Applying a five percent false-positive rate to that larger group produces four hundred and ninety-five false alerts. These

outnumber the eighty true alerts generated by the assumed eighty percent sensitivity. The practical implication is a substantial review burden, not a conclusion that screening is useless. The screen can still concentrate attention compared with reviewing every item. Its value depends on the subsequent procedure, the cost of review, and the consequences of the actions taken on those alerts.

Precision must also be interpreted at the correct stage. The illustrative 13.91 percent concerns the assumed screening condition, which might be a defined misstatement or irregularity. It is not the probability that an alerted person intended to defraud the institution. Intent requires different evidence and a different outcome definition. A sequence moving from an alert to a financial discrepancy and then to a supported account of conduct should report those stages separately. Combining them into a single fraud detection success rate would obscure the inferential steps. The schematic therefore shows alternative explanations and unresolved findings alongside escalation. It represents a process in which uncertainty is retained until the available evidence supports a narrower conclusion.

A capacity example makes the queue implications concrete. Suppose each initial alert review requires an average of two hours and the team has two hundred hours available in the review period. Under the baseline screen, five hundred and seventy-five alerts require an assumed eleven hundred and fifty hours before any consolidation or prioritisation. The team can review one hundred alerts in that period if the average time assumption holds. This creates a backlog, but it does not establish which one hundred should be selected. A documented prioritisation rule would consider severity, evidence preservation, immediate payment exposure, expected review value, and available skills. The two-hour estimate itself needs later comparison with actual effort across different case types.

Case consolidation can reduce repeated effort while preserving the original alert count. Suppose five duplicate-payment alerts relate to one supplier episode involving the same invoice reference and payment batch. Reviewing the episode once can avoid repeating the same document retrieval five times. However, consolidation must not erase distinctions among payments or conceal a wider pattern across suppliers. The case file retains all linked alert identifiers and the reason for consolidation. Evaluation can then report both alerts and unique cases. This allows management to distinguish a change in rule sensitivity from a change in workflow efficiency. A falling case count might reflect better consolidation rather than fewer suspicious conditions or a more accurate screening rule.

Loss exposure and the cost of investigation

Under the loss-and-cost assumptions, a case with q of 0.10 produces $V = 0.10 \times 0.50 \times \$50,000 - \$1,000 - 0.90 \times \$200 = \$1,320$. A case with q of 0.02 produces $V = \$500 - \$1,000 - \$196 = -\696 . These amounts arise from assumed calibrated probabilities. The first case passes the economic review threshold and the second does not. Neither calculation establishes whether misconduct occurred. The exercise identifies the value of acquiring and evaluating further evidence.

The remediation fraction changes this decision. If r falls to 0.10 while other assumptions remain fixed, the probability threshold rises to $1,200/5,200$, or 23.08 percent. A case previously selected at $q = 0.10$ would no longer pass the economic threshold. A recent payment to a recoverable counterparty differs from an old irreversible loss, even when both have similar likelihoods of irregularity. The model therefore requires the adviser to distinguish potential exposure, preventable future loss, and realistic recovery.

A queue with unequal review times requires another adjustment. When case A has higher expected benefit but takes substantially longer, reviewing A first might prevent completion of several smaller cases. Under a fixed time budget, the advisory problem becomes selecting cases to maximise aggregate expected value subject to review hours. Benefit per hour supplies a useful starting heuristic, but mandatory cases and linked investigations complicate simple ranking. The adviser should preserve these constraints rather than presenting a score as a complete allocation decision.

The decision threshold is also sensitive to exposure magnitude. Under the baseline r , K , and H , increasing L from \$50,000 to \$100,000 lowers the threshold to $\$1,200/\$50,200$, or 2.39 percent. Reducing L to \$10,000 raises the threshold to $\$1,200/\$5,200$, or 23.08 percent. Cases with equal estimated probability therefore need not have equal review priority. This arithmetic supports separating likelihood and consequence, while still requiring evidence for both values before applying the rule in practice.

A higher review cost changes the same comparison. If K rises from \$1,000 to \$3,000, the threshold becomes $\$3,200/\$25,200$, or 12.70 percent. A case at $q = 0.10$ then fails the purely economic rule. The outcome does not justify ignoring a legal duty or urgent evidence-preservation need. The model instead indicates when management needs another reason for review beyond the assumed financial payoff. Explicit reasoning improves the defensibility of both escalation and deferral.

The expected-value threshold is also specific to the contemplated action. The illustrative 4.76 percent threshold uses the stated loss, recovery, review expense, and avoidable-harm assumptions. It does not authorise every type of investigative intervention whenever an estimated probability exceeds that value. A low-cost request for routine supporting records differs from a payment suspension or an intrusive inquiry. An institution should define the action attached to the calculation and check whether it is authorised. The financial ranking can inform the allocation of review effort, while evidential and organisational requirements govern the action itself. Keeping those roles separate prevents a favourable expected return from being mistaken for proof or unrestricted permission.

Recovery timing can change the result without changing the nominal loss. Suppose the assumed recoverable amount is \$25,000 but recovery occurs after two years. At an illustrative ten percent annual discount rate, its present value is approximately \$20,661. With review cost of \$1,000 and avoidable harm of \$200, the corresponding probability threshold becomes about 5.75 percent rather than 4.76 percent. This is another deterministic scenario, not an estimate of recovery duration. It shows why the decision record needs expected timing as well as expected amount. A large possible recovery delayed for several years can support a different priority from an immediately preventable payment of the same nominal value.

A second extension concerns partial information about loss. Suppose a flagged payment is \$50,000, but records indicate that genuine goods worth \$30,000 were delivered. Treating the full payment as loss would overstate the financial discrepancy. The relevant amount depends on the supported mechanism, contractual value, and evidence of delivery. If the illustrative recoverable loss were instead \$20,000 and the assumed recovery fraction remained one half, the threshold with the same costs would be approximately 11.76 percent. This calculation does not determine the legal or contractual valuation. It shows that a defensible triage estimate requires a separate, supported loss schedule rather than automatic use of the gross transaction amount.

Review harm also has an organisational dimension. Repeated unsupported suspicions can consume supplier time, delay genuine payments, or damage confidence in routine controls. Some consequences will be difficult to quantify, and the model should not give a misleading impression that assigning \$200 captures them fully. The useful contribution is to require that such effects be considered and recorded. Management can define actions requiring a higher evidence threshold regardless of expected recovery. Review communications should state the

discrepancy being examined without claiming that the alert establishes misconduct. This practice protects the quality of the inquiry because people can respond to a specific question and supply records relevant to an alternative explanation.

A hypothetical queue example shows how evidential access changes scheduling without changing suspicion. Suppose two cases have the same assumed expected net review value and each initially requires four hours. The first has accessible source records. The second requires a separate extraction taking six additional hours. Under an eight-hour review budget, completing both is infeasible unless the extraction resource sits outside the budget. The scheduling decision should therefore specify which resource is constrained. Ranking both cases by the original four-hour estimate understates the second case's burden. This does not justify abandoning a severe case or weakening its risk classification. The model reports the access dependency and distinguishes urgent preservation from subsequent review, so management sees the additional resource needed to support the contemplated action.

Case linkage changes the same comparison. Suppose three alerts arise from one supplier episode and require two hours each when investigated independently. A consolidated episode review might need three hours because the same contract and bank record support all three alerts. The assumed saving is three hours, but the file should retain all original alerts and any differences among them. If one alert relates to a different delivery or entity, consolidation might be inappropriate. The example therefore demonstrates an opportunity contingent on documented linkage, rather than a universal consolidation benefit. Performance reporting should preserve both the raw alert count and the number of reviewed episodes. Otherwise, an apparent decrease in alerts might reflect administrative grouping instead of improved screening or fewer irregularities in the underlying transaction population.

The cost of a false escalation also depends on the contemplated action. Requesting a routine supporting document differs from suspending a supplier, withholding payment, or issuing an adverse report. A single assumed harm value cannot describe all these consequences. The review plan should therefore connect each action with its evidence requirement, approval, resource cost, and potential consequence. A preliminary document request might proceed while a stronger intervention waits for corroboration. The present model does not set a jurisdiction-specific evidentiary standard. It requires the institution to preserve the relationship between the action and the support available for that action. This

prevents a favourable expected-value calculation for information gathering from being interpreted as authority for a consequential allegation or commercial sanction.

Loss estimation requires the same separation. Suppose a hypothetical invoice of \$60,000 contains an unsupported \$10,000 component while the remaining \$50,000 corresponds to verified delivery. The exposure schedule should preserve the whole transaction and identify the component requiring examination. Reporting the entire invoice as loss would overstate the supported discrepancy. If the \$10,000 component later receives a legitimate explanation, the schedule records the revision and its evidence. If the discrepancy persists, recovery prospects still require their own assessment. A disputed amount is not automatically recoverable, and a recovered amount does not establish intent. This example uses assumed values to show the accounting relationship among transaction amount, questioned amount, supported discrepancy, and realised recovery. Each quantity should retain its separate evidential basis.

Review timing introduces another limit on benefit claims. A control response preventing a scheduled duplicate payment produces a different financial consequence from detecting the duplicate after settlement. The first might prevent an outflow if the intervention changes the payment decision before execution. The second creates a recovery task whose outcome depends on subsequent action and counterparty circumstances. Evaluation should therefore record the date of the alert, date of review, date of contemplated payment, and date of any correction or receipt. Without these dates, an engagement might count exposure detected as loss prevented even though the payment had already occurred. The model's expected remediation fraction represents this difference only as an assumption. Authentic evaluation needs the events and their sequence.

Control responses and sector interpretation

The model also highlights the difference between detection and remediation. A reviewer might confirm an error but lack authority to correct records or pursue recovery. A financially important finding therefore needs a responsible decision owner after investigation. An error corrected in the ledger differs from money recovered, future exposure prevented, or an allegation referred for further assessment. The benefit record should preserve these categories separately. Otherwise, confirmed exposure becomes an unsupported recovery claim.

An illustrative control redesign makes this distinction concrete. Suppose a review identifies duplicate supplier records which allow a similar invoice to be entered twice. The recommendation might require stable supplier identifiers, invoice validation, and separate approval for exceptions. These controls address a process weakness regardless of whether any reviewed entry establishes intentional misconduct. Their success requires monitoring future duplicate exceptions and the time needed to resolve legitimate payments. A lower alert count alone does not prove the control works if the rule has been changed to ignore more cases.

The model's feedback design supports useful evaluation. Every reviewed case records the original score, selected rule, review time, outcome, and evidentiary basis. The original score remains fixed for evaluation even when later investigation supplies new information. Outcomes also record when they became available. A future model might use these records for development, but assessment should use a separate period and a clear target condition. Without this separation, the same case history would serve as both the basis for prediction and proof of performance.

Control learning should extend beyond the detected case. A payroll error might reveal a missing approval process, an inconsistent staff register, or a weak reconciliation. These mechanisms call for different corrective actions. Organisational learning and knowledge-transfer research support documenting the relationship between the observed issue and the revised process [75] [76]. The adviser should test whether staff reproduce the correction and whether the new process handles legitimate exceptions. A policy document alone demonstrates neither capability nor effectiveness.

Forensic advisory is also a management service. Information systems success, data analytics competency, and organisational performance research support examining how information leads to decisions [77] [78] [79]. The present model uses that connection narrowly. An improved screen matters when a reviewer understands the alert, obtains the necessary evidence, and supports an appropriate response. Detection statistics, report completion, and implemented controls should therefore appear as different outcomes in engagement evaluation.

The methodological contribution does not depend on a particular algorithm. A rule-based screen, logistic model, or flexible classifier might enter the same triage process when its output and limitations are documented. More complex prediction is useful only if improved

ranking or calibration offsets added data, maintenance, and review burdens. The model gives the client a comparison based on relevant findings and resources. This focus prevents algorithm choice from becoming a substitute for defining the decision and available evidence.

Anomaly-detection models, internal industrial audit, commercial-facility governance, joint-venture alignment, and operating procedures concern different forms of organisational control. Their shared design relevance is the need to connect an exception with an accountable response. A high-risk score creates a question, while a procedure specifies who obtains records and evaluates the answer. Industrial control concepts are transferred here only at that organisational level. The paper does not assume industrial safety findings establish accounting-fraud detection. In the forensic register, a reviewer records scope, evidence access, unresolved gaps, and the approved next action. This structure makes the case inspectable and identifies when a specialised review is required. It also prevents an incomplete file from receiving an adverse conclusion simply to close the queue. [80] [81] [82] [83] [84]

Privacy governance, impact assessment, IT general controls, business intelligence, and compliance-oriented automation define important conditions for handling financial evidence. A forensic engagement should collect the records authorised for its question and preserve their provenance. The availability of an employee communication or system log does not justify unrelated collection. The model records access authority separately from analytical relevance and uses traceable transformation rules. The cited frameworks motivate this separation without establishing the legal requirements for every jurisdiction. Case findings still depend on the records reviewed and the applicable engagement conditions. The practical result is a narrower and more reproducible investigation, with evidence gaps disclosed rather than filled by unverified assumptions about what a system or employee probably did. [85] [86] [87] [88] [89]

Technology governance, identity controls, access integration, configuration governance, and supply-chain systems affect the reliability of accounting extracts. A role change might alter who enters or approves a transaction. A software configuration change might alter how the same transaction appears in a report. These possibilities become alternative explanations for an anomaly and reasons to preserve rule versions. The proposed model asks whether the data are complete, how the extract was produced, and which changes occurred during the examined period. It does not presume a technology weakness proves financial misconduct.

The screen and the evidence process therefore retain separate control checks, so an analyst does not interpret a formatting or integration problem as an unsupported finding about the underlying economic event. [90] [91] [92] [93] [94]

Communications forensics, anti-corruption analytics, data mining, banking fraud detection, and programme-audit prediction directly inform the screening problem. These sources cover different data populations and target conditions, so their results should not be treated as interchangeable. The present model uses their topical contributions to define a follow-up process, while the numerical exercise declares its own assumed prevalence and error rates. A communication pattern might prompt questions, a transaction model might rank cases, and a programme audit might require documentary confirmation. Each requires its own denominator and outcome definition. The engagement should report which mechanism was used, what evidence followed, and what remained unresolved. This prevents an aggregate detection claim from combining unrelated methods and obscuring the strength of individual findings. [95] [96] [97] [98] [99]

Payroll-and-receivable analytics, customer-service measures, accountability KPIs, unstructured-record analysis, and the power-durability study provide a final connection between financial records and operating explanations. A repair payment, delayed service, or interrupted system might be legitimate but require supporting records. The energy study is used only as context for operational interruption, rather than proof of a particular client's cause or cost. The reviewer should obtain maintenance, approval, and payment records before accepting the explanation. Similarly, narrative records require source checking rather than uncritical interpretation by a text model. The proposed register keeps the explanation, corroboration, and disposition distinct. This supports a defensible follow-up process without treating technical context or organisational indicators as conclusive evidence of accounting misconduct. [100] [101] [102] [103] [104]

Performance reporting should avoid rewarding the team solely for the number of cases escalated. Such a measure can encourage weak escalation and make well-supported closure appear unproductive. Useful measures include review time, reproducibility, correction accuracy, unresolved evidence gaps, realised recovery, and supported explanations. A missed-case measure requires an appropriate review of unalerted items, as described in the method. A prevention benefit requires a stated comparison because an avoided payment is not automatically an avoided loss. The institution should preserve the assumptions underlying

any estimate. These measures remain candidates for empirical evaluation, and the paper does not claim that their implementation will produce a particular improvement in investigative performance.

The proposed acceptance exercise should include cases capable of revealing specific workflow defects. A reversed payment checks whether the rule interprets reversal history. A linked set of invoices checks whether the case process consolidates appropriately. A missing delivery document checks whether the disposition preserves uncertainty. A legitimate repeated payment checks whether corroboration prevents an unsupported conclusion. A revised source extract checks whether the reviewer recognises a version difference. These scenarios test different requirements and should receive separate outcomes. Passing all scenarios supports readiness against the stated requirements. It does not demonstrate sensitivity, specificity, fairness, or economic benefit across the client population. Those questions require an evaluated population with credible outcomes and a comparison appropriate to the claim being made.

Forensic accounting advisory in energy businesses requires an account of the operating events behind financial exceptions. Nwangwu and colleagues' drilling-integrity framework links technical barriers, competence, compliance learning and integrated governance [105]. Their technical learning-center review emphasizes employer participation, workplace learning and shared governance [106]. Together, these conceptual studies inform interpretation of operational records and staff responsibilities. Investigators should establish who understood a measurement, who authorized a classification and who had access to the underlying records. A training deficit or process failure offers a possible explanation for an exception, but neither establishes an absence of misconduct. The investigator must test explanations against contemporaneous evidence.

The dynamic-filtration model coauthored by Nwangwu supplies a technical example of nonlinear operating behavior [107]. Investigators examining drilling costs or production shortfalls should distinguish an engineering explanation from evidence of manipulated reporting. The reservoir forecasting review by Nwangwu and colleagues emphasizes calibration integrity, uncertainty and scenario coherence [108]. Those concerns support preserving forecast versions and the technical assumptions which informed financial estimates. Their capital budgeting review connects sanctioned assumptions with staged investment and execution readiness [109]. A forensic review should consequently compare

authorization records, purchase commitments and observed execution rather than treating a cost overrun as proof of fraud. The connection to forensic accounting is this paper's interpretation; the cited studies do not test fraud-detection performance.

Nwangwu and colleagues' resource-discovery review emphasizes the curation and interpretation of longitudinal asset records [110]. This supports careful reconstruction of operational history when changing reserve or recovery assumptions affect asset valuations. Their energy-acquisition review brings technical characterization, commercial obligations, supply-chain integration and post-close governance into one conceptual account [112]. Forensic advisers examining acquisitions should reconcile technical reports with contract obligations, assumed liabilities and valuation revisions. Differences need a dated explanation and a responsible reviewer. Neither conceptual review establishes whether a particular reserve revision, impairment or transaction represents misstatement. Those conclusions require the transaction records and an independent assessment of the specific evidence.

Adebajo and colleagues, including Nwangwu, describe Oando Energy Resources' production-capacity digitalization workflow [111]. The platform brings multidisciplinary asset information together with validation, diagnosis and analysis to support production-restoration decisions. Its relevance to forensic advisory concerns the traceability of deferred-volume classifications and their supporting inputs. Investigators should distinguish technical potential, reported production and recoverable value, then retain the source measurements and classification changes behind each estimate. This published operational implementation is not a validation of the present forensic framework. No detection rate, fraud reduction or financial return is inferred from the description. The practical lesson is to connect financial exceptions with the underlying operational record before deciding whether further investigation is warranted.

Outcome labels and comparative evaluation

Feedback creates a measurement challenge. Reviewed cases receive labels; most unreviewed cases remain unknown. Training a future model solely on the reviewed queue risks learning the previous reviewers' selection habits. An independent, authorised sample of lower-ranked cases would provide information about missed relevant conditions. Review outcomes should also distinguish a lack of evidence from a verified negative finding. Without these distinctions, performance statistics overstate certainty and obscure false negatives.

Time introduces another source of bias. Investigations sometimes resolve long after the transaction date. Evaluating a screen using labels unavailable at the time of prediction creates an unrealistic information set. A future assessment should preserve the prediction date, the information available then, and the date on which outcomes became known. A later test period should follow an earlier development period. Records from the same linked case should remain together to prevent leakage between development and evaluation sets.

Three propositions follow from the model. At constant sensitivity and false-positive rate, lower prevalence reduces expected precision. At fixed loss and review cost, lower effective remediation raises the minimum probability required for an economically justified investigation. At fixed capacity, better discrimination within an alert queue should improve relevant findings per review hour. The first two are mathematical implications of declared assumptions. The third concerns empirical ranking quality and requires evidence from an appropriate evaluation.

The model also has practical limits. A monetary objective does not capture all reporting obligations, reputational damage, or the need to preserve evidence urgently. Probabilities, losses, and recoveries are difficult to estimate. The sensitivity exercise holds screening properties constant even though they often change across populations. No original transaction data have been analysed. The paper therefore provides a transparent advisory design and reproducible calculations, without reporting validated fraud detection, recovered losses, or disciplinary outcomes.

A subsequent study should define the target condition before analysis, use authorised historical data, and document how independent reviewers adjudicate outcomes. Evaluation should compare the proposed triage sequence with an existing workflow using relevant findings per review hour, unresolved cases, missed conditions in sampled unreviewed records, and evidence completeness. Such a study should keep screening staff and outcome adjudication sufficiently distinct to limit confirmation bias. Client-specific legal and professional requirements would need assessment within the actual engagement.

In a future field assessment, error and fraud outcomes should remain separate. The relevant target might be a material misstatement regardless of intent, an unauthorised payment, or independently substantiated misconduct. Mixing these labels changes prevalence and obscures performance. An authorised independent adjudication process should define sufficient evidence for each category and record disagreements. The resulting dataset would

support meaningful estimates of precision, missed conditions, review efficiency, and remedial outcomes for the specified engagement.

Future validation should examine how reviewers use the proposed process, not only whether an algorithm separates labelled records. An authorised study could compare case files prepared under different documented review procedures, using blinded assessment of reproducibility and explanation quality where feasible. It could measure the time needed to reconstruct a conclusion and the frequency with which alternative explanations are considered. Outcome labels would need independent adjudication and an explicit treatment of unresolved cases. Institutional differences in records, review powers, and recovery options would also affect transferability. The present model supplies definitions and conditional arithmetic for that work, while leaving measured detection rates, financial returns, and causal claims for a properly designed evaluation.

Future empirical evaluation should define the unit of comparison before work begins. A comparison between two screening rules requires the same eligible population, target condition, and review basis. A comparison between two investigation procedures requires comparable cases, reviewer resources, evidence access, and outcome assessment. Comparing a high-risk selected queue with a routine ledger review would otherwise confound selection with procedure. The study should also retain cases not completed during the observation period and explain how unresolved outcomes affect interpretation. Where independent outcome adjudication is feasible, adjudicators should receive the evidence needed for the conclusion without unnecessary exposure to the original screening score. This design would test whether the process improves findings and reproducibility rather than whether reviewers agree with the tool supplying their original suspicion.

CONCLUSION

The principal finding is an evidential boundary: a numerical alert supplies a reason to review records, while a supported disposition requires corroboration and a documented assessment of alternatives. Under the declared baseline, 575 alerts include 80 relevant positives. Precision is 13.91 percent even though overall accuracy reaches 94.85 percent. The discrepancy demonstrates why queue composition and available review hours deserve explicit attention.

Investigation planning should retain unresolved cases, distinguish error from substantiated misconduct and identify the action behind each cost or recovery estimate. The economic rule informs resource allocation alongside the engagement's evidence and escalation requirements. Evaluation of the proposed process requires authorised records, independent outcome assessment and a defined comparison workflow. Until those observations exist, the contribution is a case-review design and its conditional arithmetic.

REFERENCES

- [1] Beneish, M. D. (1999). The Detection of Earnings Manipulation. *Financial Analysts Journal*, 55(5), 24-36. <https://doi.org/10.2469/faj.v55.n5.2296>
- [2] Dechow, P. M., Ge, W., Larson, C. R., & Sloan, R. G. (2011). Predicting Material Accounting Misstatements. *Contemporary Accounting Research*, 28(1), 17-82. <https://doi.org/10.1111/j.1911-3846.2010.01041.x>
- [3] Perols, J. (2011). Financial Statement Fraud Detection: An Analysis of Statistical and Machine Learning Algorithms. *AUDITING: A Journal of Practice & Theory*, 30(2), 19-50. <https://doi.org/10.2308/ajpt-50009>
- [4] Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*, 50(3), 559-569. <https://doi.org/10.1016/j.dss.2010.08.006>
- [5] West, J., & Bhattacharya, M. (2016). Intelligent financial fraud detection: A comprehensive review. *Computers & Security*, 57, 47-66. <https://doi.org/10.1016/j.cose.2015.09.005>
- [6] Saito, T., & Rehmsmeier, M. (2015). The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLOS ONE*, 10(3), e0118432. <https://doi.org/10.1371/journal.pone.0118432>
- [7] Bolton, R. J., & Hand, D. J. (2002). Statistical Fraud Detection: A Review. *Statistical Science*, 17(3). <https://doi.org/10.1214/ss/1042727940>
- [8] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection. *ACM Computing Surveys*, 41(3), 1-58. <https://doi.org/10.1145/1541880.1541882>
- [9] Marais, A., Vermaak, C., & Shewell, P. (2023). Predicting financial statement manipulation in South Africa: A comparison of the Beneish and Dechow models. *Cogent Economics & Finance*, 11(1), 2190215. <https://doi.org/10.1080/23322039.2023.2190215>
- [10] Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, 3(4), 305-360. [https://doi.org/10.1016/0304-405x\(76\)90026-x](https://doi.org/10.1016/0304-405x(76)90026-x)
- [11] Eisenhardt, K. M. (1989). Agency Theory: An Assessment and Review. *The Academy of Management Review*, 14(1), 57. <https://doi.org/10.2307/258191>
- [12] Healy, P. M., & Palepu, K. G. (2001). Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *Journal of*

- Accounting and Economics, 31(1-3), 405-440. [https://doi.org/10.1016/s0165-4101\(01\)00018-0](https://doi.org/10.1016/s0165-4101(01)00018-0)
- [13] Dechow, P., Ge, W., & Schrand, C. (2010). Understanding earnings quality: A review of the proxies, their determinants and their consequences. *Journal of Accounting and Economics*, 50(2-3), 344-401. <https://doi.org/10.1016/j.jacceco.2010.09.001>
- [14] Dechow, P. M., & Dichev, I. D. (2002). The Quality of Accruals and Earnings: The Role of Accrual Estimation Errors. *The Accounting Review*, 77(s-1), 35-59. <https://doi.org/10.2308/accr.2002.77.s-1.35>
- [15] Brazel, J. F., Jones, K. L., & Zimbelman, M. F. (2009). Using Nonfinancial Measures to Assess Fraud Risk. *Journal of Accounting Research*, 47(5), 1135-1166. <https://doi.org/10.1111/j.1475-679x.2009.00349.x>
- [16] Farber, D. B. (2005). Restoring Trust after Fraud: Does Corporate Governance Matter? *The Accounting Review*, 80(2), 539-561. <https://doi.org/10.2308/accr.2005.80.2.539>
- [17] Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32. <https://doi.org/10.1023/a:1010933404324>
- [18] Bao, Y., Ke, B., Li, B., Yu, Y. J., & Zhang, J. (2020). Detecting Accounting Fraud in Publicly Traded U.S. Firms Using a Machine Learning Approach. *Journal of Accounting Research*, 58(1), 199-235. <https://doi.org/10.1111/1475-679x.12292>
- [19] Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206-215. <https://doi.org/10.1038/s42256-019-0048-x>
- [20] Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2019). A Survey of Methods for Explaining Black Box Models. *ACM Computing Surveys*, 51(5), 1-42. <https://doi.org/10.1145/3236009>
- [21] Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Communications of the ACM*, 45(4), 211-218. <https://doi.org/10.1145/505248.506010>
- [22] Redman, T. C. (1998). The impact of poor data quality on the typical enterprise. *Communications of the ACM*, 41(2), 79-82. <https://doi.org/10.1145/269012.269025>
- [23] Eppler, M. J., & Mengis, J. (2004). The Concept of Information Overload: A Review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines. *The Information Society*, 20(5), 325-344. <https://doi.org/10.1080/01972240490507974>
- [24] Beasley, M. S. (1996). An Empirical Analysis of the Relation Between the Board of Director Composition and Financial Statement Fraud. *The Accounting Review*, 71(4), 443-465. <https://doi.org/10.2308/tar-9611271988>
- [25] Abbott, L. J., Park, Y., & Parker, S. (2000). The effects of audit committee activity and independence on corporate fraud. *Managerial Finance*, 26(11), 55-68. <https://doi.org/10.1108/03074350010766990>
- [26] Jones, J. J. (1991). Earnings Management During Import Relief Investigations. *Journal of Accounting Research*, 29(2), 193. <https://doi.org/10.2307/2491047>

- [27] Gupta, M., & George, J. F. (2016). Toward the development of a big data analytics capability. *Information & Management*, 53(8), 1049-1064. <https://doi.org/10.1016/j.im.2016.07.004>
- [28] Mikalef, P., Pappas, I. O., Krogstie, J., & Giannakos, M. (2018). Big data analytics capabilities: a systematic literature review and research agenda. *Information Systems and e-Business Management*, 16(3), 547-578. <https://doi.org/10.1007/s10257-017-0362-y>
- [29] Mikalef, P., & Gupta, M. (2021). Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance. *Information & Management*, 58(3), 103434. <https://doi.org/10.1016/j.im.2021.103434>
- [30] Raisch, S., & Krakowski, S. (2021). Artificial Intelligence and Management: The Automation-Augmentation Paradox. *Academy of Management Review*, 46(1), 192-210. <https://doi.org/10.5465/amr.2018.0072>
- [31] Grant, R. M. (1996). Toward a knowledge-based theory of the firm. *Strategic Management Journal*, 17(S2), 109-122. <https://doi.org/10.1002/smj.4250171110>
- [32] Cohen, W. M., & Levinthal, D. A. (1990). Absorptive Capacity: A New Perspective on Learning and Innovation. *Administrative Science Quarterly*, 35(1), 128. <https://doi.org/10.2307/2393553>
- [33] Obukadata, V. O., Ezeako, R., Okoli, I. E., & Awulu, I. (2022). Measuring Change Program Success: A Review of Multi-Dimensional Evaluation Frameworks in Program Management. *International Journal of Multidisciplinary Evolutionary Research*, 3(1), 158-176. <https://doi.org/10.54660/ijmer.2022.3.1.158-176>
- [34] Okoli, I. E., Uzundu, C. S., Alugwe, S., & Awulu, I. (2022). Quality Management Systems in Large-Scale Operations: A Conceptual Framework Integrating Six Sigma, Lean, and Data Analytics. *International Journal of Multidisciplinary Evolutionary Research*, 3(1), 139-157. <https://doi.org/10.54660/ijmer.2022.3.1.139-157>
- [35] Okoli, I. E., Uzundu, C. S., Ibeh, A. V., & Alugwe, S. (2023). First Pass Yield and Defect-Driver Analysis in Distributed Operations; A Conceptual Framework for Controllable-Variable Isolation. *International Journal of Multidisciplinary Futuristic Development*, 4(1), 125-141. <https://doi.org/10.54660/ijmfd.2023.4.1.125-141>
- [36] Mensah, K. O., Kinyanjui, A., Okoli, I., Agboada, P., & Pedro, O. (2023). Advances in Asset Liability Management and Interest Rate Risk Modeling in Retail and Investment Banking. *Iconic Research and Engineering Journals*, 6(9), 583-618. <https://doi.org/10.64388/irev6i9-1723268>
- [37] Mensah, K. O., Pedro, O., Kinyanjui, A., Okoli, I., & Agboada, P. (2019). A Review of Capital Adequacy Frameworks and Risk Weighted Asset Optimization in Modern Banking Systems. *Iconic Research and Engineering Journals*, 3(6), 731-768. <https://doi.org/10.64388/irev3i6-1723267>
- [38] Mensah, K. O., Pedro, O., Kinyanjui, A., Okoli, I., & Agboada, P. (2019). A Conceptual Model for Integrated Enterprise Risk Management and Capital Allocation in Financial Institutions. *Iconic Research and Engineering Journals*, 3(1), 684-719. <https://doi.org/10.64388/irev3i1-1723266>

- [39] Oyediji, K. E., Adesanya, O. S., & Okoli, I. E. (2019). A Forensic Investigation Model for Detecting Vendor Collusion and Theft in Nigerian LNG Procurement. *International Journal of Humanities and Information Technology*, (Special 3), 67-78.
- [40] Oyediji, K. E., Adesanya, O. S., & Okoli, I. E. (2018). An Evidence Based Framework for Integrity Due Diligence on Board Nominees in Nigerian Banks. *International Journal of Technology Management & Humanities*, 4(4), 24-35.
- [41] Agboada, P., Okoli, I., Kinyanjui, A., Akintobi, A., & Pedro, O. (2022). Correspondent Banking and Trade Finance Risk Management in West African Banking Networks. *International Journal of Science, Technology & Society*, 6(1), 29-59.
- [42] Pedro, O., Agboada, P., Okoli, I., Kinyanjui, A., & Akintobi, A. (2023). Operational Risk and Regulatory Compliance in Third Party Fund Administration Services. *International Journal of Science, Technology & Society*, 7(1), 1-27.
- [43] Eboh, E. E., & Aliliele, C. (2023). Portfolio Risk Optimization Models for Venture Capital Investment Strategy and Fraud Exposure Analysis. *Iconic Research and Engineering Journals*, 7(4), 759-792. <https://doi.org/10.64388/irev7i4-1715305>
- [44] Aliliele, C., Mbonu, I. S., & Iwuanyanwu, U. (2023). A Review of API Governance and Risk Prioritization Frameworks in Modern Financial Institutions. *International Journal of Scientific Research in Computer Science Engineering and Information Technology*, 395. <https://doi.org/10.32628/cseit2361072>
- [45] Uchechi Mary-Linda Unamma, U. M. L. U., & Funmilayo Ashore-Onisemo, F. A. O. (2023). Fairness and Bias Auditing of Decision Models in Credit and Service Delivery: A Review. *Gyanshauryam International Scientific Refereed Research Journal*, 594. <https://doi.org/10.32628/gisrrj236341>
- [46] Ashore-Onisemo, F., Unamma, U. M. L., & Iziduh, E. F. (2021). Reconceptualizing Cost of Poor Quality as a Cross-Sector Governance Metric. *International Journal of Multidisciplinary Research and Growth Evaluation*, 2(6), 1119-1136. <https://doi.org/10.54660/ijmrge.2021.2.6.1119-1136>
- [47] Okwesa, I. J., & Ashore-Onisemo, F. (2023). Data Governance and Auditability in High-Volume Financial Operations: A Review. *Iconic Research and Engineering Journals*, 6(8), 450-488. <https://doi.org/10.64388/irev6i8-1722630>
- [48] Akomolafe, O., Uzoma Agu, M., & Bello, A. (2023). A Conceptual Model for Implementing Risk-Based Auditing in Strategic Financial Management. *International Journal of Advanced Multidisciplinary Research and Studies*, 3(6), 2274-2286. <https://doi.org/10.62225/2583049x.2023.3.6.5359>
- [49] Okonkwo, C. S., Ogunwole, O., Okeke, O. T., & Mayo, W. (2019). Conceptual Framework for Cost Reduction Through Contract Negotiation and Vendor Governance. *Iconic Research and Engineering Journals*, 2(9), 468-482. <https://doi.org/10.64388/irev2i9-1713121>
- [50] Ike, P. N., Ogbuefi, E., Nnabueze, S. B., Olatunde-Thorpe, J., Aifuwa, S. E., Oshoba, T. O., & Akokodaripon, D. (2021). Supplier Relationship Management Strategies Fostering Innovation, Collaboration, and Resilience in Global Supply Chain Ecosystems. *International Journal of Multidisciplinary Evolutionary Research*, 2(2), 52-62. <https://doi.org/10.54660/ijmer.2021.2.2.52-62>

- [51] Adelanwa, A., Basnet, A., & Nkechinyere Anene, U. (2023). Predictive Analytics Models for Financial Risk Detection and Fraud Prevention in Public Systems. *International Journal of Advanced Multidisciplinary Research and Studies*, 3(6), 2663-2679. <https://doi.org/10.62225/2583049x.2023.3.6.5969>
- [52] Akomolafe, O., Olaogun, B. O., Adesuyi, M. O., Ndukwe, V. U., & Sakyi, J. K. (2022). Smart Contract Automation Model for Supplier Payment Systems and Performance Benchmarking. *International Journal of Multidisciplinary Research and Growth Evaluation*, 3(6), 827-836. <https://doi.org/10.54660/ijmrge.2022.3.6.827-836>
- [53] Aifuwa, S. E., Oshoba, T. O., Ogbuefi, E., Ike, P. N., Nnabueze, S. B., & Olatunde-Thorpe, J. (2020). Predictive Analytics Models Enhancing Supply Chain Demand Forecasting Accuracy and Reducing Inventory Management Inefficiencies. *International Journal of Multidisciplinary Research and Growth Evaluation*, 1(3), 171-181. <https://doi.org/10.54660/ijmrge.2020.1.3.171-181>
- [54] Ebhojie, O., A Dogbatsey, E., & Oluwafemi Oyeleye, A. (2023). Audit Liaison, Corrective Action Planning, and Control Compliance in Multinational Organisations: A Systematic Literature Review. *International Journal of Advanced Multidisciplinary Research and Studies*, 3(6), 2839-2850. <https://doi.org/10.62225/2583049x.2023.3.6.6200>
- [55] Uzoma Agu, M., Akomolafe, O., & Bello, A. (2023). A Comparative Review of SOX Compliance Frameworks in Cross-Border Financial Auditing. *International Journal of Advanced Multidisciplinary Research and Studies*, 3(6), 2297-2306. <https://doi.org/10.62225/2583049x.2023.3.6.5362>
- [56] Ejimah, A. A., & Aregbe, O. O. (2020). Assessment of solar energy-based electronic system for sustainable rural electrification in Giwa Local Government Area of Kaduna State, Nigeria. *Global Journal of Applied, Management and Social Sciences*, 20, 264-275.
- [57] Ejimah, A. A., & Aregbe, O. O. (2022). Feasibility study and design of a piezoelectric energy harvesting system for self-powered street lighting on Nigerian high-traffic highways. *Global Journal of Applied, Management and Social Sciences*, 24, 339-347.
- [58] Jaakkola, E. (2020). Designing conceptual articles: four approaches. *AMS Review*, 10(1-2), 18-26. <https://doi.org/10.1007/s13162-020-00161-0>
- [59] Hand, D. J. (2006). Classifier Technology and the Illusion of Progress. *Statistical Science*, 21(1). <https://doi.org/10.1214/0883423060000000060>
- [60] Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM Computing Surveys*, 46(4), 1-37. <https://doi.org/10.1145/2523813>
- [61] Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104, 333-339. <https://doi.org/10.1016/j.jbusres.2019.07.039>
- [62] Torraco, R. J. (2005). Writing Integrative Literature Reviews: Guidelines and Examples. *Human Resource Development Review*, 4(3), 356-367. <https://doi.org/10.1177/1534484305278283>

- [63] 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. <https://doi.org/10.1111/1467-8551.00375>
- [64] Wang, R. Y., & Strong, D. M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, 12(4), 5-33. <https://doi.org/10.1080/07421222.1996.11518099>
- [65] Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). Data quality in context. *Communications of the ACM*, 40(5), 103-110. <https://doi.org/10.1145/253769.253804>
- [66] Mosonyi, S., Empson, L., & Gond, J. P. (2020). Management Consulting: Towards an Integrative Framework of Knowledge, Identity, and Power. *International Journal of Management Reviews*, 22(2), 120-149. <https://doi.org/10.1111/ijmr.12218>
- [67] Sturdy, A. (2011). Consultancy's Consequences? A Critical Assessment of Management Consultancy's Impact on Management. *British Journal of Management*, 22(3), 517-530. <https://doi.org/10.1111/j.1467-8551.2011.00750.x>
- [68] Funmilayo Ashore-Onisemo, F. A. O., Ifeanyichukwu Jeffrey Okwesa, I. J. O., & Uchechi Mary-Linda Unamma, U. M. L. U. (2020). Machine Learning for Credit Risk and Default Prediction in Retail and SME Lending : A Systematic Review. *International Journal of Scientific Research in Computer Science Engineering and Information Technology*, 477. <https://doi.org/10.32628/cseit2066446>
- [69] Bello, A., Akomolafe, O., & Uzoma Agu, M. (2023). A Conceptual Framework for Data-Driven Procurement Risk Management in Multinational Enterprises. *International Journal of Advanced Multidisciplinary Research and Studies*, 3(6), 2307-2316. <https://doi.org/10.62225/2583049x.2023.3.6.5363>
- [70] Eyetsemitan, R. A., Ambali, K. B., Oyeleye, A. O., & Fadayomi, O. (2023). User Acceptance Testing in Small Business Technology Deployment: A Structured Validation Framework for Lean Operational Environments. *International Journal of Multidisciplinary Research and Growth Evaluation*, 4(6), 1512-1531. <https://doi.org/10.54660/ijmrge.2023.4.6.1512-1531>
- [71] Oluwasanjo Ladapo, O., A Dosunmu, A., Jooda, D., & O Abolaji, T. (2023). An IT Service Management Literature Review: Challenges, Benefits, Opportunities, and Implementation Practices. *International Journal of Advanced Multidisciplinary Research and Studies*, 3(6), 2875-2889. <https://doi.org/10.62225/2583049x.2023.3.6.6205>
- [72] Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157), 1124-1131. <https://doi.org/10.1126/science.185.4157.1124>
- [73] Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2), 263. <https://doi.org/10.2307/1914185>
- [74] Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82-115. <https://doi.org/10.1016/j.inffus.2019.12.012>
- [75] Argyris, C. (1977). Organizational learning and management information systems. *Accounting, Organizations and Society*, 2(2), 113-123. [https://doi.org/10.1016/0361-3682\(77\)90028-9](https://doi.org/10.1016/0361-3682(77)90028-9)

- [76] Szulanski, G. (1996). Exploring internal stickiness: Impediments to the transfer of best practice within the firm. *Strategic Management Journal*, 17(S2), 27-43. <https://doi.org/10.1002/smj.4250171105>
- [77] DeLone, W. H., & McLean, E. R. (2003). The DeLone and McLean Model of Information Systems Success: A Ten-Year Update. *Journal of Management Information Systems*, 19(4), 9-30. <https://doi.org/10.1080/07421222.2003.11045748>
- [78] Ghasemaghaei, M., Ebrahimi, S., & Hassanein, K. (2018). Data analytics competency for improving firm decision making performance. *The Journal of Strategic Information Systems*, 27(1), 101-113. <https://doi.org/10.1016/j.jsis.2017.10.001>
- [79] Melville, N., Kraemer, K., & Gurbaxani, V. (2004). Review: Information Technology and Organizational Performance: An Integrative Model of IT Business Value1. *MIS Quarterly*, 28(2), 283-322. <https://doi.org/10.2307/25148636>
- [80] Maryann Inimfon Atakpa, M. I. A. I. A., & Nyiauwung Fobellah Abetoh, N. F. A. (2023). Anomaly Detection in Financial Time-Series Data : A Conceptual Model for Healthcare and Banking Applications. *International Journal of Scientific Research in Computer Science Engineering and Information Technology*, 967. <https://doi.org/10.32628/cseit2342441>
- [81] Obogo, S. F., Arumosoye, O. M., & Obriki, O. D. (2020). Advances in Internal QHSE Audit Systems for Industrial Engineering Operations. *Iconic Research and Engineering Journals*, 4(4), 399-417. <https://doi.org/10.64388/irev4i4-1715499>
- [82] Obogo, S. F., Obriki, O. D., & Arumosoye, O. M. (2022). Conceptual Safety Governance Model for Large Commercial Facility Operations. *Gyanshauryam International Scientific Refereed Research Journal*, 383. <https://doi.org/10.32628/gisrrj225639>
- [83] Eyetsemitan, R. A., Ambali, K. B., Oyeleye, A. O., & Fadayomi, O. (2020). Multi-Stakeholder Governance Alignment in Joint Venture Operations: A Conceptual Framework for Coordinating Business Processes in Highly Regulated Environments. *Iconic Research and Engineering Journals*, 4(4), 418-441. <https://doi.org/10.64388/irev4i4-1716955>
- [84] Ruth Arogbeoritse Eyetsemitan, R. A. E., Ajibola Oluwafemi Oyeleye, A. O. O., Kazeem Babatunde Ambali, K. B. A., & Oladapo Fadayomi, O. F. (2022). Standard Operating Procedures as Strategic Assets in Small Business Operations : A Systematic Review and Implementation Framework. *Gyanshauryam International Scientific Refereed Research Journal*, 438. <https://doi.org/10.32628/gisrrj225356>
- [85] Mbonu, I. S., Aliliele, C., Iwuanyanwu, U., & Oluoha, O. M. (2018). A Conceptual Framework for Legal and Ethical Risk Modeling in Enterprise Data Protection Governance Systems. *Iconic Research and Engineering Journals*, 2(2), 207-226. <https://doi.org/10.64388/irev2i2-1714911>
- [86] Mbonu, I. S., Iwuanyanwu, U., Aliliele, C., & Uzoka, E. (2022). A Review of Data Protection Impact Assessment Models in Multi Cloud Financial Infrastructure Systems. *International Journal of Scientific Research in Computer Science Engineering and Information Technology*, 589. <https://doi.org/10.32628/cseit25442>
- [87] Mbonu, I. S., Aliliele, C., Iwuanyanwu, U., & Uzoka, E. (2022). A Conceptual Framework for AI Enabled IT General Controls and SOX Audit Automation Processes.

- Gyanshauryam International Scientific Refereed Research Journal, 384.
<https://doi.org/10.32628/gisrrj2256239>
- [88] Stephanie Mbonu, I., Aliliele, C., Iwuanyanwu, U., & Uzoka, E. (2021). A Conceptual Framework for Risk Based Business Intelligence Architecture in Financial Technology Platforms. *International Journal of Multidisciplinary Research and Growth Evaluation*, 2(6), 731-746. <https://doi.org/10.54660/ijmrge.2021.2.6.731-746>
- [89] Sanni, J. O., Iwuanyanwu, U. A., Essien, M. A., Atima, M. E., & Attah, A. (2022). Adaptive Control Models for AI-Driven Marketing Automation in Financial Compliance Environments. *Shodhshauryam International Scientific Refereed Research*, 243. <https://doi.org/10.32628/shisrrj247133>
- [90] Fadero, S., Oyesiji, S. O., Nwakamma, S., & Jooda, D. (2018). IT Governance Frameworks (COBIT/ITIL) and Their Impact on Regulatory Compliance in Nigerian Insurance Firms. *Iconic Research and Engineering Journals*, 2(5), 478-496. <https://doi.org/10.64388/irev2i5-1722619>
- [91] Mbonu, I. S., Iwuanyanwu, U., Aliliele, C., & Uzoka, E. (2022). Advances in Cloud Identity and Access Governance Optimization in Large Scale AWS Enterprise Environments. *Shodhshauryam International Scientific Refereed Research*, 403. <https://doi.org/10.32628/shisrrj225490>
- [92] Stephanie Mbonu, I., Aliliele, C., Iwuanyanwu, U., & Uzoka, E. (2020). A Review of Identity and Access Management Integration Strategies in Hybrid and Multi Cloud Environments. *International Journal of Multidisciplinary Research and Growth Evaluation*, 1(5), 795-810. <https://doi.org/10.54660/ijmrge.2020.1.5.795-810>
- [93] Stephanie Mbonu, I., Iwuanyanwu, U., Aliliele, C., & Uzoka, E. (2020). Advances in Infrastructure as Code Governance for Secure Terraform Based Enterprise Cloud Deployments. *International Journal of Multidisciplinary Research and Growth Evaluation*, 1(5), 811-828. <https://doi.org/10.54660/ijmrge.2020.1.5.811-828>
- [94] Mbonu, I. S., Iwuanyanwu, U., Aliliele, C., & Uzoka, E. (2020). A Conceptual Framework for Agile Supply Chain Digital Transformation with Embedded IT Risk and ISO Compliance Controls. *Iconic Research and Engineering Journals*, 3(11), 566-593. <https://doi.org/10.64388/irev3i11-1714916>
- [95] Stephanie Mbonu, I., Iwuanyanwu, U., Aliliele, C., & Uzoka, E. (2021). A Review of VoIP Forensic Analytics Models for Financial Fraud Detection and Regulatory Compliance Monitoring. *International Journal of Multidisciplinary Research and Growth Evaluation*, 2(6), 711-730. <https://doi.org/10.54660/ijmrge.2021.2.6.711-730>
- [96] Ashore-Onisemo, F., & Unamma, U. M. L. (2022). Predictive Analytics in Anti-Corruption and Public Financial Oversight: Advances and Opportunities. *International Journal of Multidisciplinary Research and Growth Evaluation*, 3(6), 1163-1182. <https://doi.org/10.54660/ijmrge.2022.3.6.1163-1182>
- [97] Uchechi Mary-Linda Unamma, U. M. L. U., Ebehiremen Faith Iziduh, E. F. I., & Ifeanyichukwu Jeffrey Okwesa, I. J. O. (2021). Data-Mining Techniques for Financial Fraud Detection : A Review. *International Journal of Scientific Research in Science and Technology*, 760. <https://doi.org/10.32628/ijrst2152555>
- [98] Maryann Inimfon Atakpa, M. I. A., & Nyiawung Fobellah Abetoh, N. F. A. (2022). A Systematic Review of Machine Learning Advances in Financial Fraud Detection for

- Banking Systems. *International Journal of Scientific Research in Computer Science Engineering and Information Technology*, 771. <https://doi.org/10.32628/cseit23906220>
- [99] Abetoh, N. F., & Atakpa, M. I. (2022). Predictive Analytics for Fraud Detection in Donor-Funded Health Program Audits: A Conceptual Framework for Sub-Saharan African Contexts. *International Journal of Multidisciplinary Research and Growth Evaluation*, 3(6), 1036-1054. <https://doi.org/10.54660/ijmrge.2022.3.6.1036-1054>
- [100] Ifeoma Okoli, I. O., Ann Kinyanjui, A. K., Kofi Okyere Mensah, K. O. M., Priscilla Agboada, P. A., & Omotayo Pedro, O. P. (2022). Data-Driven Auditing in Payroll and Receivables: Predictive Models for Enhancing Accuracy in Financial Statement Validation. *International Journal of Scientific Research in Computer Science Engineering and Information Technology*, 851. <https://doi.org/10.32628/cseit23902204>
- [101] Sakyi, J. K., Nnabueze, S. B., Filani, O. M., Okojie, J. S., & Okereke, M. (2022). Customer Service Analytics as a Strategic Driver of Revenue Growth and Sustainable Business Competitiveness. *Journal of Frontiers in Multidisciplinary Research*, 3(2), 109-123. <https://doi.org/10.54660/ijfmr.2022.3.2.109-123>
- [102] Sakyi, J. K., Filani, O. M., Nnabueze, S. B., Okojie, J. S., & Ogedengbe, A. O. (2022). Developing KPI Frameworks to Enhance Accountability and Performance across Large-Scale Commercial Organizations. *Journal of Frontiers in Multidisciplinary Research*, 3(2), 81-93. <https://doi.org/10.54660/ijfmr.2022.3.2.81-93>
- [103] Iziduh, E. F., Ashore-Onisemo, F., & Unamma, U. M. L. (2020). Natural Language Processing for Unstructured Records in Finance and Public Services: A Review. *International Journal of Multidisciplinary Research and Growth Evaluation*, 1(5), 1215-1229. <https://doi.org/10.54660/ijmrge.2020.1.5.1215-1229>
- [104] Ejimah, A. A., & Aregbe, O. O. (2021). Impact of unstable power supply on the performance and durability of electronic devices in Lagos State, Nigeria. *Global Journal of Applied, Management and Social Sciences*, 23, 299-306.
- [105] Nwangwu, C. V., Duvbiama-Owasanoye, O. A., & Onwumere, A. (2018). A framework for operational integrity in high-risk drilling: Balancing productivity and safety compliance. *Iconic Research and Engineering Journals*, 2(5), 436-453. <https://www.irejournals.com/paper-details/1717207>
- [106] Nwangwu, C. V., Duvbiama-Owasanoye, O. A., & Onwumere, A. (2019). A strategic review of the industry-academia collaboration model for modernizing technical learning centers. *Iconic Research and Engineering Journals*, 3(6), 565-583. <https://www.irejournals.com/paper-details/1717208>
- [107] Akinsete, O. O., Nwangwu, C. V., & Isehunwa, S. O. (2020). An analytical model for characterising dynamic filtration of drilling fluid in vertical well. *Petroleum and Coal*, 62(1), 41-49. https://www.researchgate.net/publication/343049711_Article_Open_Access_An_Analytical_Model_for_Characterising_Dynamic_Filtration_of_Drilling_Fluid_in_Vertical_Well
- [108] Nwangwu, C. V., Duvbiama-Owasanoye, O. A., & Onwumere, A. (2020). High-fidelity reservoir forecasting: A technical review of multimodal modeling for long-term production estimates. *Iconic Research and Engineering Journals*, 4(5), 390-409. <https://www.irejournals.com/paper-details/1717209>

- [109] Nwangwu, C. V., Duvbiama-Owasanoye, O. A., & Akano, O. A. (2021). Aligning capital budgeting with field development execution in the upstream petroleum sector. Gyanshauryam, International Scientific Refereed Research Journal, 4(4), 259-282. <https://gisrrj.com/paper/GISRRJ214424.pdf>
- [110] Nwangwu, C. V., Duvbiama-Owasanoye, O. A., & Akano, O. A. (2022). Resource discovery in mature asset portfolios: A review of longitudinal data trends for subsurface value extraction. Gyanshauryam, International Scientific Refereed Research Journal, 5(3), 503-524. <https://gisrrj.com/GISRRJ225390>
- [111] Adebajo, A. K., Duruzor, E. D., Oyakhire, B. O., Onasanya, B. A., Nwangwu, C. V., Michael, E. O., Akpedeye, D. O., Afere, O. O., & Ayalogu, D. I. (2023). Oil & gas production capacity digitalization: A strategic approach to production deferment management. SPE Nigeria Annual International Conference and Exhibition, SPE-217224-MS. <https://doi.org/10.2118/217224-MS>
- [112] Nwangwu, C. V., Duvbiama-Owasanoye, O. A., & Akano, O. A. (2023). Technical due diligence and supply chain integration in energy M&A: A review of large-scale asset acquisitions. Shodhshauryam, International Scientific Refereed Research Journal, 6(2), 502-528. <https://shisrrj.com/SHISRRJ122606>