SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Directional Asymmetry in banking Journal Entries: A Machine Learning Perspective on Audit Risk Prioritization

Awni Rawashdeh[1], Mohammad Kamal Kamel Afaneh[2], Badi Alrawashdeh[3], Mohammad Motasem Alrfai[4], Nahed Habis Alrawashedh[5]

[1]Department of AI in Auditing and Accounting, Faculty of Business, Applied Science Private University (ASU), Jordan.
[2]Department of Finance, Collage of Business, Imam Mohammad Ibn Saud Islamic University (IMSIU).
[3]Department of Accounting, Faculty of Administrative and Financial Science Irbid National University (INU), Irbid, Jordan.
E-mail: b.alrwashdeh@inu.edu.jo
[4]Department of Accounting, Faculty of Administrative and Financial Science Irbid National University (INU), Irbid, Jordan.
[5]Department of Accounting , Faculty of Business, Islamic University of Minnesota.US.

Abstract
Banking journal entry testing needs to provide auditors the ability to evaluate large sets of transactions and assess which transactions should receive the most focused attention. The purpose of this paper was to determine if a journal-entry risk-ranking model (AARR-XGBoost), based on a directional asymmetry-aware representation of entry features, could produce better risk rankings than traditional methods. Thirty individual simulated audit environments were created, including 240 companies with varying characteristics, 231,773 reference-year documents, and 231,986 target-year documents. The features used to represent account roles, account-pair relationships, creator paths, timing patterns, and transaction amounts were all extracted from the historical ledger records of each company. With a limited review budget of 3%, AARR-XGBoost captured 24.83% of the scenario-positive documents (95% CI [20.79%, 29.07%]), outperforming conventional XGBoost, Isolation Forest, transparent practice-informed rules, and random selection. When feature-set analyses were conducted to identify the specific components that contributed to AARR-XGBoost's performance, directional role violation and account-pair surprisal were the largest contributors to ranking performance. Additionally, an independent evaluation using a publicly accessible ERP fraud dataset provided descriptive evidence of how well AARR-XGBoost performed relative to conventional XGBoost in a different environment. Specifically, AARR-XGBoost identified 30 of 32 labelled positive documents, while conventional XGBoost identified 24. Collectively, the results demonstrate that including directional asymmetry within the representation improved journal-entry risk-ranking performance in controlled simulations when audit review capacity was constrained. The results also support the development of a reproducible framework for assessing journal-entry analytics and suggest future research using either operational data or de-identified ledger data.

Keywords: audit analytics; journal entry testing; directional asymmetry; controlled simulation; risk ranking; synthetic ledger; XGBoost; audit sampling; external evaluation; public ERP data

1. Introduction

A number of studies and auditing standards support the use of sampling when less than the full transaction population is tested. Therefore, prior to undertaking audit sampling, there should be a reasonable and defensible rationale for limiting sample size (Public Company Accounting Oversight Board, 2026a, 2026b). The rationale will depend upon factors such as the nature of the assertions being tested, the type of substantive procedures to be performed, and the level of assessed risk. When performing substantive testing of journal entries, it may be desirable to examine a broad population of entries. Entries and other adjustments may be related to the risk of management override, and PCAOB guidance emphasizes that both the completeness of the underlying population and examination of entries for evidence of possible material misstatement require attention (Public Company Accounting Oversight Board, 2025, 2026b). ISA 240 (Revised), like PCAOB guidance, also addresses the auditor's response to fraud risk within a financial statement audit (International Auditing and Assurance Standards Board, 2025).

Prior research has identified ways in which auditors can use audit data analytics to help prioritize complete populations for further review. However, even with the assistance provided by audit data analytics, the reliability and relevance of the underlying data, the completeness of the population, and the continued importance of professional judgment remain fundamental concerns (Brown-Liburd et al., 2015; Cao et al., 2015; Earley, 2015; International Auditing and Assurance Standards Board Data Analytics Working Group, 2016; Jans et al., 2014; Yoon et al., 2015). Previous research has discussed how large-scale analytics can complement audit evidence and support process-level analyses, whereas more recent research has examined formalized audit tasks, supervised-learning research designs,

and organizational influences on analytics adoption (Almufadda & Almezeini, 2022; Appelbaum et al., 2017; Brown-Liburd et al., 2015; Cao et al., 2015; Earley, 2015; Gepp et al., 2018; Issa et al., 2016; Jans et al., 2014; Kokina & Davenport, 2017; Krahel & Titera, 2015; Krieger et al., 2021; Law & Shen, 2025; Manita et al., 2020; Yoon et al., 2015). Across these studies, analytics are treated primarily as tools for helping select items for follow-up and supporting audit evidence.

Because journal entries may contain unusual structural patterns, they provide a logical context for employing analytical techniques for screening purposes. Double-entry bookkeeping not only preserves equal monetary values within each valid journal document, but recurring business processes also produce conditional directional patterns in which specific debit-credit combinations recur, individual journal-document creators follow established posting paths, time-based patterns emerge from closing procedures, and document amounts occupy firm-specific distributions. Similar arguments have been made in accounting graph topology and general-ledger anomaly detection, where journal entries are viewed as having relational and directional structure beyond what can be captured through global rarity measures (Gronewald et al., 2024; Guo et al., 2022; Hemati et al., 2022; Huang et al., 2026; Mashiko et al., 2025).

This study views directional asymmetry as a structural aspect of accounting data, that is, as a symmetry/asymmetry characteristic of accounting data. Directional asymmetry is distinct from amount imbalance because each validly generated document remains double-entry balanced. It is also distinct from generic anomaly-detection approaches. The appropriate reference point is not simply a document's distance from a global center, but whether its ordered account roles, creator path, timing, and amount behavior deviate collectively from the directional patterns observed during the company's completed period (Chandola et al., 2009; Guo et al., 2022; Hemati et al., 2022; West & Bhattacharya, 2016). When directional asymmetry is modeled explicitly, this permits enhanced predictive capability in machine-learning models used to rank high-risk journal entries under limited review budgets.

Comparative evaluations of analytical methods often rely on simulated environments when field data are unavailable because of confidentiality restrictions or incomplete outcome labels. To be effective, simulations should have an identifiable data-generating mechanism, meaningful overlap between atypical and typical entries, and separation between organizational settings used for development and evaluation. This paper assesses AARR in a controlled simulation framework combining legitimate atypical entries, operating-family holdout, feature-set analysis, and cross-mechanism evaluation. The results illustrate comparative ranking performance under defined simulation conditions and provide a foundation for future empirical evaluation (Altman et al., 2023; Becerra-Suarez et al., 2025; Belgodere et al., 2024; Morris et al., 2019).

2. Related Work and Study Positioning

2.1. Journal Entry Testing and Audit Data Analytics

Auditing research has evolved beyond questions about whether data are available for use as evidence or whether the data provide complementary evidence. The focus has shifted to more specific issues such as the degree of formalization required for audit tasks, what evidence supports the auditor's assessment, and how auditors interact with results provided by algorithms.

Early foundational studies described how auditing can be enhanced through the use of analytics and how analytical tools can alter the manner in which audit evidence is evaluated (Appelbaum et al., 2017; Brown-Liburd et al., 2015; Cao et al., 2015; Earley, 2015; Gepp et al., 2018; Jans et al., 2014; Yoon et al., 2015). As subsequent studies identified audit analytics as a distinct empirical area, methodologies developed around labeling, feature identification, sampling, and evaluation targets (Almufadda & Almezeini, 2022; Issa et al., 2016; Kokina & Davenport, 2017; Krahel & Titera, 2015; Law & Shen, 2025; Ruhnke, 2023). These issues are directly related to journal entry testing because determining a class label is not the only task; it also involves determining which subset of entries should be examined further when a complete examination of all entries is impractical.

Additionally, the journal-entry literature illustrates the diversity of methods that can be used for analysis. Studies concerning financial-statement fraud have employed statistical and machine-learning approaches to rare and heterogeneous outcomes (J. Perols, 2011; J. L. Perols et al., 2017; Ravisankar et al., 2011; West & Bhattacharya, 2016). On the other hand, studies examining anomalies in general-ledger data have employed unsupervised, deep-learning, hybrid, and graph-based techniques that use knowledge of accounting structure (Gronewald et al., 2024; Guo et al., 2022; Hemati et al., 2022; Huang et al., 2026; Mashiko et al., 2025). There are differences among these approaches, including levels of data access, label specificity, and objectives. Some models seek anomalous entries, others attempt to identify latent representations, and others model graph relationships. A high anomaly score is not equivalent to fraud.

2.2. From Generic Anomaly Detection to Directional Asymmetry

Anomaly detection is generally based on identifying entries that are unusual relative to an observed population, for example, entries that are rare, isolated, or difficult to reconstruct (Chandola et al., 2009; Liu et al., 2008). This type of

detection provides useful baseline performance in auditing because confirmed labels are scarce and audit populations are highly imbalanced. However, a potential drawback of anomaly-detection techniques is that global or latent-space anomaly scores may not capture the accounting direction that gives an entry its substantive meaning. For example, while a particular entry may be an outlier with respect to an overall distribution, it is more informative when its debit-credit roles, creator-pair history, posting time, and amount jointly conflict with the firm's completed-period patterns.
AARR is proposed as a conditional structural representation for journal-entry risk ranking. In addition to incorporating conventional audit indicators, AARR incorporates six historical departures: account-pair surprisal, amount asymmetry conditional on the account pair, creator-pair surprisal, temporal surprisal conditional on the business process, directional role violation, and a bounded concordance count. As such, AARR can potentially leverage advances in graph-based representations of accounting data and continual learning to treat accounting data as both structurally dependent and temporally dynamic rather than simply as independent rows of data (Guo et al., 2022; Hemati et al., 2022; Huang et al., 2026; Mashiko et al., 2025). Within this study, AARR makes these structures measurable in simulated audits alongside more complex relational models and substantive audit evidence.

2.3. Synthetic Data and Simulation Studies

Synthetic data allow researchers to evaluate models under audit conditions that are rare, confidential, or deliberately controlled while establishing the outcome process, information boundaries, comparison methods, and performance criteria before model behavior is evaluated (Morris et al., 2019). This type of data is useful when researchers want to establish how outcomes are generated before comparing model behavior against those outcomes. Recent research on synthetic data has also identified four important considerations: realism, privacy, controllability, and task validity (Altman et al., 2023; Becerra-Suarez et al., 2025; Belgodere et al., 2024).
In addition to using a single study-generated ledger as a means of evaluating analytical representations, there are publicly available labelled ERP event data generated from occupational-fraud scenarios. These datasets can provide another basis for evaluation beyond a single study. Tritscher et al. developed a publicly available ERP dataset that includes fraudulent activity and legitimate operational activity within a simulated virtual ERP environment, with documentation of normal business processes and fraud scenarios (Tritscher et al., 2022).
A controlled simulation of journal entries should clearly identify the relevant information provided to each model, include examples of legitimate atypical documents, use different operating environments during development and evaluation, and measure how performance changes based on scenario structure. The current study addresses these requirements through exclusion of administrative attributes, a cross-family holdout method, univariate association diagnostics, cross-mechanism evaluation, and feature-set analysis. The results represent comparative ranking performance under specified conditions and support subsequent testing using operational ledgers.

2.4. Research Gap and Study Positioning

There are many existing papers on audit analytics that provide significant insights into the use of data, process mining, deep anomaly models, graph structures, and organizational adoption of artificial intelligence (De Santis & D'Onza, 2021; Fedyk et al., 2022; Gronewald et al., 2024; Guo et al., 2022; Hemati et al., 2022; Huang et al., 2026; Jans et al., 2014; Kokina & Davenport, 2017; Krahel & Titera, 2015; Krieger et al., 2021; Law & Shen, 2025; Manita et al., 2020; Mashiko et al., 2025; Ruhnke, 2023; Samiolo et al., 2024). However, there is less knowledge about how different forms of journal-entry representation perform in controlled studies using organizational holdout, scenario-positive outcomes, fixed review budgets, and transparent information boundaries. This paper addresses this gap by comparing directional asymmetry-aware ranking with conventional and unsupervised approaches and rule-based alternatives in repeated ledger simulations.

2.5. Research Objective and Contribution

The primary research question is: Does an asymmetry-aware risk-ranking representation improve the capture of simulated scenario-positive journal documents within a fixed audit review budget, compared with conventional feature models, unsupervised anomaly detection, transparent rule-based systems, and random selection?
Six contributions are made in this research. First, this work examines the process of identifying journal entries as a constrained ranking problem instead of a generic classification problem. Second, it represents directional relationships among selected accounts using information from completed reference-period documents. Third, it includes legitimate atypical document examples with features similar to those observed in scenario-positive documents. Fourth, it conducts the assessment in 30 independent simulation environments, applies operating-family holdout, performs cross-mechanism evaluations, and examines different feature sets. Fifth, the research documents the data generator, model-exclusion requirements, and uncertainty intervals so that the study can be repeated and extended. Sixth, it tests a structural representation on an independently developed publicly available ERP fraud dataset using run-holdout

evaluation. The study therefore provides evidence on risk prioritization in both controlled and public benchmark settings for selecting items for substantive audit procedures.

3. Materials and Methods

3.1. Controlled Simulation Design and Data Construction

The study followed the simulation-study approach, in which the data-generating mechanism, estimand, methods, and performance measures should be explicitly defined before the results are interpreted (Morris et al., 2019). Thirty independent simulation environments were created. Each environment included eight synthetic companies drawn from four operating families: manufacturing, distribution, services, and hybrid operations. Each company had its own chart of accounts, document-code mapping, user roster, process mix, amount distribution, manual-posting propensity, close-process pattern, and volume profile. Two companies from one operating family were reserved for testing, while six companies from the remaining three operating families were used for development.

For each company, 2023 was designated as the completed reference year and 2024 as the target year. All features used to rank target-year documents were created using only reference-year records from the respective company. Therefore, the model did not use target-year outcomes, labels, or future-period information when constructing the ranking features. Grouped cross-validation was used for hyperparameter selection over development companies, separating model selection from evaluation on the held-out operating family (Varma & Simon, 2006).

Across all target documents, there were 3,236 scenario-positive documents out of 231,986, equivalent to 1.39%. The performance metrics described below were calculated by averaging at the environment level, where the mean held-out prevalence was 1.30%. This environment-level prevalence was used to calculate the reported mean enrichment ratio. Table 1 details the simulation corpus and evaluation framework.

Table 1. Simulation corpus and evaluation structure.

Dataset component	Value
Independent simulation environments	30
Companies	240
Operating families	4
Historical documents, 2023	231,773
Target documents, 2024	231,986
Scenario-positive target documents	3,236
Scenario-positive prevalence	1.39%
Held-out family rotations	4

The target variable, labelled scenario-positive, represents a simulated journal-document scenario that includes partial departures in directional account roles, account-pair history, creator-path history, timing, and amount behavior. The variable is not a legal or professional determination of fraud. There were four positive mechanism families: directional path, timing override, amount-timing, and creator path. A candidate mechanism was assigned before the Bernoulli outcome was generated. Hidden approval gaps, documentation weakness, legitimate supporting evidence, company-level shifts, and random disturbance also influenced the final probability. The unobserved support variable was designed to produce controls with risk-like observed attributes that nevertheless remained non-positive. Table 2 outlines the distribution and construction of scenario-positive documents.

Table 2. Distribution and construction logic of simulated scenario-positive target documents.

Scenario mechanism	Documents	Share	Structural description
Directional path	989	30.56%	Jointly unusual account-pair, creator-pair, and debit-credit role direction.
Creator path	801	24.75%	Rare creator-path and account-pair combination with delayed posting context.
Timing override	730	22.56%	Manual and post-close timing deviation combined with temporal irregularity.
Amount-timing	716	22.13%	Amount deviation combined with period-end and rounded-amount context.

Raw document type, business process, account identifiers, company code, creator identifiers, scenario type, the legitimate-atypical-entry indicator, and generator-only variables were excluded from all predictive models. Each document type and business process observed among scenario-positive documents was also present in the comparison population in every environment.

The average maximum single-feature ROC-AUC across environments was 0.739 (95% CI 0.713 to 0.765; range 0.592 to 0.871). This shows useful univariate information without a feature that universally separates the outcome. The complete range is reported in the supplementary material. Legitimate atypical entries represented 17.98% of non-positive documents and 10.03% of scenario-positive documents, showing overlap in risk-like observable patterns.

3.2. Directional Asymmetry-Aware Representation

The AARR model includes seven conventional audit-oriented variables and six measures of asymmetry. Conventional variables include log amount, manual posting, post-close posting, posting delay, weekend posting, period-end posting, and rounded amount. The asymmetry variables include account-pair surprisal, amount asymmetry, creator-pair surprisal, temporal surprisal, directional role violation, and asymmetry concordance. All probability and frequency terms below are based on the completed reference period only.

For company c and an ordered debit-credit account pair p, account-pair surprisal is defined as:

$S\_pair = -\log[(n_{c,p} + 1) / (N_c + K_c + 1)]$

where $n_{c,p}$ is the historical frequency of pair p, N_c is the number of historical documents, and K_c is the number of distinct historical pairs. The additive constants provide Laplace smoothing for unseen but possible pairs.

Amount asymmetry is the absolute standardised departure of the log amount from the historical pair-specific mean:

$A\_amt = |\log(1 + amount) - \mu_{c,p}| / \max(\sigma_{c,p}, 0.35)$

Creator-pair surprisal and temporal surprisal are defined as:

$S\_creator = -\log[(n_{c,u,p} + 1) / (N_{c,u} + K_c + 1)]$

$S\_time = -\log[(n_{c,b,t} + 1) / (N_c / 5 + 5)]$

where u is the creator, b is the business process, and t is one of five relative posting-date buckets. Directional role violation uses debit and credit tendencies estimated for each account a:

$r_a = \log[(n_{debit,a} + 2) / (n_{credit,a} + 2)]; \quad V\_dir = \max(0, -(r_{debit} - r_{credit}))$

Finally, asymmetry concordance is a bounded count of conditions that exceed pre-specified thresholds: pair surprisal > log(22), creator-pair surprisal > log(14), amount asymmetry > 1.35, temporal surprisal > 1.45, and directional role violation > 0.85. The count is included as one feature, not treated as a fraud rule or an audit conclusion.

3.3. Models, Audit Baselines, and Feature-Set Analysis

The primary model was AARR-XGBoost. Three-fold grouped cross-validation was run within each development dataset. Two compact candidate configurations were tested: (i) 110 trees, maximum depth 2, learning rate 0.06, and minimum child weight 4; and (ii) 130 trees, maximum depth 3, learning rate 0.05, and minimum child weight 5. Both configurations used histogram tree construction, subsample 0.88, column subsample 0.88, gamma 0.04, L2 regularization 2.0, and a positive-class weight equal to the negative-to-positive ratio.

XGBoost, random forest, and Isolation Forest were chosen because they are established supervised and unsupervised methods for tabular, imbalanced anomaly problems (Breiman, 2001; Chen & Guestrin, 2016; Krawczyk, 2016; Liu et al., 2008). Comparator models included conventional XGBoost, logistic regression, and random forest trained using conventional variables only; AARR-random forest trained using the full representation; Isolation Forest trained on the unlabeled target population; a transparent rule-based approach; and expected random sampling. The rule-based approach gave one point each for manual posting, post-close posting, period-end posting, amount asymmetry greater than 1.75, delay of four or more days, and temporal surprisal greater than 1.55. This is a simple transparent baseline and has no connection to proprietary audit-firm methodologies.

Feature-set analysis evaluated the impact of individual asymmetry variables compared with the full representation. The same parsimonious XGBoost configuration, 110 trees, maximum depth 2, learning rate 0.06, and minimum child weight 4, was used across all feature sets to isolate differences due to feature content rather than hyperparameters. The comparisons were made among conventional features only, conventional features plus each individual asymmetry feature, the full representation without concordance, and the full AARR representation.

3.4. Evaluation Protocol, Statistical Analysis, and Reproducibility

Capture rate at a 3% review budget is the proportion of all scenario-positive documents contained in the highest-ranked 3% of target documents. This captures the operational question: What does an audit team find when detailed-testing capacity is limited? Secondary metrics are capture rates at 1% and 5%, precision at 3%, precision-recall area under the curve (PR-AUC), receiver operating characteristic area under the curve (ROC-AUC), and the review workload required to capture 50% and 80% of scenario-positive documents. Precision-recall metrics are used because the target event is rare and ROC-AUC alone can be incomplete under class imbalance (Davis & Goadrich, 2006; Krawczyk, 2016; Saito & Rehmsmeier, 2015).

The unit of statistical inference is the independent simulated environment. Mean metrics and 95% confidence intervals were estimated using nonparametric bootstrap resampling of the 30 simulated environments (Efron & Tibshirani, 1993; Morris et al., 2019). Pairwise comparison of capture rate at 3% across environments was conducted using a two-

sided Wilcoxon signed-rank test (Demšar, 2006). Effect-size and confidence-interval estimates are treated as primary evidence, while p-values are supplementary. A four-way cross-mechanism analysis omits each scenario mechanism in turn from development positives while retaining all controls and evaluates the omitted mechanism against controls in the held-out operating family. This produces 120 mechanism-environment combinations. One cell contained zero positive documents for the mechanism of interest and was excluded only from mechanism-specific metric averages, leaving 119 evaluable cells. Under this design, AARR can perform better than, similarly to, or worse than comparator models, and its benefit can be concentrated in one component or persist when a mechanism is left out of development positives. Feature-set and cross-mechanism analyses examine these possibilities and provide incremental rank-ordering information beyond conventional audit indicators within a controlled simulation study (Altman et al., 2023; Becerra-Suarez et al., 2025; Belgodere et al., 2024; Morris et al., 2019).

Reproducibility materials include environment seed values, company heterogeneity profiles, document generator code, feature-construction routines, hyperparameter candidates, the model-input exclusion manifest, environment-level outcomes, bootstrap procedures, and figure-generation scripts. Scenario labels, generator variables, company and account identifiers, document types, business processes, and creator identifiers that would not be available to a real-time ranking tool are excluded from training. The source code reproduces the reported results and simulation assumptions.

3.5. Independent Public-ERP Evaluation

To assess whether the structural representation could be evaluated beyond the controlled ledgers used in the main study, an independent out-of-source assessment was conducted using the publicly released ERP fraud data from Tritscher et al. (Tritscher et al., 2022). The database was independently developed and contains fraud-annotated ERP runs generated in a simulated make-to-stock environment. Its use therefore represents an external public benchmark rather than field validation using a client general ledger.

The three fraud-annotated datasets contained 34,045 accounting documents, including 32 positive documents identified because at least one underlying line item had a source label other than NonFraud. Only preceding documents within each run were included when calculating structural scores. Because preparer identifiers, a manual-posting flag, and an accounting-period close field were unavailable, creator-pair surprisal was replaced with an explicitly labelled counterparty-path surprisal based on vendor or material identifiers and the primary debit-credit general-ledger path. Raw document numbers, source labels, transaction categories, account codes, vendor identifiers, and material identifiers were excluded from direct model inputs.

Leave-one-fraud-run-out testing was used for evaluation; the model was fitted on two public fraud runs and tested on the remaining run without held-out-run tuning. As with the main study, the primary measure remained capture at a 3% review budget. Because only three fraud runs and 32 positive documents were available, the assessments are reported descriptively by run and include means, ranges, and pooled capture rather than formal hypothesis testing.

Table 3. External public-ERP evaluation design and information boundary.

Component	External public-ERP evaluation
Data source	Public ERP fraud data (Tritscher et al., 2022)
Unit of analysis	Accounting document (Belegnummer)
Fraud-annotated runs	3
Documents / positive documents	34,045 / 32
Feature history	Preceding documents in the same run only
Test protocol	Leave-one-fraud-run-out; no test-run tuning
Unavailable original fields	Preparer identity; manual posting; period-close indicator

4. Results

4.1. Main Ranking Performance and Pairwise Comparisons

AARR-XGBoost demonstrated the greatest overall average capture at each stated review budget. At a 3% review budget, AARR-XGBoost captured 24.83% of scenario-positive documents (95% CI 20.79% to 29.07%) and achieved an average precision of 10.64%. The mean prevalence of scenario-positive documents within the held-out environments was 1.30%, whereas the pooled corpus prevalence was 1.39%. Because the enrichment calculation uses the environment-level mean prevalence, the 10.64% precision represents approximately 8.2-fold enrichment over random selection at the same review budget.

Table 4 presents the principal ranking results, while Figure 1 illustrates capture at the primary 3% review budget. The operational distinction among the methods is clear because the amount of data being reviewed is held constant. Under the same 3% review-capacity constraint, AARR-XGBoost captured more than six times the share captured by

conventional XGBoost and more than twice the share captured by Isolation Forest. AARR-random forest also showed strong performance, with a capture rate of 21.85%, indicating that much of the improvement is associated with the asymmetry-aware representation rather than being unique to tree boosting.

By contrast, conventional XGBoost, logistic regression, the transparent rule score, and expected random selection remained in the low single-digit range. The workload measure shows the same relative ordering: AARR-XGBoost required a mean review workload of 19.7% to capture 50% of the scenario-positive documents, whereas conventional XGBoost required 44.7% and random selection required 50.0%.

Therefore, the primary result reflects both a higher concentration of scenario-positive documents at the top of the ranking and a material reduction in the proportion of documents that need to be reviewed to reach a fixed capture target.

Table 4. Mean performance over 30 independent environments. Capture at a fixed 3% review budget is the primary metric. The table reports environment-level means and bootstrap confidence intervals.

Method	PR-AUC	ROC-AUC	Capture @ 3%	95% CI	Review workload for 50% capture
AARR-XGBoost	0.080	0.694	24.83%	20.79% to 29.07%	19.7%
AARR-Random forest	0.079	0.685	21.85%	18.12% to 25.82%	20.7%
Isolation Forest	0.032	0.632	9.45%	7.07% to 11.87%	30.8%
Conventional XGBoost	0.018	0.523	3.71%	2.35% to 5.12%	44.7%
Logistic regression	0.023	0.548	4.19%	2.76% to 5.67%	43.0%
Practice-informed rules	0.016	0.509	2.49%	1.22% to 4.02%	47.0%
Random sampling	0.013	0.500	3.00%	3.00% to 3.00%	50.0%

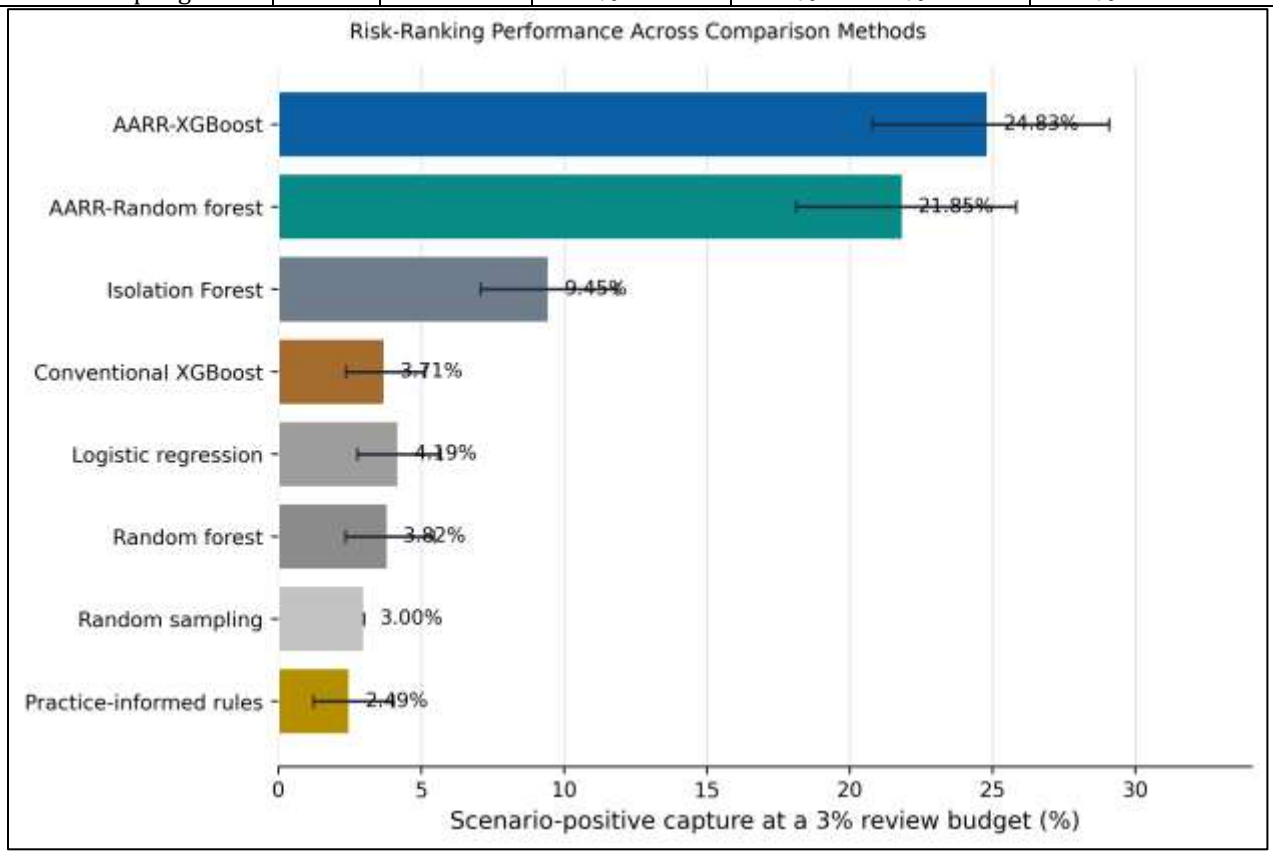

Figure 1. Scenario-positive capture at a fixed 3% review budget across all comparison methods. Error bars represent 95% bootstrap confidence intervals over independent simulation environments.

The results show that AARR-XGBoost performed better across the 30 simulation environments than conventional XGBoost at the 3% review budget. AARR-XGBoost outperformed conventional XGBoost by 21.12 percentage points (95% CI 17.08 to 25.41 percentage points; two-sided Wilcoxon $p < .001$). Likewise, AARR-XGBoost outperformed Isolation Forest by 15.38 percentage points (95% CI 11.50 to 19.15; $p < .001$) and AARR-random forest by 2.97

percentage points (95% CI 0.88 to 5.17; p = .011). The comparisons are presented in Table 5. These results indicate variation in performance across methods while preserving the same environment-level unit of comparison.

Table 5. Pairwise comparisons of capture rate at a 3% review budget. Positive differences favour AARR-XGBoost. Two-sided Wilcoxon signed-rank tests are reported as supplementary evidence.

Comparison	Mean difference	95% CI	Two-sided p
AARR-XGBoost vs Conventional XGBoost	21.12 pp	17.08 to 25.41 pp	<0.001
AARR-XGBoost vs AARR-Random forest	2.97 pp	0.88 to 5.17 pp	0.011
AARR-XGBoost vs Random forest	21.01 pp	17.08 to 25.26 pp	<0.001
AARR-XGBoost vs Logistic regression	20.63 pp	16.50 to 25.00 pp	<0.001
AARR-XGBoost vs Isolation Forest	15.38 pp	11.50 to 19.15 pp	<0.001
AARR-XGBoost vs Practice-informed rules	22.34 pp	18.11 to 26.41 pp	<0.001
AARR-XGBoost vs Random sampling	21.83 pp	17.70 to 26.11 pp	<0.001

4.2. Representation-Component Analysis

Feature-set analysis employed a predefined parsimonious XGBoost configuration to separate the contribution of feature content. Therefore, the full-AARR estimate of 25.36% differs slightly from the 24.83% reported in Table 4, where the primary model used a configuration selected separately within each development dataset. The two most informative additions to conventional features were directional role violation and account-pair surprisal. The representation could not be reduced to the concordance count: AARR without concordance captured 24.15% of scenario-positive documents, while conventional features alone captured 4.71%. Full AARR captured 25.36%, but the additional value of concordance was not statistically significant at the conventional 5% level (p = .087). Thus, the ranking increase appears to come from the combined structural representation rather than from a single count-based method. The results are presented in Table 6, and Figure 2 illustrates the ablation pattern.

The component analysis also shows that the asymmetry variables are not interchangeable. Adding directional role violation to the conventional feature set raised capture to 21.44%, while adding account-pair surprisal raised capture to 10.93%. Creator-pair surprisal and asymmetry concordance produced smaller improvements, whereas amount asymmetry and temporal surprisal did not improve capture when added individually. Their limited standalone performance does not mean that they are irrelevant in the full representation, because tree-based models can use conditional interactions among features. The full 13-feature representation achieved 25.36% capture, while the 12-feature version without concordance achieved 24.15%. The relatively small difference supports the interpretation that performance is largely due to the underlying structural measures rather than a simple count-based representation.

Table 6. Feature-set analysis under a common parsimonious XGBoost configuration. Hyperparameters are fixed across specifications to isolate changes in feature content.

Feature set	Features	Capture @ 3%	95% CI	PR-AUC
Conventional features	7	4.71%	3.17% to 6.39%	0.019
Conventional + account-pair surprisal	8	10.93%	8.53% to 13.40%	0.042
Conventional + amount asymmetry	8	3.29%	1.85% to 4.92%	0.017
Conventional + creator-pair surprisal	8	6.14%	4.16% to 8.24%	0.027
Conventional + temporal surprisal	8	4.16%	2.87% to 5.46%	0.019
Conventional + directional role violation	8	21.44%	18.00% to 25.08%	0.071
Conventional + asymmetry concordance	8	6.94%	4.97% to 9.08%	0.029
AARR without concordance	12	24.15%	19.86% to 28.49%	0.082
Full AARR	13	25.36%	21.05% to 29.78%	0.082

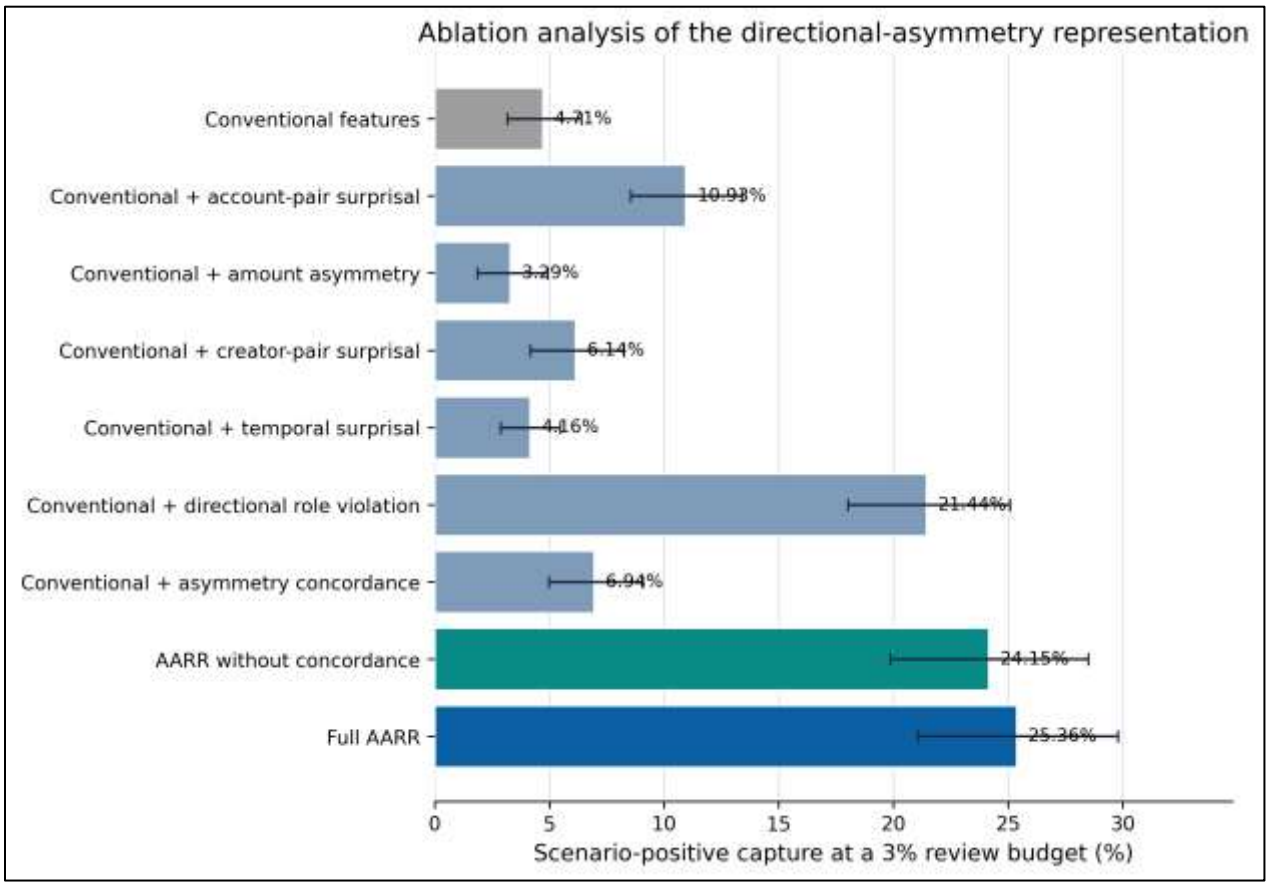

Figure 2. Feature-set ablation at a fixed 3% review budget. Error bars represent 95% bootstrap confidence intervals over independent simulation environments.

4.3. Feature Overlap and Cross-Mechanism Performance

No raw administrative category or identity field was used as input to any model. In addition, every scenario-positive document type and business process appeared in the comparison population. The maximum single-feature diagnostic had a mean ROC-AUC of 0.739. Thus, there is meaningful univariate structural information, but the structural features do not universally separate the outcome. Additional information concerning the complete distribution, including upper-tail values, is available in the supplementary material.

In cross-mechanism evaluation, AARR-XGBoost captured approximately 24.00% of scenario-positive documents associated with an omitted mechanism at a 3% review budget across 119 evaluable cells (95% CI 20.46% to 27.60%). Performance varied by mechanism, ranging from 22.24% for amount-timing to 25.96% for creator-path. This analysis examines how the representation performs across four mechanism families when the corresponding mechanism family is absent from development positives. Table 7 presents the cross-mechanism results.

Cross-mechanism performance remained comparatively stable across the omitted mechanism families. Directional-path capture was 25.28%, timing-override capture was 22.57%, amount-timing capture was 22.24%, and creator-path capture was 25.96%. The difference between the lowest and highest mechanism averages was therefore modest relative to the overall improvement over conventional models in the main evaluation. This is important for interpretation because, in each case, the positive mechanism was absent from development data. The model therefore ranked documents using structural properties shared across the representation, which is consistent with AARR's role as a conditional description of posting behavior rather than a set of mechanism-specific rules.

Table 7. Cross-mechanism evaluation. One cell with zero positive documents for the relevant mechanism was not evaluable for capture; 119 evaluable cells remain.

Held-out mechanism	**Evaluable cells**	**Mean positives**	**Capture @ 3%**	**95% CI**	**ROC-AUC**
directional_path	29	7.9	25.28%	18.08% to 32.76%	0.731
timing_override	30	5.7	22.57%	16.15% to 29.29%	0.639

amount_timing	30	6.3	22.24%	15.58% to 29.30%	0.691
creator_path	30	5.9	25.96%	18.70% to 33.37%	0.704
All mechanisms	119	6.4	24.00%	20.46% to 27.60%	0.691

4.4. Independent Public-ERP Evaluation

Public ERP results are reported descriptively across the three held-out runs because only 32 positive documents were available. With a 3% review budget, AARR-XGBoost captured 30 of the 32 labelled positive documents (93.75% pooled capture), compared with 24 of 32 (75.00%) for conventional XGBoost. Isolation Forest also captured 30 of 32 positives (93.75%). This shows that the directional asymmetry-aware representation can be applied across an independently developed ERP schema and labeling process while improving on the conventional XGBoost baseline in this benchmark. It does not demonstrate transfer of a model trained on the controlled simulation data because the public benchmark used leave-one-fraud-run-out fitting within the public dataset. The strong Isolation Forest result also suggests that the public benchmark contains substantial anomaly information.

The external benchmark shows a different comparative relationship from the controlled simulations. Mean AARR-XGBoost capture across held-out public runs was 92.06%, with a range from 83.33% to 100.00%, while conventional XGBoost averaged 73.41%. Logistic regression also performed strongly at 86.90%, and Isolation Forest had the highest mean capture per run at 95.24%. Although AARR-XGBoost and Isolation Forest had different run means, both captured 30 of 32 positives in pooled results. These results should be interpreted descriptively because the public benchmark contains only three annotated fraud runs. The adapted representation nevertheless performed better than conventional supervised XGBoost under a different account structure, transaction schema, set of available fields, and labeling process, while the Isolation Forest result shows that the asymmetry-aware supervised representation does not necessarily outperform a strong unsupervised anomaly signal in every setting.

Table 8. External run-held-out performance on independently developed public ERP fraud data. Scores are descriptive; formal tests are not reported because only three public runs were available.

Method	Mean capture @3%	Range across runs	Pooled capture	PR-AUC	ROC-AUC
AARR-XGBoost	92.06%	83.33% to 100.00%	30/32 (93.75%)	0.200	0.990
Conventional XGBoost	73.41%	66.67% to 78.57%	24/32 (75.00%)	0.132	0.852
Logistic regression	86.90%	75.00% to 100.00%	27/32 (84.38%)	0.256	0.914
Isolation Forest	95.24%	85.71% to 100.00%	30/32 (93.75%)	0.180	0.991
Practice-informed rules	76.19%	50.00% to 100.00%	23/32 (71.88%)	0.137	0.910
Expected random selection	3.00%	3.00%	Expected 3.00%	0.001	0.500

5. Discussion and Conclusion

5.1. Interpretation and Scope of Findings

When audit review capacity was constrained, directional asymmetry enhanced the prioritization of scenario-positive documents across repeated environments. With a 3% review budget, AARR-XGBoost captured around one quarter of the positive documents, while conventional feature models, transparent rule sets, and random selection captured roughly 2% to 4%. The finding also supports audit-analytics research that views analytical outputs as evidence-selection aids rather than replacements for audit evidence or professional judgment (Almufadda & Almezeini, 2022; Appelbaum et al., 2017; Brown-Liburd et al., 2015; Cao et al., 2015; Earley, 2015; Gepp et al., 2018; International Auditing and Assurance Standards Board Data Analytics Working Group, 2016; Issa et al., 2016; Jans et al., 2014; Kokina & Davenport, 2017; Yoon et al., 2015).

The feature-set results show how much value is added by the asymmetry construct. Account-pair surprisal and directional role violation improved performance over conventional indicators. AARR without the concordance count retained 24.15% capture, while full AARR captured 25.36%. Therefore, the improvement in performance does not arise simply from counting flags.

The structure of the feature set shows the importance of capturing the order in which accounts post debit and credit transactions. Conventional indicators include manual postings, period-end timing, delayed posting, and rounded values, but these indicators provided limited separation without structural context. Directional role violation represents departures from established debit-credit ordering, while account-pair surprisal represents how rarely an ordered pair of accounts has appeared in the company's historical records. A combined model can therefore assign a high score when directionality, pathway, timing, and magnitude are jointly unusual rather than because one

transaction characteristic is unusually extreme. Removing the concordance count from the full representation supports this interpretation.

Cross-mechanism evaluation extends the analysis beyond a single mechanism type. The four mechanisms were each removed from development positives in turn, and the representation continued to generate useful ranking performance for the respective held-out mechanisms. In an audit environment, a high score should direct attention to source documentation, approvals, business rationale, accounting treatment, related postings, and contradictory evidence. Such use is consistent with PCAOB guidance for journal entry testing, which emphasizes evidence and professional judgment (Public Company Accounting Oversight Board, 2025, 2026b).

The results illustrate comparative performance in simulated environments in which journal-entry attributes, labels, and information boundaries are explicitly defined. In practice, auditing involves assessing risk through accounting policy, management incentives, internal controls, available evidence, and the auditor's knowledge of the engagement. These factors motivate field studies examining how the representation interacts with audit plans and professional judgment. The transparent design also allows researchers to vary operating assumptions, increase complexity, and test alternative ranking models without access to confidential ledgers.

Scenario-positive documents are generated from predetermined combinations of partial deviations in directionality, pathways, timing, and amount behavior. The analysis therefore examines whether retaining this structural information improves ranking when observations share some characteristics and labels remain probabilistic. The results provide evidence about this representation within the described simulation family; they do not provide estimates of the prevalence, causes, or legal status of financial misconduct outside that simulated environment.

Several features of the study design support this interpretation. Scenario positivity incorporates unobserved approval, documentation, and support conditions. Legitimate atypical entries appear in the comparison population. No single feature fully separates positive and non-positive outcomes, and removing the concordance count still maintains strong performance for the full AARR specification. Cross-mechanism evaluation also examines performance when each mechanism family is excluded from development positives. Taken together, these analyses describe the contribution of the representation under controlled conditions.

The design also establishes limits to generalizability. Scenario-positive status is derived from a specified data-generating process and is not an estimate of real-world fraud prevalence or an audit conclusion. Overlap exists because of legitimate atypical entries and unobserved support conditions, but no simulation can reproduce every source of ambiguity in operational ledgers. Management incentives, internal controls, documentation practices, policy changes, and auditor knowledge are only partly represented or absent. Therefore, the supported claim is that directional asymmetry adds incremental ranking information under controlled conditions, not that the study estimates field false-positive costs, behavioral responses, or the sufficiency of audit evidence.

5.2. Implications for Journal Entry Testing

In future operational implementations, a directional asymmetry score could serve as a means of prioritizing documents. A high score would indicate that a document requires examination of source documentation, approvals, business rationale, accounting treatment, related postings, and contradictory evidence. The score would support the auditor's process for selecting documents for further investigation and should be considered together with other audit evidence. This human-in-the-loop position is consistent with audit-analytics research on judgment, algorithmic insights, and maintaining professional skepticism when advanced analytical tools are used (Brown-Liburd et al., 2015; Fedyk et al., 2022; Kokina & Davenport, 2017; Law & Shen, 2025; Samiolo et al., 2024).

Operational implementation would require the score to be part of an auditable review process. When a document receives a high rank, it would trigger examination of source documentation, approvals, business rationale, related postings, and potentially conflicting evidence rather than determine an audit conclusion. The fixed-budget design translates model performance into workload terms; however, there are costs associated with both false positives and missed items, so each engagement would need to consider those costs together with available resources and evidence.

5.3. External Public-ERP Evidence and Boundary Conditions

This external public-ERP benchmark provides descriptive evidence of representation transferability. The pooled capture rates of AARR-XGBoost (93.75%) and conventional XGBoost (75.00%), together with Isolation Forest matching AARR-XGBoost at 93.75%, indicate that the structural representation can be adapted to a different account structure, transaction schema, and labeling process. The results also identify a boundary condition: structural asymmetry does not necessarily perform better than strong unsupervised anomaly-detection baselines in every environment.

The public-ERP analysis further expands the scope of this study beyond the ledger generator used in the primary simulations without turning the analysis into operational validation. The benchmark was independently developed using a different account format, transaction schema, and labeling process. The directional representation retained a

high capture rate after adaptation, while Isolation Forest achieved the same pooled capture. This pattern is consistent with a strong generic anomaly signal in the public labels and identifies an important boundary condition: structural asymmetry can provide significant value beyond conventional supervised features but does not necessarily outperform every unsupervised detection technique. The benchmark also lacked preparer identity and several accounting-process fields and contained relatively few positive instances. Therefore, it provides independent synthetic-benchmark evidence rather than field validation and reinforces the need for operational studies of performance, false-positive burden, and auditor judgment.

The public benchmark is also a useful boundary test because it uses a different transaction representation, omits preparer identity and some process indicators, and contains few labelled positives. The adapted AARR variables remained usable, supporting portability of the representation logic, but Isolation Forest achieved the same pooled capture. The defensible external claim is therefore narrower: structural asymmetry adds information beyond conventional supervised features and remains implementable under materially different schemas, while comparative performance depends on the target population and labeling process.

5.4. Conclusion

The study provides a framework for evaluating journal-entry risk-ranking systems using repeated environments, operating-family holdouts, feature-set analysis, cross-mechanism evaluation, environment-level inference, and public-source transfer testing. The framework can be applied to assess competing models, expand current simulation architectures, and support future research using secure operational or de-identified ledgers in which reviewed outcomes can be linked to documentary evidence and auditor assessments.

Therefore, the next empirical step is not merely to increase model complexity, but to evaluate the proposed representation using operational or de-identified ledger populations whose review outcomes can be linked to supporting documentary evidence or auditor judgments. Subsequent studies could investigate ranking stability over time, sensitivity to process changes, false-positive costs, and how auditor judgments interact with recommended journal-entry rankings. They could also assess whether directional variables continue to add incremental value once other context-rich information becomes available. Until such evidence is available, the results should be viewed primarily as a controlled demonstration that historically conditioned debit-credit direction and pathway information can improve the capture of scenario-positive journal entries within a limited review budget.

Conflicts of Interest

The author declares no conflict of interest.

References

1. Almufadda, G., & Almezeini, N. A. (2022). Artificial intelligence applications in the auditing profession: a literature review. Journal of Emerging Technologies in Accounting, 19(2), 29–42. https://doi.org/10.2308/JETA-2021-038
2. Altman, E., Blanuša, J., Niederhäusern, L. von, Egressy, B., Anghel, A., & Atasu, K. (2023). Realistic Synthetic Financial Transactions for Anti-Money Laundering Models. Advances in Neural Information Processing Systems, 36, 29851–29874. https://doi.org/10.52202/075280-1300
3. Appelbaum, D., Kogan, A., & Vasarhelyi, M. A. (2017). Big data and analytics in the modern audit engagement: Research needs. Auditing: A Journal of Practice & Theory, 36(4), 1–27. https://doi.org/10.2308/ajpt-51684
4. Becerra-Suarez, F. L., Alvarez-Vasquez, H., & Forero, M. G. (2025). Improvement of Bank Fraud Detection Through Synthetic Data Generation with Gaussian Noise. Technologies, 13(4), 141. https://doi.org/10.3390/technologies13040141
5. Belgodere, B., Dognin, P., Ivankay, A., Melnyk, I., Mroueh, Y., Mojsilović, A., Navratil, J., Nitsure, A., Padhi, I., Rigotti, M., Ross, J., Schiff, Y., Vedpathak, R., & Young, R. A. (2024). Auditing and Generating Synthetic Data with Controllable Trust Trade-Offs. IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 14(4), 773–788. https://doi.org/10.1109/JETCAS.2024.3477976
6. Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324
7. Brown-Liburd, H., Issa, H., & Lombardi, D. (2015). Behavioral implications of Big Data's impact on audit judgment and decision making and future research directions. Accounting Horizons, 29(2), 451–468. https://doi.org/10.2308/acch-51023
8. Cao, M., Chychyla, R., & Stewart, T. (2015). Big data analytics in financial statement audits. Accounting Horizons, 29(2), 423–429. https://doi.org/10.2308/acch-51068

9. Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys (CSUR), 41(3), 1–58. https://doi.org/10.1145/1541880.1541882
10. Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. https://doi.org/10.1145/2939672.2939785
11. Davis, J., & Goadrich, M. (2006). The relationship between Precision-Recall and ROC curves. Proceedings of the 23rd International Conference on Machine Learning, 233–240. https://doi.org/10.1145/1143844.1143874
12. De Santis, F., & D'Onza, G. (2021). Big data and data analytics in auditing: in search of legitimacy. Meditari Accountancy Research, 29(5), 1088–1112. https://doi.org/10.1108/MEDAR-03-2020-0838
13. Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. Journal of Machine Learning Research, 7(Jan), 1–30.
14. Earley, C. E. (2015). Data analytics in auditing: Opportunities and challenges. Business Horizons, 58(5), 493–500. https://doi.org/10.1016/j.bushor.2015.05.008
15. Efron, B., & Tibshirani, R. J. (1993). An Introduction to the Bootstrap. Chapman and Hall.
16. Fedyk, A., Hodson, J., Khimich, N., & Fedyk, T. (2022). Is artificial intelligence improving the audit process? Review of Accounting Studies, 27(3), 938–985. https://doi.org/10.1007/s11142-022-09697-x
17. Gepp, A., Linnenluecke, M. K., O'Neill, T. J., & Smith, T. (2018). Big data techniques in auditing research and practice: Current trends and future opportunities. Journal of Accounting Literature, 40(1), 102–115. https://doi.org/10.1016/j.acclit.2017.05.003
18. Gronewald, J., Rombach, A. M., Stephan, S., & Fettke, P. (2024). Anomaly Detection in General Ledger Data: Results from a Hybrid Approach. Proceedings of the 1st International Conference on Auditing and Artificial Intelligence.
19. Guo, K. H., Yu, X., & Wilkin, C. (2022). A picture is worth a thousand journal entries: accounting graph topology for auditing and fraud detection. Journal of Information Systems, 36(2), 53–81. https://doi.org/10.2308/ISYS-2021-003
20. Hemati, H., Schreyer, M., & Borth, D. (2022). Continual learning for unsupervised anomaly detection in continuous auditing of financial accounting data. https://doi.org/10.48550/arXiv.2112.13215
21. Huang, Q., Schreyer, M., Michiles Jr, N. R., & Vasarhelyi, M. A. (2026). Connecting the dots: Graph neural networks for auditing accounting journal entries. Auditing: A Journal of Practice & Theory, 1–27. https://doi.org/10.2308/AJPT-2024-058
22. International Auditing and Assurance Standards Board. (2025). ISA 240 (Revised): The Auditor's Responsibilities Relating to Fraud in an Audit of Financial Statements. International Federation of Accountants. https://www.iaasb.org/publications/isa-240-revised-auditor-s-responsibilities-relating-fraud-audit-financial-statements
23. International Auditing and Assurance Standards Board Data Analytics Working Group. (2016). Exploring the Growing Use of Technology in the Audit, with a Focus on Data Analytics. International Auditing and Assurance Standards Board. https://www.iaasb.org/publications/exploring-growing-use-technology-audit-focus-data-analytics
24. Issa, H., Sun, T., & Vasarhelyi, M. A. (2016). Research Ideas for Artificial Intelligence in Auditing: The Formalization of Audit and Workforce Supplementation. Journal of Emerging Technologies in Accounting, 13(1), 1–20. https://doi.org/10.2308/jeta-10511
25. Jans, M., Alles, M. G., & Vasarhelyi, M. A. (2014). A field study on the use of process mining of event logs as an analytical procedure in auditing. The Accounting Review, 89(5), 1751–1773. https://doi.org/10.2308/accr-50807
26. Kokina, J., & Davenport, T. H. (2017). The emergence of artificial intelligence: How automation is changing auditing. Journal of Emerging Technologies in Accounting, 14(1), 115–122. https://doi.org/10.2308/jeta-51730
27. Krahel, J. P., & Titera, W. R. (2015). Consequences of Big Data and Formalization on Accounting and Auditing Standards. Accounting Horizons, 29(2), 409–422. https://doi.org/10.2308/acch-51065
28. Krawczyk, B. (2016). Learning from imbalanced data: open challenges and future directions. Progress in Artificial Intelligence, 5(4), 221–232. https://doi.org/10.1007/s13748-016-0094-0
29. Krieger, F., Drews, P., & Velte, P. (2021). Explaining the (non-) adoption of advanced data analytics in auditing: A process theory. International Journal of Accounting Information Systems, 41, 100511. https://doi.org/10.1016/j.accinf.2021.100511
30. Law, K. K. F., & Shen, M. (2025). How Does Artificial Intelligence Shape Audit Firms? Management Science, 71(5), 3641–3666. https://doi.org/10.1287/mnsc.2022.04040

31. Alrawashedh, N. H. (2023). The reality of social responsibility accounting in commercial banks listed on the Amman Stock Exchange. Banks and Bank Systems, 18(2), 63.
32. Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008). Isolation forest. 2008 Eighth IEEE International Conference on Data Mining, 413–422. https://doi.org/10.1109/ICDM.2008.17
33. Alrawashedh, N. H., Alsaleem, E., Alhawamdeh, H., & Zraqat, O. (2026). Impact of Cash Basis of Accounting Measurement on the Effectiveness of Public Expenditure Control Procedures. In Technology and Entrepreneurship: Systems Driving Innovation (pp. 273-285). Cham: Springer Nature Switzerland.
34. Manita, R., Elommal, N., Baudier, P., & Hikkerova, L. (2020). The digital transformation of external audit and its impact on corporate governance. Technological Forecasting and Social Change, 150, 119751. https://doi.org/10.1016/j.techfore.2019.119751
35. Alrawashedh, N. H. (2021). Evaluation of continuity impact under the Covid 19 pandemic, during the preparation of 2020 financial reports, and external auditors report of public limited shares companies in Jordan. Indian Journal of Economics and Business, 20(3), 291-310.
36. Mashiko, S., Kawamata, Y., Nakayama, T., Sakurai, T., & Okada, Y. (2025). Anomaly detection in double-entry bookkeeping data by federated learning system with non-model sharing approach. Scientific Reports, 15(1), 42208. https://doi.org/10.1038/s41598-025-26120-y
37. Alrawashdeh, N. H., & Alrawashdeh, N. H. (2015). Hedging of interest rate risk with interest rate futures. Int. J. Geol. Agric. Environ. Sci, 3(2), 8-23.
38. Morris, T. P., White, I. R., & Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. Statistics in Medicine, 38(11), 2074–2102. https://doi.org/10.1002/sim.8086
39. 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
40. Al Rawashdeh, N. H. H. (2013). The role of the auditor in verified of the unethical practices in accounting. International Journal of Scientific and Research Publications, 3(7), 1-9.
41. Perols, J. L., Bowen, R. M., Zimmermann, C., & Samba, B. (2017). Finding needles in a haystack: Using data analytics to improve fraud prediction. The Accounting Review, 92(2), 221–245. https://doi.org/10.2308/accr-51625
42. Habis Alrawashedh, N. (2023). Management accounting methods for financial decisions: Case of industrial companies in Jordan. Investment Management and Financial Innovations, 20(4), 60-68.
43. Alslihat, N., Alrawashdeh, N., Al-Ali, S., Al-Omari, M., Gazzaz, A. A., & Alali, H. (2017). Accounting disclosure of both financial and non-financial information in the light of international accounting standards. International Journal of Economic Research, 14(12), 105-118.
44. Public Company Accounting Oversight Board. (2025). Audit Focus: Journal Entries. Public Company Accounting Oversight Board. https://pcaobus.org/resources/staff-publications/audit-focus/audit-focus-journal-entries
45. Alrawashedh, N. H., & Shubita, M. F. (2024). Impact of digital transformation on the organization's financial performance: a case of Jordanian commercial banks listed on the Amman stock exchange. Banks and Bank Systems, 19(1), 126-134.
46. Public Company Accounting Oversight Board. (2026a). AS 2315: Audit Sampling. https://pcaobus.org/oversight/standards/auditing-standards/details/AS2315
47. Alrawashedh, N. H., Khaddam, L. A., & Alharafsheh, M. (2021). Voluntary disclosure of intellectual capital: the case of family and non-family businesses in Jordan. Psychology and Education, 58(1), 2819-2837.
48. Public Company Accounting Oversight Board. (2026b). AS 2401: Consideration of Fraud in a Financial Statement Audit. https://pcaobus.org/oversight/standards/auditing-standards/details/AS2401
49. Alrawashedh, N. H. (2023). Factors affecting organizational intention to adopt forensic accounting practices: A case of Jordan. Problems and Perspectives in Management, 21(3), 334-351.
50. Ravisankar, P., Ravi, V., Rao, G. R., & Bose, I. (2011). Detection of financial statement fraud and feature selection using data mining techniques. Decision Support Systems, 50(2), 491–500. https://doi.org/10.1016/j.dss.2010.11.006
51. Ruhnke, K. (2023). Empirical research frameworks in a changing world: The case of audit data analytics. Journal of International Accounting, Auditing and Taxation, 51, 100545. https://doi.org/10.1016/j.intaccaudtax.2023.100545
52. 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

53. Alrawashedh, N. H., Zureigat, B. N., Rawashdeh, B. S., Zraqat, O., & Hussien, L. F. (2025). Does corporate governance affect earnings management?. Evidence from an emerging market. Heritage and Sustainable Development, 7(1), 167-178.
54. Samiolo, R., Spence, C., & Toh, D. (2024). Auditor Judgment in the Fourth Industrial Revolution. Contemporary Accounting Research, 41(1), 498–528. https://doi.org/10.1111/1911-3846.12901
55. Tritscher, J., Gwinner, F., Schlör, D., Krause, A., & Hotho, A. (2022). Open ERP System Data for Occupational Fraud Detection. arXiv. https://doi.org/10.48550/arXiv.2206.04460
56. Varma, S., & Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics, 7, 91. https://doi.org/10.1186/1471-2105-7-91
57. 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
58. Yoon, K., Hoogduin, L., & Zhang, L. (2015). Big data as complementary audit evidence. Accounting Horizons, 29(2), 431–438. https://doi.org/10.2308/acch-51076