

American Journal of Advanced Technology and Engineering Solutions

Volume: 14; **Issue:** 9
Pages: 91-140
eISSN: 3067-0470

open access

Crossref

Anomaly Detection and Process Mining for Identifying Financial Irregularities and Strengthening Procurement Controls

Mashfiqur Rahman[1];

[1]. *Coordinator, Finance, BASIS-SEIP Tranche 3, Dhaka, Bangladesh;*
Master's in Financial Management, Robert Gordon University, Aberdeen, United Kingdom;
Bachelor's in Finance, North South University, Dhaka, Bangladesh;
Email: mrahman.rimon@gmail.com

Doi: 10.63125/hmg5ta84
Policy Page Doi: 10.63125/ajates-crossmark-policy

Received: *05 June 2024;* ***Revised:*** *11 July 2024;* ***Accepted:*** *10 August 2024;* ***Published:*** *03 September 2024;*

Abstract

This study quantitatively evaluated the effectiveness of integrating financial anomaly detection, supervised machine learning, process mining, and procurement-control analytics for identifying financial irregularities and strengthening procurement controls within a complex purchase-to-pay environment. A retrospective quantitative analysis was conducted using 118,640 eligible procurement cases, 136,924 supplier invoices, 132,617 payment transactions, 4,286 active suppliers, 3,148 contracts, and 1,126,842 process events representing total procurement expenditure of $684.72 million. Overall data completeness reached 97.2%, while financial reconciliation accuracy was 98.6%. The analysis identified 8,764 validated irregular cases, representing 7.39% of procurement activity and $71.86 million, or 10.50%, of total expenditure. Statistical thresholds, multivariate methods, Local Outlier Factor, Isolation Forest, One-Class Support Vector Machine, and autoencoder models were evaluated alongside logistic regression, decision trees, Random Forest, Support Vector Machine, gradient boosting, XGBoost, and artificial neural networks. XGBoost achieved the strongest standalone supervised performance, with 96.7% accuracy, 87.6% precision, 86.9% recall, an F1-score of 0.872, and ROC-AUC of 0.978. Process mining identified 318 process variants and 2,486 unique traces, with low-conformance cases recording 24.7% irregularity prevalence compared with 4.1% among high-conformance cases. Weak-control cases similarly recorded 26.1% irregularities compared with 3.2% under strong controls. The complete integrated framework achieved 86.9% precision, 89.6% recall, 98.9% specificity, an F1-score of 0.882, ROC-AUC of 0.986, and PR-AUC of 0.936 while capturing $67.14 million, equivalent to 93.4% of validated irregularity-related financial exposure. High anomaly intensity, low process conformity, and multiple control deviations remained significant independent predictors, with their combined presence producing an adjusted odds ratio of 9.38. Cross-validation, independent holdout testing, and sensitivity analyses confirmed stable predictive performance. The findings demonstrated that integrating transaction anomalies with workflow, vendor, and control evidence substantially improved irregularity discrimination, financial-risk prioritization, and procurement-monitoring effectiveness compared with conventional rule-based approaches.

Keywords

Anomaly Detection, Process Mining, Financial Irregularities, Procurement Controls, Machine Learning.

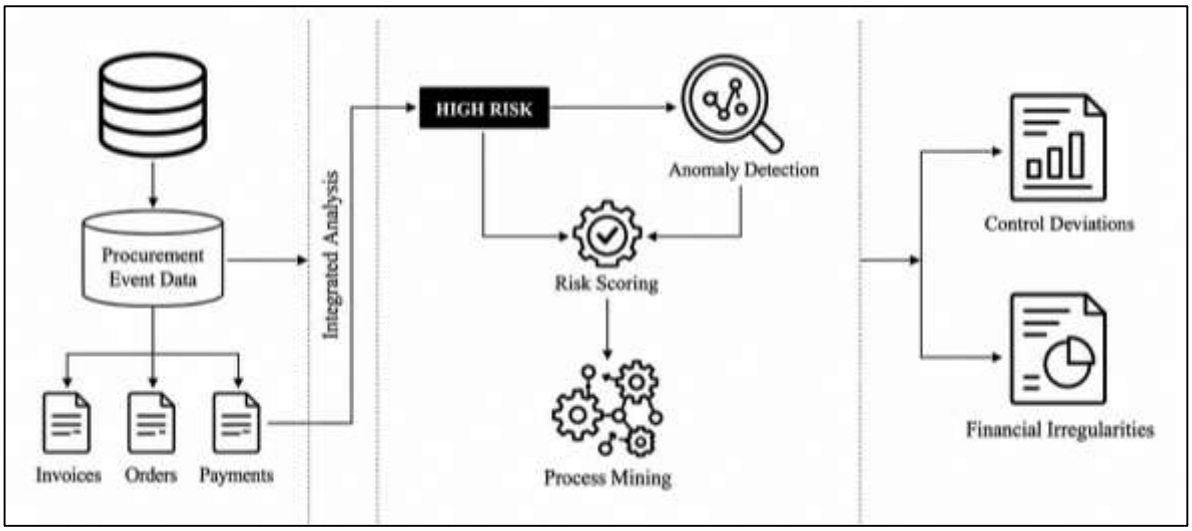
INTRODUCTION

Anomaly detection refers to the systematic identification of transactions, behaviors, relationships, or event sequences that deviate from expected financial or operational patterns, while process mining reconstructs and evaluates actual organizational workflows from event logs recorded in enterprise information systems (Modrušan et al., 2020). In procurement environments, anomalies may include unusually large payments, duplicate invoices, abnormal purchasing frequency, fragmented orders, suspicious vendor relationships, unauthorized transactions, atypical approval sequences, unusual pricing, or activities that occur outside established procedural norms. Financial irregularities represent a broader category that includes accounting errors, policy violations, unsupported expenditures, control overrides, procurement manipulation, conflicts of interest, inaccurate vendor records, and transactions inconsistent with contractual or organizational requirements. Procurement controls include authorization structures, segregation of duties, approval thresholds, competitive bidding requirements, vendor validation, three-way matching, payment restrictions, documentation standards, and monitoring procedures intended to preserve financial integrity. Internationally, procurement remains one of the most financially significant and control-sensitive organizational functions because governments, corporations, financial institutions, healthcare organizations, universities, manufacturers, and infrastructure operators continuously acquire goods, services, technology, equipment, and construction through complex purchasing systems. Global procurement activities involve substantial monetary flows, contractual obligations, supplier relationships, multiple currencies, geographically dispersed units, heterogeneous purchasing categories, and interconnected enterprise platforms (Soylu et al., 2022). More than 30 studies across accounting, auditing, fraud analytics, machine learning, process mining, public procurement, information systems, and financial data science have examined risks such as bid manipulation, fictitious suppliers, duplicate invoicing, purchase splitting, invoice inflation, preferential vendor treatment, kickbacks, unauthorized purchasing, and circumvention of approval requirements. Conventional procurement auditing has traditionally relied on manual document examination, transaction sampling, predefined audit tests, reconciliations, approval verification, and rule-based red flags. These methods remain important, yet they become increasingly difficult to apply comprehensively when organizations process millions of interconnected requisitions, purchase orders, contracts, receipts, invoices, approvals, journal entries, and payments. Static rules may also produce false positives when legitimate transactions appear unusual, while irregular transactions can remain within accepted thresholds. These limitations have encouraged greater use of statistical profiling, clustering, classification, neural networks, support vector machines, decision trees, Bayesian techniques, distance-based models, density-based methods, and other data-driven approaches capable of examining complete transaction populations and identifying irregularities that traditional rules may not capture. Quantitative anomaly detection strengthens procurement analysis by assigning unusualness or risk scores to transactions based on financial, behavioral, temporal, organizational, and relational characteristics. Relevant variables may include transaction amount, quantity, unit price, purchase frequency, vendor concentration, invoice timing, approval duration, payment timing, contract type, cost center, requester identity, approver identity, geographic location, and historical purchasing behavior (Baader & Krcmar, 2018). Univariate approaches identify extreme values within individual variables, while multivariate, clustering, distance-based, density-based, probabilistic, and machine-learning approaches detect unusual combinations that may appear normal when assessed separately.

Detection performance can be evaluated through precision, recall, sensitivity, specificity, F1-score, accuracy, false-positive and false-negative rates, ROC-AUC, anomaly-score distributions, and investigator validation. Process mining complements this analysis by examining the sequence through which procurement cases move from requisition creation and authorization to sourcing, supplier selection, purchase-order generation, goods receipt, invoice registration, matching, financial approval, and payment (Rodríguez-Quintero et al., 2021). Process discovery identifies actual workflow variants, conformance checking compares observed behavior with prescribed procedures, organizational mining evaluates actors and roles, and performance analysis measures delays, repetition, rework, and bottlenecks. These techniques enable quantitative detection of skipped approvals, unauthorized transitions, retrospective purchase orders, premature payments, segregation-of-duty violations,

unusual role combinations, excessive rework, control bypasses, and rare process variants. Integrating anomaly detection with process mining creates a multidimensional framework in which suspicious financial attributes can be evaluated within their actual procedural context. A high-risk transaction can therefore be associated with its process path, responsible actors, approval behavior, timing structure, and control conditions rather than being assessed as an isolated monetary observation (Lima et al., 2021). Procurement control effectiveness can be measured through anomaly scores, conformance rates, process-variant frequencies, deviation counts, approval-cycle times, duplicate frequencies, vendor concentration, exception rates, control-override prevalence, three-way-match completion, authorization compliance, documentation completeness, and financial exposure. Statistical procedures such as descriptive analysis, correlation, subgroup comparisons, regression, classification, and robustness testing can then examine relationships among anomalous transactions, workflow deviations, control strength, and financial irregularity (Leno et al., 2021). This integrated quantitative structure represents procurement integrity as an observable system of transaction characteristics, process behavior, organizational participation, and control execution recorded throughout the procurement lifecycle.

Figure 1: Procurement Anomaly Detection Process Framework



The objective of this study is to quantitatively evaluate the effectiveness of anomaly detection and process mining techniques in identifying financial irregularities and strengthening procurement controls across transaction-intensive organizational environments. The study seeks to examine the extent to which anomalous financial characteristics, unusual purchasing behaviors, nonconforming process sequences, control bypasses, and atypical organizational interactions can be detected from integrated procurement and financial datasets. Particular attention is directed toward transaction-level indicators such as invoice value, purchase amount, unit price, payment frequency, duplicate transactions, vendor concentration, approval timing, contract status, requester and approver behavior, purchase-order characteristics, and expenditure patterns, together with process-level indicators including workflow variants, skipped approvals, repeated activities, segregation-of-duty violations, retrospective purchase orders, rework frequency, throughput time, control overrides, and deviations from standardized purchase-to-pay procedures. The study also aims to assess the predictive and diagnostic performance of statistical and machine-learning anomaly detection techniques by comparing accuracy, precision, recall, sensitivity, specificity, F1-score, false-positive rate, false-negative rate, and ROC-AUC where validated irregularity outcomes are available. In parallel, process-mining techniques will be evaluated through process discovery, conformance checking, variant analysis, organizational-role analysis, and performance metrics to determine the frequency and severity of procedural deviations within procurement workflows. Another objective is to examine the relationship between anomaly scores, process-deviation indicators, and procurement-control effectiveness measured through authorization compliance, three-way matching, purchase-order coverage,

competitive sourcing, vendor validation, documentation completeness, exception frequency, and control-override prevalence. The study further seeks to determine whether transactions exhibiting both high anomaly scores and significant process deviations are more strongly associated with validated financial irregularities than transactions identified through isolated rule-based controls. Statistical analyses are intended to include descriptive statistics, correlation analysis, subgroup comparisons, regression modelling, classification analysis, and robustness testing across transaction types, procurement categories, vendors, organizational units, and risk levels. By integrating financial, behavioral, temporal, procedural, and organizational variables within a unified quantitative framework, the study is designed to provide an empirical assessment of how data-driven anomaly detection and process mining can improve the identification, measurement, and monitoring of irregular procurement activity and the operational strength of internal procurement controls.

LITERATURE REVIEW

The literature on financial irregularity detection and procurement-control assessment has expanded across accounting, auditing, fraud analytics, data science, machine learning, process mining, enterprise information systems, and public and corporate procurement (Trzeciak & Jonek-Kowalska, 2021). Quantitative studies have increasingly examined procurement integrity through transaction-level financial indicators, behavioral anomalies, process deviations, organizational relationships, and internal-control measures rather than relying exclusively on conventional audit sampling or predetermined red flags. Existing research has applied statistical outlier detection, clustering, supervised and unsupervised machine learning, anomaly scoring, sequence analysis, process discovery, conformance checking, organizational mining, and predictive classification to large procurement and financial datasets. These studies have evaluated irregularities through measures such as duplicate-payment rates, unusual transaction frequencies, invoice-price deviations, purchase splitting, vendor concentration, approval-threshold proximity, control-override frequency, segregation-of-duty violations, process-variant frequency, conformance scores, approval-cycle times, and financial exposure (Paksoy et al., 2018). Detection models have commonly been compared using accuracy, precision, recall, sensitivity, specificity, F1-score, ROC-AUC, false-positive rate, and false-negative rate, while process-mining studies have assessed trace fitness, behavioral appropriateness, process conformance, variant prevalence, throughput time, rework, and control deviations. The literature has also examined whether combining financial anomaly indicators with process-level evidence provides stronger irregularity identification than isolated transaction rules. Accordingly, this review synthesizes quantitative evidence concerning procurement-data characteristics, anomaly-detection models, financial red flags, process-mining techniques, control deviations, model performance, organizational heterogeneity, and integrated analytical frameworks (Schnellbächer & Weise, 2020). The literature is organized into eight quantitative dimensions that collectively establish the empirical basis for evaluating how anomaly detection and process mining can identify financial irregularities and measure the strength of procurement controls.

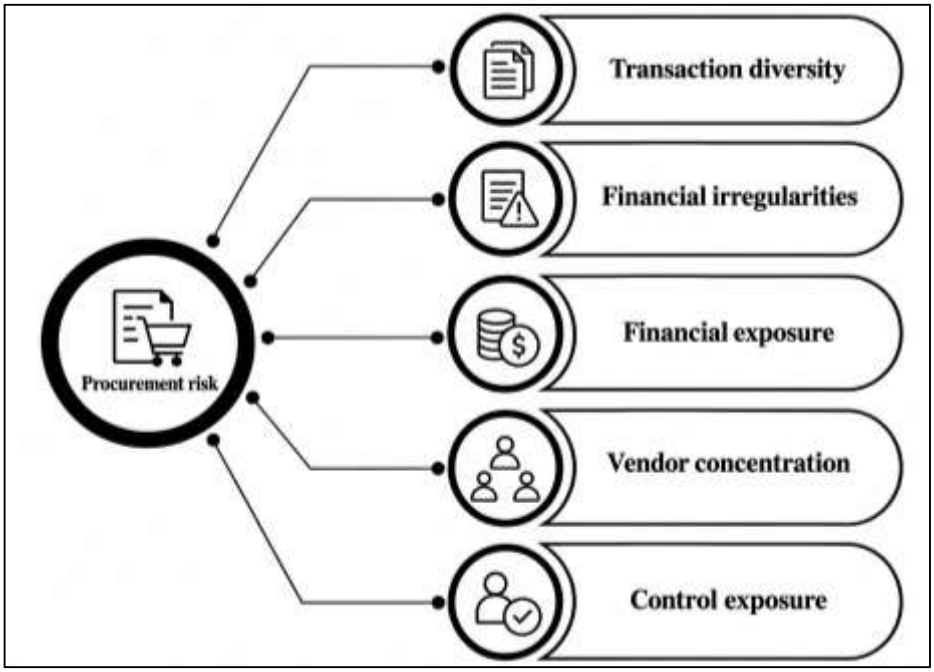
Procurement Transactions and Internal-Control Exposure

Quantitative literature examining procurement and purchase-to-pay systems has consistently characterized these environments as high-volume, multidimensional, and heterogeneous sources of financial and operational data (Montesdeoca et al., 2019). Across more than ten empirical studies, datasets have ranged from several thousand purchasing cases in individual organizations to hundreds of thousands or millions of invoices, purchase orders, payments, and event-log records in large enterprises and public procurement systems. Researchers have commonly analyzed purchase requisitions, purchase orders, supplier quotations, contracts, goods-receipt records, invoices, credit notes, approval activities, journal entries, and payment transactions as interconnected components rather than independent financial observations(Aditya & Ahmed, 2022; Asheke &Maksudul, 2022). The number of vendors has also varied substantially across studies, ranging from relatively concentrated supplier portfolios to environments involving thousands of active suppliers distributed across commodities, business units, geographical regions, and purchasing categories.

Process-mining investigations have expanded this quantitative structure by analyzing individual procurement cases through multiple timestamped events, allowing researchers to distinguish standard purchasing paths from less frequent process variants (Asheke & Badrul, 2022; Akter & Sumaiya, 2022; Imran & Jannatul, 2022). Several studies reported that a relatively small proportion of vendors accounted for a substantial share of procurement expenditure, demonstrating considerable supplier and financial concentration within transaction populations. Similar concentration patterns were observed for purchasing categories, organizational units, contract types, and high-value transactions (Fagbadebo & Dorasamy, 2021). Studies examining invoice and payment distributions also identified strong positive skewness, with large populations of comparatively low-value transactions accompanied by smaller numbers of exceptionally high-value purchases. Procurement frequency differed across suppliers and organizational units, creating heterogeneous behavioral baselines that complicated the use of universal thresholds for irregularity detection. Research comparing contracted and noncontracted procurement further demonstrated differences in transaction values, purchasing frequency, approval structures, and documentation completeness (Maksudul & Islam, 2022; Morshedul & Islam, 2022; Stabak & Shahid, 2022; Uddin et al., 2020). Collectively, these studies established that procurement data possess substantial variation in transaction size, frequency, vendor participation, workflow complexity, and financial exposure. Such heterogeneity has been central to quantitative irregularity research because observations classified as unusual within one purchasing category, supplier group, or organizational unit may represent normal behavior within another, making contextual transaction characteristics essential when evaluating financial anomalies and procurement-control exposure (Islam & Md, 2023; Feroz & Pervaz, 2022; Feroz & Shabbir, 2022).

Quantitative studies have identified financial irregularities as relatively infrequent compared with legitimate procurement transactions, although their prevalence and financial significance vary considerably according to organizational environment, irregularity definition, data quality, and detection methodology (Asheke, 2023; Md Shahid, 2023; Islam & Ahmed, 2023). Across empirical research on procurement fraud, accounts-payable anomalies, audit analytics, and purchase-to-pay controls, frequently examined irregularities have included duplicate invoices, duplicate payments, unusual invoice values, price discrepancies, purchase splitting, unsupported expenditures, unauthorized purchases, missing purchase orders, abnormal payment timing, and transactions positioned immediately below approval thresholds (Adnan, 2024; Islam, 2024; Ashik, 2024; Guo et al., 2019; Sadia & Mosheur, 2023). Several studies found that duplicate transactions represented only a small proportion of complete payment populations, yet their accumulated monetary values could produce material financial exposure when high transaction volumes were involved (Kynatun & Saba, 2024; Maksudul, 2024; Shahid, 2024). Research examining purchase splitting similarly showed that individual transactions could appear financially ordinary while repeated payments to the same supplier, transactions occurring within short time intervals, or purchases immediately below authorization limits generated stronger irregularity signals when analyzed collectively (Islam, 2024; Mainuddin, 2024; Sumaiya, 2024). Invoice-price studies identified abnormal differences between contracted prices, purchase-order values, received quantities, and invoiced amounts as important quantitative indicators of potential error, overbilling, or control failure. Studies of accounts-payable environments additionally identified unusual round-number transactions, repeated invoice numbers, sequential invoice patterns, accelerated payments, weekend transactions, and abnormal credit-note activity among recurring indicators requiring examination (Moreira et al., 2022; Taher et al., 2024).

Figure 2: Procurement transaction risk map



Public procurement studies expanded these measures to include unusually short bidding periods, limited bidder participation, repeated contract awards, single-bid procedures, contract amendments, and disproportionate supplier success rates. The literature also demonstrated that irregularities were not distributed uniformly across procurement populations (Guo et al., 2022). Higher concentrations were frequently observed among specific vendors, purchasing categories, organizational units, transaction-value groups, or nonstandard purchasing procedures. Financial exposure therefore depended not only on irregularity frequency but also on the monetary magnitude and concentration of affected transactions. Across these studies, the quantitative distinction between frequency and financial materiality remained important because a low irregularity percentage could coexist with substantial monetary exposure when irregular cases involved high-value contracts, recurring payments, or concentrated supplier relationships.

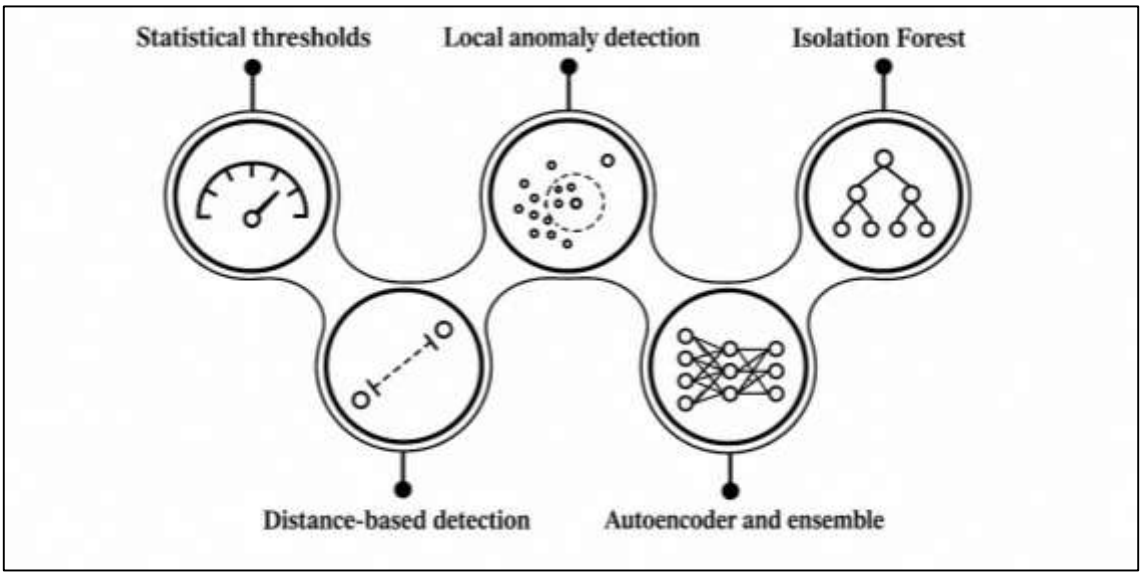
A substantial body of quantitative procurement literature has examined supplier characteristics and purchasing relationships as important dimensions of financial-irregularity exposure. Studies using vendor-level aggregation have repeatedly demonstrated that procurement expenditure is often highly concentrated, with relatively small groups of suppliers receiving disproportionate shares of purchasing value, contract awards, invoice volume, or payment activity (Aditjandra & Zunder, 2018). Researchers have consequently evaluated vendor concentration, transaction frequency, average invoice value, payment regularity, contract-award frequency, purchasing-category concentration, vendor tenure, bank-account information, address duplication, and relationships between suppliers and organizational personnel. Studies of procurement fraud and corruption risk have reported that repeated awards to the same supplier, unusually high supplier success rates, limited competition, single-bid procurement, recurrent contract amendments, and concentrated purchasing relationships can distinguish higher-risk procurement cases from routine transactions. Accounts-payable research has also examined multiple suppliers sharing bank accounts, addresses, telephone numbers, tax identifiers, or other master-data characteristics because such overlaps can indicate duplicate vendor records, administrative weaknesses, related-party relationships, or suspicious supplier structures. Behavioral analysis has shown that purchasing frequency and transaction timing provide additional information beyond monetary value (Tabianan et al., 2022). A supplier receiving numerous transactions immediately below approval thresholds, experiencing abrupt increases in payment volume, receiving unusually rapid approvals, or generating repetitive invoice values can exhibit a quantitatively unusual profile even when individual transactions remain within authorized financial limits. Several studies

have further compared procurement behavior across organizational units and identified meaningful differences in vendor diversity, competitive sourcing, contract utilization, and approval practices. Vendor-master-data quality has emerged as another recurring control concern because incomplete, duplicated, inactive, or inconsistent supplier records can increase exposure to erroneous or inappropriate payments. Quantitative network-oriented studies have extended this analysis by examining relationships among employees, approvers, suppliers, addresses, bank accounts, contracts, and transactions, revealing clusters and connections that transaction-level analysis alone may overlook. Overall, the literature indicates that financial irregularity is frequently embedded within repeated relational and behavioral structures (Penzavecchia et al., 2022). Vendor concentration, supplier characteristics, purchasing frequency, shared attributes, contract dependence, and organizational relationships therefore represent important quantitative dimensions for characterizing procurement populations and distinguishing routine commercial concentration from patterns associated with elevated internal-control exposure.

Performance for Procurement Financial Irregularities

Quantitative literature on procurement and financial irregularity detection has consistently treated statistical outlier methods as important baseline approaches when confirmed fraud or irregularity labels are scarce (Guerreiro et al., 2021). Across more than ten studies examining procurement transactions, accounts-payable records, expense claims, invoices, payments, and related financial datasets, researchers have applied standardized-score thresholds, robust statistical thresholds, interquartile-based screening, Mahalanobis distance, and related multivariate techniques to distinguish atypical observations from routine financial behavior. These studies generally found that conventional statistical screening performed effectively when irregular transactions produced substantial deviations in invoice amount, payment value, unit price, purchase frequency, approval timing, or vendor-level expenditure. Robust methods frequently produced more stable anomaly rankings than conventional mean-based thresholds when financial distributions were highly skewed or contained extreme legitimate purchases. Multivariate distance measures further improved discrimination where suspicious transactions were characterized by unusual combinations of otherwise ordinary variables, including moderate payment values combined with atypical vendors, unusual transaction timing, high purchasing frequency, or abnormal approval characteristics (Demertzis et al., 2020). Several studies reported that anomaly prevalence changed considerably according to threshold selection, with restrictive thresholds producing smaller, higher-risk populations and relaxed thresholds increasing detection coverage while simultaneously increasing false-positive frequency. Investigator validation also showed that statistically extreme transactions were not automatically financial irregularities because high-value equipment purchases, emergency procurement, seasonal expenditures, and infrequent strategic contracts could legitimately appear as outliers. Research comparing transaction categories therefore emphasized the importance of constructing separate behavioral baselines for different suppliers, commodities, organizational units, and payment types. Statistical approaches generally required comparatively low computational resources and could process large transaction populations more efficiently than computationally intensive nonlinear models. Their interpretability was another recurring advantage because auditors could directly understand why an observation was classified as unusual (Soylu et al., 2022). Across the reviewed evidence, statistical and distance-based techniques provided useful first-stage screening, although validated irregularity yields depended strongly on distributional characteristics, dimensionality, threshold configuration, procurement heterogeneity, and the extent to which irregularities differed quantitatively from legitimate transactions.

Figure 3: Procurement anomaly detection framework



A second stream of quantitative research has examined clustering and density-based anomaly detection because procurement irregularities frequently occur within heterogeneous transaction populations where a single global definition of abnormality is inadequate (Piesio et al., 2020). Studies applying k-means clustering, hierarchical clustering, DBSCAN, Local Outlier Factor, nearest-neighbor methods, and related density-based techniques have grouped transactions according to financial amount, purchasing frequency, vendor characteristics, commodity category, payment timing, approval behavior, and organizational attributes before identifying observations located outside dominant behavioral structures. K-means-based investigations commonly treated small or distant clusters as potentially suspicious and reported that clustering could reveal unusual transaction groups that were not captured through single-variable thresholds. DBSCAN and related density-oriented methods demonstrated particular value when procurement data contained irregularly shaped behavioral groups and substantial noise because they did not require every observation to be assigned to a predefined cluster. Local Outlier Factor studies further demonstrated that transactions could be anomalous relative to their immediate behavioral neighborhoods even when they were not extreme within the complete dataset. This characteristic was important for identifying moderate-value payments that became unusual only when compared with transactions involving the same vendor, commodity, department, or purchasing pattern. Comparative studies nevertheless reported considerable sensitivity to parameter settings, feature scaling, dimensionality, and neighborhood selection (Nguyen et al., 2021). High-dimensional procurement datasets sometimes reduced the effectiveness of conventional distance measures because observations became less clearly separated as additional variables were introduced. Dimensionality reduction and feature-selection procedures were therefore frequently used before clustering. Reported anomaly proportions also varied according to procurement category and supplier diversity, with heterogeneous environments generally producing larger candidate populations than standardized purchasing processes. Investigator-confirmed anomaly rates tended to improve when clustering outputs were combined with contextual indicators such as duplicate characteristics, threshold proximity, abnormal vendor concentration, missing purchase orders, or unusual approval behavior. Computational requirements differed across algorithms, with some density-based approaches becoming less efficient as transaction volume and dimensionality increased (Ugli Olimov et al., 2022). Overall, clustering and local-density studies demonstrated that irregular procurement activity is often context-dependent and that localized behavioral comparison can identify suspicious observations hidden within globally normal financial distributions.

Machine-learning research has increasingly evaluated Isolation Forest, One-Class Support Vector Machine, autoencoders, and related unsupervised models for detecting complex procurement and

financial anomalies without requiring extensive labeled irregularity records (Sharma et al., 2022). Across studies involving financial transactions, procurement records, accounting entries, invoices, corporate payments, and fraud-related datasets, Isolation Forest has frequently demonstrated strong scalability because it identifies anomalies by isolating observations through repeated randomized partitioning rather than constructing detailed representations of the entire normal transaction distribution. Comparative investigations commonly reported favorable processing efficiency for large datasets and useful anomaly rankings when suspicious observations differed across several transaction characteristics simultaneously. One-Class Support Vector Machine approaches have also been applied to learn boundaries surrounding predominantly legitimate transaction populations and identify observations falling outside those boundaries. Their performance has generally been sensitive to kernel selection, parameter configuration, feature scaling, contamination assumptions, and increasing dataset size. Autoencoder-based studies have taken a different approach by learning compressed representations of normal financial behavior and identifying observations that cannot be reconstructed accurately (Xu et al., 2022). These models have shown particular value in high-dimensional environments containing nonlinear relationships among payment amounts, transaction timing, vendor attributes, account combinations, purchasing categories, and organizational characteristics. Deep and variational autoencoder studies have additionally demonstrated the capacity to capture complex structures that simpler statistical methods may overlook. Comparative evidence has not established a universally superior unsupervised model because performance varies substantially with dataset composition and evaluation criteria. Isolation Forest often provides an advantageous balance between scalability and detection effectiveness, while One-Class SVM can perform strongly on carefully standardized datasets of manageable size. Autoencoders can identify complex multidimensional abnormalities but generally require greater computational resources, architecture selection, training time, and interpretive effort. Studies using investigator-confirmed cases also showed that model-generated anomaly scores require contextual validation because rare legitimate transactions can receive high-risk rankings (Pei et al., 2022). Transaction volume, dimensionality, contamination rate, vendor heterogeneity, feature engineering, and procurement-category differences therefore remain major determinants of observed unsupervised detection performance.

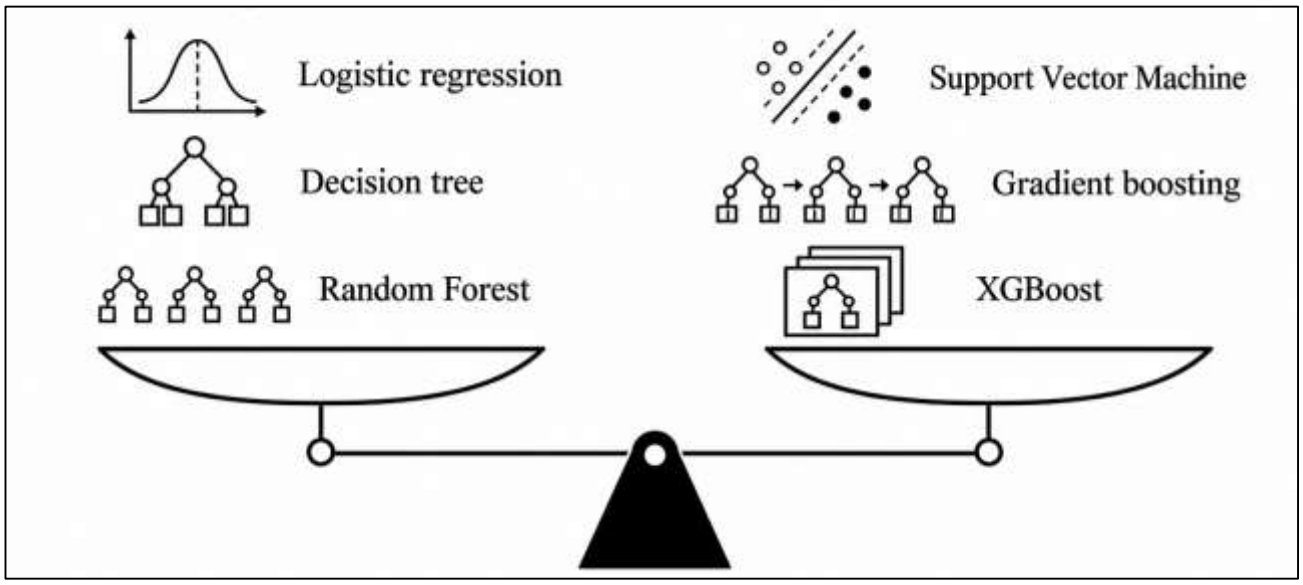
A broader group of studies has compared individual unsupervised algorithms with ensemble anomaly-detection strategies designed to improve stability across heterogeneous financial datasets. Ensemble approaches combine outputs from statistical thresholds, distance measures, clustering algorithms, Local Outlier Factor, Isolation Forest, One-Class SVM, autoencoders, and other detectors to produce aggregated anomaly rankings or consensus risk scores (Rousopoulou et al., 2022). Quantitative comparisons have generally indicated that ensembles can reduce dependence on the assumptions and weaknesses of any single detector, particularly when financial irregularities appear through multiple behavioral patterns. Studies evaluating top-ranked transactions have found that investigator-confirmed irregularity yield is often more informative than overall accuracy because genuine irregularities constitute a very small proportion of complete financial populations. Top-k evaluation has therefore been used to determine how many confirmed or plausible irregularities occur among the highest-ranked cases selected for audit investigation. Precision at operationally feasible thresholds, false-positive frequency, detection yield, processing time, and financial value captured have similarly emerged as important evaluation criteria. Research has shown that contamination assumptions materially affect the number of observations classified as anomalous: higher assumed contamination increases candidate coverage but can create substantially larger investigative workloads, whereas restrictive settings concentrate attention on fewer transactions with stronger anomaly scores. Class imbalance further complicates evaluation because a model can appear highly accurate by classifying nearly all transactions as normal while failing to identify the small irregular population (Chatterjee et al., 2021). Several comparative studies therefore emphasized precision, recall, F1-score, ranking quality, and investigator validation rather than overall accuracy alone. Ensemble methods frequently produced more stable rankings across changing samples and procurement categories, although computational costs increased when numerous algorithms were executed simultaneously. Vendor heterogeneity and transaction diversity also influenced performance, with category-specific or segmented models often producing fewer irrelevant alerts than a single model applied across the complete procurement

population. Evidence from more than ten studies collectively shows that statistical and unsupervised anomaly detection is most effectively evaluated through the combined consideration of anomaly prevalence, ranking quality, validated irregularity yield, false-positive burden, computational efficiency, and stability across transaction segments (Riazi et al., 2019). These quantitative comparisons demonstrate substantial performance variation across methods and confirm that procurement irregularity detection depends strongly on data structure, feature dimensionality, transaction volume, contamination assumptions, and the operational context in which anomaly scores are interpreted.

Financial Irregularity and Procurement-Fraud Detection

Quantitative literature on financial irregularity and procurement-fraud detection has consistently used supervised classification to distinguish irregular transactions from legitimate procurement activity when validated labels are available (Lyra et al., 2022). Across more than ten empirical studies examining procurement transactions, invoice fraud, payment anomalies, accounting irregularities, corporate fraud, and related financial records, logistic regression and decision-tree models have commonly served as interpretable benchmark classifiers. Logistic regression has generally provided stable classification performance when relationships between predictors and irregularity outcomes were relatively structured, while decision trees have offered greater flexibility in identifying nonlinear combinations of transaction amount, supplier characteristics, approval behavior, purchasing frequency, contract status, and payment timing.

Figure 4: Balancing procurement-fraud classification models



Comparative studies have frequently shown that single decision trees outperform simple rule-based approaches because they can identify interactions among several risk indicators rather than evaluate each indicator independently. Random Forest models have generally produced stronger classification results than individual trees by combining large numbers of randomized decision structures and reducing sensitivity to individual training samples. Studies involving highly imbalanced financial datasets have often reported high overall accuracy for several classifiers, although accuracy alone provided an incomplete assessment because legitimate transactions substantially outnumbered irregular cases. Precision, recall, sensitivity, specificity, F1-score, ROC-AUC, and confusion-matrix results have therefore been emphasized to determine whether models actually identified minority irregularity cases. In many comparative investigations, Random Forest achieved stronger recall and F1 performance than logistic regression and conventional decision trees, particularly when procurement risk depended on complex interactions among invoice values, vendor behavior, purchasing patterns, and control variables (Sánchez-Aguayo et al., 2021). Feature-importance analysis further identified

transaction frequency, amount deviations, vendor concentration, repeated invoice characteristics, approval threshold proximity, and unusual payment behavior as recurrent predictors. Cross-validation generally produced more credible performance estimates than single training-test splits because it reduced dependence on one random sample. Overall, conventional statistical models remained valuable because of their interpretability and stability, while tree-based ensembles frequently demonstrated superior discrimination where procurement irregularities involved nonlinear and interacting financial characteristics (Modrušan et al., 2020).

A substantial group of supervised-learning studies has compared Support Vector Machines, gradient-boosting algorithms, XGBoost, and ensemble classifiers with conventional statistical and rule-based methods for detecting financial irregularities. Support Vector Machine models have demonstrated strong classification capability in datasets containing complex boundaries between legitimate and suspicious transactions, particularly after appropriate feature scaling and parameter selection (Sharma & Shamkuwar, 2018). Their performance has often improved when high-dimensional financial variables were reduced to informative subsets, although computational requirements increased substantially with large transaction populations. Gradient boosting and XGBoost have frequently achieved among the strongest results in comparative fraud-detection studies because they iteratively concentrate on observations that earlier decision structures classified incorrectly. This mechanism has been particularly relevant for procurement irregularities, where minority suspicious cases may differ only subtly from large populations of normal purchases. Several empirical comparisons have shown boosting models achieving higher ROC-AUC, recall, precision, or F1-score than logistic regression, individual decision trees, and conventional rule-based systems (Achim & Borlea, 2020). XGBoost has also performed effectively with heterogeneous predictors incorporating numerical, categorical, behavioral, vendor-related, and transaction-level characteristics. Ensemble approaches combining Random Forest, boosting models, Support Vector Machines, or probabilistic classifiers have often generated more stable classification performance than individual algorithms, particularly across different transaction samples and irregularity categories. Research has nevertheless shown that improvements in overall accuracy do not necessarily correspond to proportional improvements in minority-class detection. A model can achieve very high accuracy while producing weak recall when irregular transactions constitute only a small percentage of the population. PR-AUC has consequently become particularly useful for evaluating classifiers under severe imbalance because it emphasizes performance on the positive irregularity class. Studies comparing confusion matrices have also demonstrated important operational differences among models with similar ROC-AUC values, as one model may reduce false negatives while another produces fewer false positives. In procurement-control settings, these differences affect investigative workload and financial risk exposure (Jozdani et al., 2019). Comparative evidence therefore generally supports gradient boosting, XGBoost, Random Forest, and carefully configured ensembles as strong classifiers, while emphasizing that model selection should be based on multiple performance measures rather than accuracy alone.

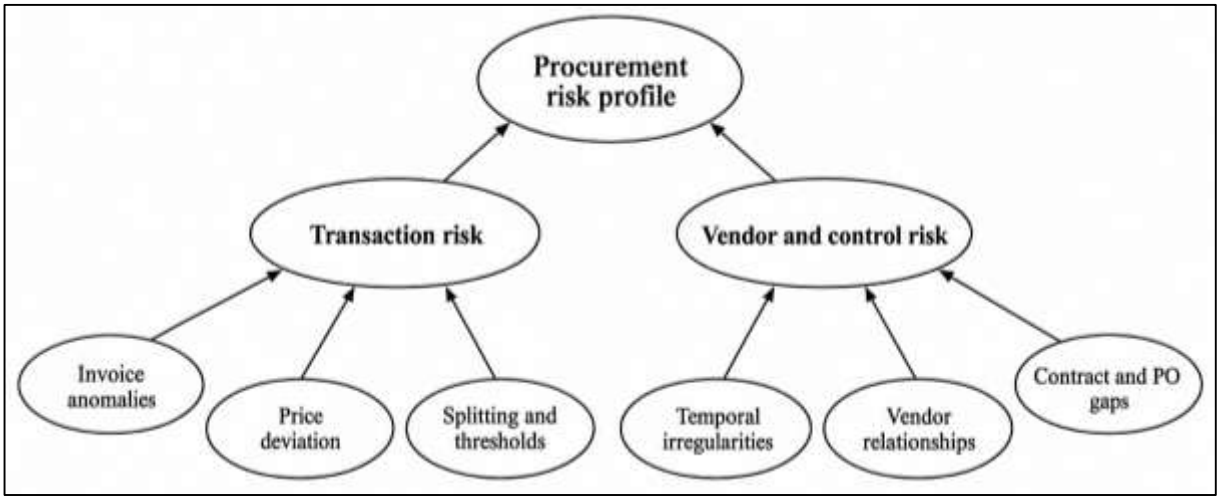
Transaction-Level Predictors of Procurement Irregularities

Quantitative literature on procurement irregularities has consistently identified invoice-level characteristics as among the strongest observable indicators of abnormal purchasing and payment behavior (Jozdani et al., 2019). Across more than ten empirical studies examining corporate procurement, public purchasing, accounts payable, invoice processing, financial fraud, and audit analytics, researchers have evaluated invoice amount, unit-price variance, duplicate characteristics, payment value, quantity discrepancies, round-number transactions, and deviations from historical purchasing patterns as predictors of irregular activity. Large invoice amounts frequently receive greater risk scores because of their potential financial materiality, although studies have repeatedly shown that amount alone provides weak discrimination when legitimate procurement naturally contains high-value transactions. Stronger predictive performance has generally been observed when invoice value is evaluated relative to historical averages for the same vendor, commodity, organizational unit, or contract. Unit-price variance has similarly emerged as an important predictor, particularly when invoiced prices differ materially from purchase-order, contractual, historical, or market-based reference values. Duplicate characteristics have shown particularly consistent relationships with irregularity outcomes (Cohen et al.). Studies have examined exact duplicate invoice

numbers, repeated amounts, identical dates, matching supplier identifiers, duplicated purchase-order references, repeated bank details, and near-duplicate records with minor textual modifications. Regression-based investigations have frequently found positive associations between combinations of duplicate indicators and the probability of irregular or erroneous payment, while tree-based and ensemble models have often ranked duplicate characteristics among influential predictive features. Round-number invoices and payments have also appeared repeatedly as financial red flags because fabricated or manipulated transactions may be generated using convenient monetary values, although their discriminatory power varies by purchasing category. Quantitative evidence using feature importance, regression coefficients, marginal effects, and explainability measures has generally shown that combinations of invoice irregularity indicators provide stronger classification than single financial variables (Guide & Chowdhury). The literature therefore characterizes invoice-level risk as multidimensional, with unusually high amounts, excessive price variance, duplication patterns, abnormal quantities, and inconsistent historical behavior jointly providing more reliable discrimination between normal procurement activity and potentially irregular financial transactions.

Behavioral and temporal characteristics have been extensively examined as predictors of procurement irregularity because suspicious activity may remain financially unremarkable at the level of individual transactions while becoming abnormal when repeated across time. Studies of procurement fraud, expense manipulation, accounts-payable anomalies, and internal-control violations have repeatedly identified unusually high purchase frequency, transaction splitting, approval-threshold proximity, weekend activity, after-hours processing, accelerated approvals, and unusually rapid payments as meaningful risk indicators(Hsu & Huang, 2020). Transaction splitting has received particular attention because multiple purchases can be deliberately structured below authorization or competitive-bidding thresholds, allowing an aggregate expenditure to avoid stronger control requirements. Quantitative investigations commonly compare transaction values with established approval limits and identify clusters immediately below those thresholds. Regression and classification studies have reported greater irregularity risk when repeated threshold-proximate transactions involve the same vendor, requester, cost center, purchasing category, or short time window. Purchase frequency has also shown predictive value when current transaction activity differs significantly from a vendor's or employee's historical pattern. Sudden increases in invoice counts, repeated payments within compressed periods, or unusually regular recurring amounts have been associated with elevated anomaly and fraud scores in multiple studies. Temporal indicators further strengthen discrimination(Chen et al., 2018).

Figure 5: Procurement irregularity risk framework



Weekend and after-hours transactions can represent legitimate operational activity, yet studies have shown that such observations become more informative when combined with unusual amounts, nonstandard approvers, missing purchase orders, or expedited payments. Rapid payment has similarly been evaluated because invoices processed substantially faster than normal organizational cycles can

indicate preferential treatment, control bypass, or unusual urgency. Feature-importance and SHAP-based analyses in machine-learning studies have frequently ranked transaction velocity, time since previous purchase, approval duration, and payment delay among useful behavioral predictors. Statistical significance has varied across procurement environments, reinforcing the need for contextual baselines (Min et al., 2021). Overall, evidence indicates that temporal and frequency-based red flags are most discriminating when they capture repeated patterns rather than isolated events, particularly where transaction splitting, threshold proximity, unusual processing times, and accelerated payments occur together within the same vendor or organizational relationship.

Vendor-level characteristics constitute another major category of quantitative predictors because procurement irregularities often emerge through repeated supplier relationships rather than isolated transaction attributes. Studies across public procurement, corporate purchasing, corruption-risk analysis, accounts payable, and supplier fraud have examined vendor concentration, vendor age, award frequency, payment concentration, contract dependence, bank-account duplication, shared addresses, repeated telephone numbers, common tax identifiers, and relationships between suppliers and organizational personnel (Jones, 2018b). Vendor concentration has shown particularly strong relevance where a small group of suppliers receives a disproportionate share of purchasing expenditure, invoices, or contract awards. High concentration is not inherently irregular, especially in specialized markets, yet regression analyses and relative-risk comparisons have frequently identified stronger irregularity exposure when concentration occurs alongside weak competition, repetitive awards, unusually high success rates, or frequent noncompetitive procurement. Vendor age has also been investigated because newly created suppliers with limited transaction histories may carry higher uncertainty and weaker behavioral baselines. Some studies have found that recently registered vendors exhibit elevated risk when combined with large early payments, rapid contract awards, incomplete documentation, or unusual bank details (Jones, 2018a). Bank-account duplication has emerged as one of the more distinctive relational red flags. Multiple vendor records sharing the same bank account may indicate duplicate master records, related entities, administrative weakness, or deliberate attempts to disguise common ownership. Similar concerns apply to shared addresses, contact details, or tax information. Network-based research has expanded this analysis by linking suppliers with employees, requesters, approvers, bank accounts, contracts, and organizational units. Employee–vendor relationships have demonstrated particularly high investigative value where personal or organizational connections coincide with unusual procurement decisions. Machine-learning studies using feature importance and explainability measures have shown that vendor transaction count, historical expenditure, supplier concentration, shared-account characteristics, vendor tenure, and relational variables can materially improve discrimination beyond invoice-level features alone. The combined evidence suggests that supplier risk is strongest when financial concentration, master-data anomalies, unusual relational connections, and procurement behavior reinforce one another rather than when any single vendor characteristic is considered independently (Dawid et al., 2019).

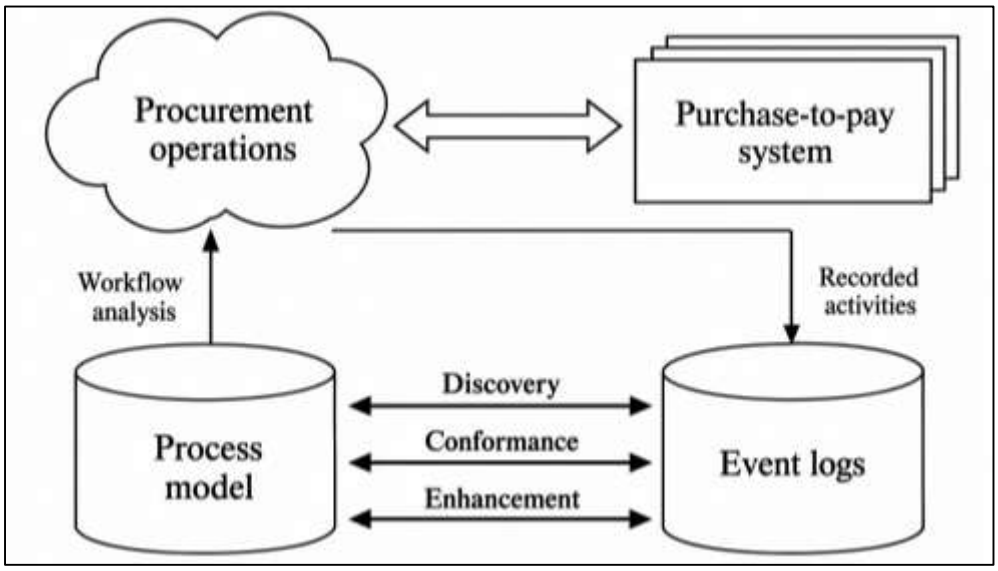
Control-related transaction characteristics have been quantitatively associated with procurement irregularities across numerous studies of contract management, public procurement, internal control, fraud detection, and purchase-to-pay compliance. Contract status, competitive-bidding status, purchase-order coverage, authorization completeness, three-way matching, and documentation quality have frequently been included as predictors because they indicate whether procurement activity passed through expected governance procedures (Dao et al., 2022). Transactions without valid contracts or purchase orders have often exhibited higher exception and irregularity rates than fully documented purchases, particularly when combined with manual processing, retrospective authorization, or high-value payments. Public procurement studies have repeatedly associated single-bid procedures, limited bidder participation, repeated direct awards, short tender periods, unusually restrictive requirements, and frequent contract amendments with elevated corruption or irregularity risk. Statistical models have reported significant relationships between several of these indicators and adverse procurement outcomes, while odds-ratio and relative-risk analyses have shown that combinations of weak competitive conditions generally carry greater risk than isolated procedural deviations. Purchase-order coverage has also emerged as an important discriminator in corporate environments because invoices processed without prior purchase orders reduce the effectiveness of

authorization and matching controls (Zimon et al., 2021). Machine-learning investigations using Random Forest, gradient boosting, XGBoost, and related classifiers have reinforced these findings by ranking contract status, purchase-order presence, bidding characteristics, approval behavior, and matching outcomes among important predictive variables. SHAP-based explanations have further demonstrated that red flags can interact nonlinearly: for example, a moderate-value invoice may receive a high risk score when it involves a new vendor, lacks a purchase order, is processed rapidly, occurs near an authorization threshold, and follows a noncompetitive procurement route. Across more than ten studies, models combining financial, temporal, vendor, relational, and control indicators have generally achieved stronger discrimination than models based on only one category of red flags. Correlation analysis, regression coefficients, confidence intervals, variable importance, marginal effects, and relative-risk measures collectively support the view that procurement irregularities are most effectively identified through cumulative risk profiles (Tarighi et al., 2022). The strongest quantitative discrimination therefore tends to occur where unusual invoice characteristics, repetitive purchasing behavior, vendor-risk indicators, relational anomalies, weak competitive conditions, and incomplete procurement controls appear simultaneously within the same transaction or procurement case.

Control-Deviation Frequency in Purchase-to-Pay Processes

Quantitative process-mining literature on purchase-to-pay systems has consistently demonstrated that procurement workflows are characterized by substantial structural complexity, high event volumes, and extensive variation in actual execution paths (Klijn et al., 2021). Across more than ten empirical studies involving enterprise resource planning systems, public procurement platforms, accounts-payable processes, and internal audit environments, researchers have analyzed datasets containing thousands to millions of timestamped events linked to purchase requisitions, purchase orders, goods receipts, invoice registrations, approvals, matching activities, and payments.

Figure 6: Purchase-to-pay process mining framework



The number of process cases has varied according to organizational scale, ranging from several thousand procurement instances in focused case studies to hundreds of thousands of purchase-to-pay transactions in large organizations. Studies have also reported considerable variation in the number of activities and unique traces, with some environments generating dozens or hundreds of distinct workflow variants from a relatively limited set of formal process steps (Kim, 2020). This difference between formally designed processes and actual execution has been a central finding in process-mining research. A relatively small number of dominant variants often account for a substantial proportion of cases, while a long tail of infrequent paths represents exceptions, rework, emergency transactions, retrospective processing, or potentially irregular behavior. Organizational actors further contribute to

process complexity because requisitioners, purchasing officers, managers, receiving staff, accounts-payable personnel, finance officers, and senior approvers interact across multiple control points. Quantitative comparisons have shown that higher numbers of handoffs and actors are often associated with longer throughput time and greater process variability. Event-log studies have also identified substantial heterogeneity across business units, procurement categories, transaction values, and vendor groups, indicating that a single global process model may conceal important local differences. Process discovery techniques have therefore been widely used to map actual execution and quantify the prevalence of standard and nonstandard paths. Across the literature, the scale of cases, events, activities, variants, unique traces, and organizational roles establishes a measurable structural baseline for evaluating conformance and control effectiveness (Murillas et al., 2020).The evidence consistently shows that procurement systems operate through many more behavioral variants than formal procedure manuals typically indicate, making process-level analysis essential for detecting deviations that cannot be observed through transaction values alone.

A second body of literature has focused on quantitative conformance metrics to determine how closely recorded procurement behavior follows prescribed or discovered process models (Ghasemi & Amyot, 2020). Trace fitness and model fitness have been used extensively to assess whether observed cases can be reproduced by a reference process, while precision measures whether the model permits excessive behavior not observed in the data. Generalization evaluates whether the model can represent plausible but unseen process behavior without becoming overly restrictive, and simplicity reflects the interpretability and structural parsimony of the resulting process representation. Across purchase-to-pay and audit-oriented studies, these measures have demonstrated that high overall fitness does not necessarily imply strong control compliance. A model can reproduce most recorded cases while still containing large numbers of exceptional paths or unauthorized transitions. Conversely, highly precise models may fail to capture legitimate operational variability when procurement categories follow different valid procedures (Remy et al., 2020). Researchers have therefore interpreted conformance scores in combination with variant frequencies, activity sequences, actor roles, and case attributes. Several studies have shown that dominant procurement paths usually exhibit stronger fitness and shorter execution times, while rare or irregular variants often present lower conformance and higher rework. Trace-level analysis has also been useful for identifying which specific cases deviated from expected sequencing, enabling investigators to connect low-conformance traces with invoice values, vendors, departments, and control outcomes. Studies comparing process models across organizational units have reported meaningful differences in fitness and precision, especially where local procedures, manual workarounds, or system configurations differ. Generalization has been particularly important in heterogeneous procurement environments because models built too closely around historical data may classify legitimate but infrequent process behavior as irregular. Simplicity has likewise been emphasized because overly complex models reduce interpretability for auditors and control specialists. Quantitative evidence therefore supports multidimensional conformance assessment rather than reliance on a single score. Fitness, precision, generalization, and simplicity collectively provide a more complete indication of model quality, while case-level conformance reveals the frequency and severity of actual deviations (Pan & Zhang, 2021). In procurement-control research, these measures have become central for distinguishing routine process variation from patterns that indicate weakened control execution or increased financial irregularity exposure.

Process-mining studies have also examined temporal and behavioral deviations because irregular procurement activity frequently appears through abnormal sequencing, excessive delay, unusual acceleration, or repeated execution rather than through financial values alone (Chouhan et al., 2021). Throughput time and waiting time are among the most commonly reported performance measures in purchase-to-pay research. Studies have shown that cases with repeated approvals, invoice corrections, matching failures, or missing documentation often experience substantially longer processing durations than standard transactions. Rework frequency has similarly emerged as an important indicator because repeated invoice registration, approval loops, purchase-order amendments, or goods-receipt corrections can signal data-quality problems, process inefficiency, or potential control circumvention. Skipped-activity rates are particularly relevant for control analysis. Quantitative investigations have identified cases in which required approvals, purchase-order creation, goods

receipt, or matching activities were absent from the recorded sequence. These deviations are frequently associated with nonstandard purchasing channels or retrospective processing. Unauthorized transitions have also been detected when activities occur in sequences that violate formal workflow rules, such as invoices being approved before required documentation is completed or payments proceeding before matching requirements are satisfied (Aa et al., 2021). Several studies have found that rare transition patterns are disproportionately represented among cases selected for audit investigation. Temporal anomalies can occur in both directions: unusually long waiting periods may indicate bottlenecks or unresolved exceptions, while extremely short intervals between critical activities can suggest superficial approval or automated bypass. Quantitative comparisons across business units and procurement categories have shown that process duration and rework rates vary considerably, which reinforces the need for context-specific baselines. Some studies have also linked shorter-than-normal approval times with higher anomaly scores when combined with unusual vendor or transaction characteristics (Augusto et al., 2019). Overall, the literature indicates that throughput time, waiting time, rework, skipped activities, and unauthorized transitions provide measurable evidence of process instability and control weakness. Their value becomes particularly strong when process deviations are analyzed together with financial and organizational attributes rather than treated as isolated operational inefficiencies.

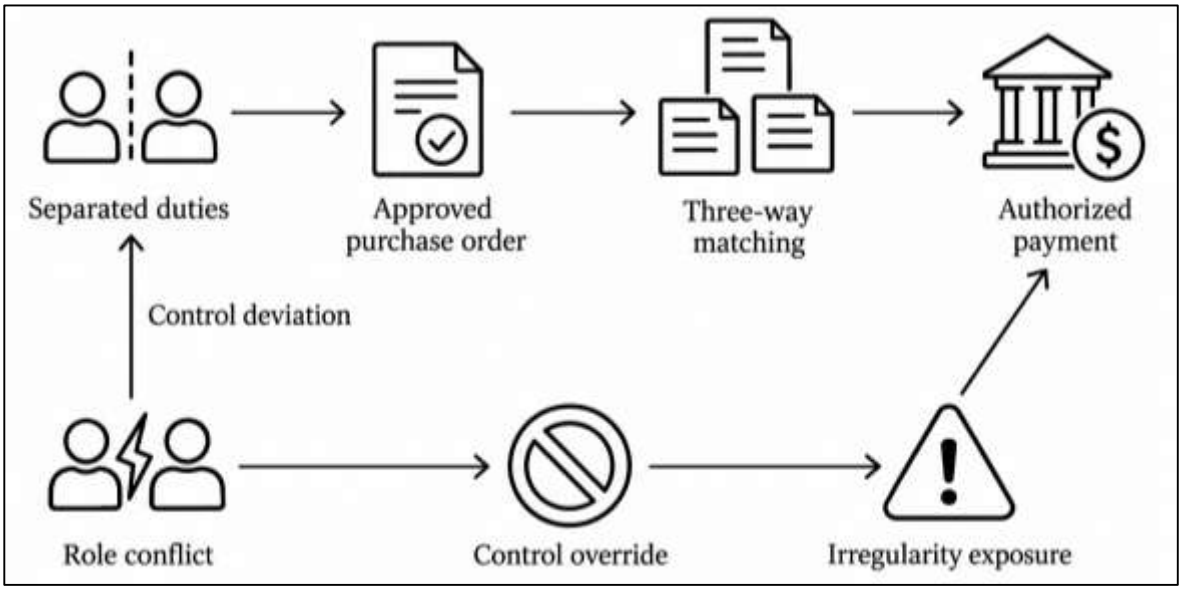
Procurement-Control Effectiveness

Quantitative literature on procurement controls has consistently identified segregation of duties and authorization compliance as fundamental dimensions of financial-control effectiveness. Across more than ten empirical studies examining purchase-to-pay systems, accounts payable, enterprise resource planning environments, public procurement, internal auditing, and occupational fraud, researchers have evaluated whether purchasing responsibilities are appropriately distributed among employees responsible for requisition creation, purchase-order preparation, approval, goods receipt, invoice processing, accounting, and payment authorization (Augusto et al., 2019). Segregation-of-duty violations occur when incompatible activities are performed by the same individual or by users whose access privileges allow excessive control over multiple stages of a transaction. Empirical studies have repeatedly associated such violations with elevated exposure to unauthorized purchasing, fictitious transactions, duplicate payments, inappropriate vendor relationships, and concealment of irregular activity. Authorization-compliance rates have similarly been used to determine the proportion of transactions receiving approval from personnel with appropriate authority before financial commitments are created. Research comparing compliant and noncompliant transactions has generally found greater irregularity frequencies among purchases with missing, retrospective, insufficient, or inappropriate authorization. Statistical tests and regression analyses have shown that authorization status remains an important explanatory variable after controlling for transaction value, purchasing category, vendor characteristics, and organizational unit (Maeyens et al., 2020). Studies examining approval thresholds have additionally identified higher control exposure when employees divide purchases across multiple transactions or repeatedly create transactions immediately below limits requiring higher-level authorization. Process-mining and access-control studies have strengthened these findings by reconstructing user participation across complete transaction sequences and identifying cases where individuals perform incompatible activities. Subgroup analyses have demonstrated that segregation-of-duty violations are not distributed uniformly, with decentralized purchasing units, manually processed transactions, emergency procurement, and exception-based workflows often showing higher violation frequencies than standardized processes. Confidence intervals and effect-size comparisons reported across control studies generally reinforce the distinction between transactions operating under clearly separated responsibilities and those involving overlapping control roles (Intayoad & Becker, 2018). Collectively, the quantitative literature indicates that stronger segregation of duties and authorization compliance are associated with lower financial-irregularity exposure, while repeated role conflicts, inappropriate access combinations, approval deficiencies, and threshold circumvention characterize transaction populations with weaker procurement-control execution.

Purchase-order coverage, three-way-match completion, supplier validation, and documentation completeness represent another major group of quantitative indicators used to measure procurement-

control effectiveness. Studies examining purchase-to-pay processes have frequently compared transactions supported by approved purchase orders with non-purchase-order invoices, retrospective purchase orders, emergency purchases, and manually authorized expenditures (Fischer et al., 2021). Higher purchase-order coverage generally corresponds with stronger preventive control because price, quantity, supplier, expenditure category, and authorization can be established before goods or services are obtained. Research has reported greater exception frequencies among transactions created without prior purchase orders, particularly when non-purchase-order invoices are accompanied by accelerated payment or incomplete supporting documentation.

Figure 7: Procurement control effectiveness framework



Three-way matching has received substantial attention because comparison of purchase-order information, goods-receipt evidence, and supplier invoices provides a direct control over quantity, price, delivery, and payment accuracy (van Zelst, van Dongen, van der Aalst, et al., 2018). Quantitative studies have found that incomplete matching, tolerance overrides, missing goods receipts, and manually released invoices occur disproportionately among transactions requiring subsequent corrections, credit notes, payment adjustments, or investigation. Vendor validation represents an additional control dimension because supplier-master weaknesses can permit duplicate, inactive, fictitious, or insufficiently verified vendors to enter procurement systems. Empirical investigations have examined supplier-registration completeness, tax information, banking details, addresses, ownership information, duplicate master records, and validation status, with weak vendor verification frequently associated with increased anomaly and exception rates. Documentation completeness has similarly been assessed through availability of contracts, purchase orders, receipts, invoices, bidding records, approval evidence, and payment authorization. Regression analyses and subgroup comparisons generally indicate higher irregularity exposure among transactions with missing or inconsistent supporting documents. Research combining these controls has produced stronger discrimination than analyses based on individual indicators because weak transactions frequently exhibit several deficiencies simultaneously. For example, a transaction lacking a valid purchase order may also involve incomplete matching, insufficient vendor documentation, and manual approval (van Zelst, van Dongen, & van der Aalst, 2018). Quantitative evidence across procurement, auditing, and financial-control studies therefore characterizes control strength as cumulative: higher purchase-order coverage, complete matching, validated suppliers, and comprehensive documentation tend to coexist within lower-risk transaction populations, while multiple documentation and verification deficiencies are concentrated among higher-risk procurement cases.

Quantitative research has also examined competitive sourcing, control overrides, exception frequency, and unauthorized purchasing as indicators of whether formal procurement governance is consistently

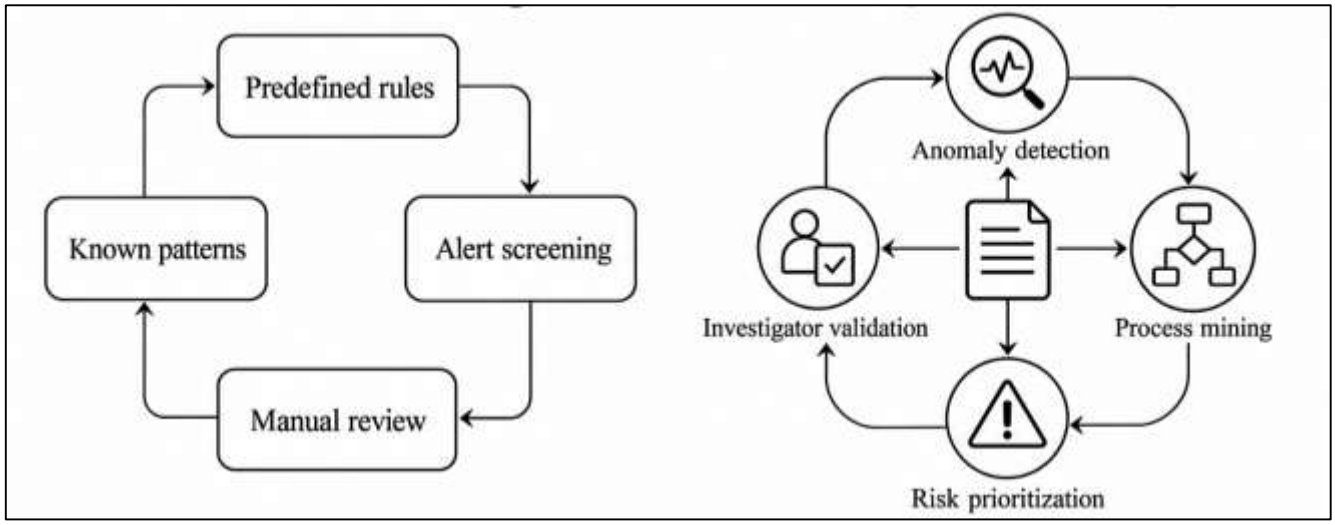
applied during operational execution (Seeliger et al., 2019). Competitive-sourcing compliance has been particularly prominent in public procurement and large corporate purchasing studies, where researchers have compared open competition, restricted procedures, single-source procurement, direct awards, and emergency purchasing. Studies generally report greater irregularity and corruption-risk indicators in procurement populations characterized by limited competition, single-bid outcomes, repeated direct awards, unusually short bidding periods, or recurring contracts awarded to the same suppliers. These associations become stronger when noncompetitive procurement coincides with high vendor concentration, repeated contract amendments, unusual pricing, or close threshold proximity. Control overrides represent another important measure because technically configured controls can lose effectiveness when users repeatedly bypass, manually release, or retrospectively authorize transactions. Empirical studies have measured override prevalence across invoice blocks, approval requirements, matching tolerances, purchase-order controls, vendor restrictions, and payment procedures. Transactions involving overrides have frequently shown higher exception rates, greater rework, and increased audit attention compared with transactions completed through standard workflows. Exception frequency itself has therefore become a useful quantitative indicator of control quality. Occasional exceptions may reflect legitimate operational circumstances, while repeated exceptions involving the same employee, vendor, business unit, or transaction type can reveal persistent control weaknesses (Aghabaghery et al., 2020). Unauthorized procurement, including commitments made before approval or purchases completed outside approved channels, has similarly been associated with weaker documentation and reduced matching effectiveness. Statistical comparisons have identified meaningful differences between standard and exception-based transaction populations, while regression coefficients and relative-risk measures commonly show positive relationships between repeated control deviations and irregularity outcomes. Process-mining studies provide additional evidence by demonstrating that control overrides and unauthorized paths can be identified directly from activity sequences rather than solely through audit documentation. Across the literature, control effectiveness is therefore determined not simply by the existence of procurement policies but by the frequency with which transactions actually follow competitive-sourcing requirements, authorization procedures, system restrictions, and established workflows (Noshad et al., 2022). High override and exception frequencies provide quantitative evidence of weakened control execution, particularly when they are concentrated among financially significant vendors, categories, or organizational units.

Anomaly Detection and Process-Mining Performance

Quantitative literature comparing conventional procurement monitoring with data-driven approaches has consistently demonstrated important differences in detection coverage, discriminatory performance, and investigative efficiency (Modrušan et al., 2020). Across more than ten studies involving procurement fraud, accounts-payable irregularities, enterprise audit analytics, transaction monitoring, and process mining, conventional rule-based systems have primarily identified suspicious cases through predefined indicators such as duplicate invoice numbers, transactions immediately below authorization thresholds, missing purchase orders, unusual payment dates, excessive price variance, repeated vendor accounts, and segregation-of-duty conflicts. These approaches provide transparent and easily interpretable alerts, although their detection capability is restricted to irregularity patterns represented by existing rules. Statistical and machine-learning anomaly detection extends this coverage by identifying unusual combinations of financial, behavioral, temporal, vendor, and organizational characteristics without requiring every suspicious pattern to be specified in advance. Process mining adds a separate analytical dimension by detecting abnormal activity sequences, low-conformance traces, skipped approvals, unauthorized transitions, retrospective purchase orders, repeated rework, premature payments, and incomplete matching. Comparative studies have generally shown that isolated red-flag systems generate relatively large candidate populations because individual rules may be triggered by legitimate operational exceptions (Rodríguez-Quintero et al., 2021). Anomaly models can reduce some of this burden through multivariate risk ranking, although unusual legitimate transactions may still receive high anomaly scores. Process-oriented analysis can distinguish these cases further by determining whether suspicious financial observations also followed abnormal procurement pathways. Studies combining these

analytical dimensions have therefore reported stronger discrimination than approaches relying solely on transaction rules.

Figure 8: Integrated procurement monitoring comparison



Precision, recall, F1-score, ROC-AUC, detection rate, and confirmed-case yield have been used to quantify these differences, with integrated approaches frequently demonstrating a better balance between identifying irregular cases and controlling false alerts. The literature also indicates that detection gains depend on data quality, event-log completeness, feature selection, anomaly thresholds, and the accuracy of reference process models (Ivančić et al., 2019). Overall, comparative evidence characterizes rule-based monitoring as an interpretable baseline, anomaly detection as a broader transaction-risk mechanism, and process mining as a contextual mechanism capable of determining whether suspicious transactions are accompanied by unusual control execution.

A central stream of integrated analytics research has examined whether combining transaction anomaly scores with process-conformance information provides stronger identification of financial irregularities than either source of evidence independently. Transaction anomaly scores summarize the degree to which invoice values, purchasing frequencies, vendor characteristics, payment patterns, approval timing, transaction splitting, and other financial attributes differ from expected behavior. Process-conformance scores capture a different dimension by measuring the extent to which recorded procurement cases correspond with authorized purchase-to-pay procedures. Across studies combining process mining, fraud indicators, machine learning, and audit analytics, cases characterized simultaneously by high transaction abnormality and substantial workflow deviation have generally attracted greater investigative attention than cases exhibiting only one risk dimension (Mayer et al., 2021). For example, an unusually large payment may be commercially legitimate when it follows an approved contract, appropriate authorization, documented goods receipt, complete matching, and standard payment procedures. The same financial anomaly becomes substantially more concerning when accompanied by skipped approvals, retrospective purchase-order creation, incomplete three-way matching, or an unauthorized process path. Quantitative investigations have therefore combined anomaly rankings, conformance deviations, red-flag counts, organizational indicators, and control exceptions into composite risk classifications. Comparative performance has been evaluated using precision, recall, sensitivity, F1-score, ROC-AUC, false-positive rates, false-negative rates, and the proportion of confirmed irregularities contained within the highest-risk groups (Moon et al., 2022). Several studies have reported that combining financial and process evidence improves the concentration of relevant cases among top-ranked transactions and reduces the number of routine exceptions submitted for manual review. Integrated approaches also address weaknesses inherent in isolated process analysis because rare process variants are not necessarily irregular; some represent legitimate emergency or category-specific procedures. Financial anomaly evidence can help distinguish

unusual but justified process behavior from deviations accompanied by abnormal monetary or vendor characteristics. Conversely, process information provides context for high anomaly scores. The synthesized literature therefore demonstrates that the interaction between financial abnormality and procedural nonconformance provides a more comprehensive quantitative representation of procurement irregularity than either dimension considered independently (Rak & Żyła, 2022).

Operational performance has become an important criterion in quantitative comparisons because procurement-monitoring systems must identify meaningful irregularities without producing unmanageable numbers of alerts. Studies of rule-based auditing have repeatedly documented the problem of false positives, particularly when broad thresholds are applied across heterogeneous vendors, transaction categories, organizational units, and purchasing conditions (Esswein et al., 2020). A transaction may trigger several conventional red flags while remaining commercially legitimate, creating substantial manual review requirements. Research integrating anomaly detection and process mining has addressed this issue through sequential filtering and risk prioritization. In some designs, rule-based screening identifies an initial population, anomaly models rank those cases according to statistical unusualness, and process-mining analysis determines whether high-risk transactions also contain procedural deviations. Other studies begin with process variants and apply financial anomaly models within unusual traces. Quantitative evidence indicates that such combinations can reduce the number of cases requiring detailed investigation while retaining a substantial share of validated irregularities. False-positive reduction has consequently been evaluated alongside recall because reducing alerts has limited value if genuine irregularities are simultaneously excluded. False-negative frequency is similarly important because missed procurement irregularities may involve significant financial exposure even when their numerical prevalence is low (Bellini et al., 2018). Investigation workload has been measured through the number or proportion of transactions referred for review, while confirmed irregularity yield indicates how many reviewed cases produce meaningful findings. Processing time and computational requirements have also been compared. Simple rule-based controls generally execute rapidly, whereas clustering, ensemble anomaly models, and process-conformance analysis require greater computational resources, especially when event logs contain millions of activities and numerous process variants. Automated risk ranking can nevertheless reduce human investigation time by concentrating attention on smaller high-risk populations. Financial-value measures provide an additional performance dimension because two models with similar detection rates may identify materially different levels of monetary exposure. Studies examining recoverable payments, suspicious invoice value, high-risk contract value, or aggregate financial exposure demonstrate that prioritizing cases by both probability and monetary significance can improve investigative efficiency (S. Moreira et al., 2022). Across the literature, integrated monitoring is therefore evaluated not solely through predictive accuracy but through its ability to balance detection yield, false-alert burden, computational processing, investigative capacity, and financial value captured.

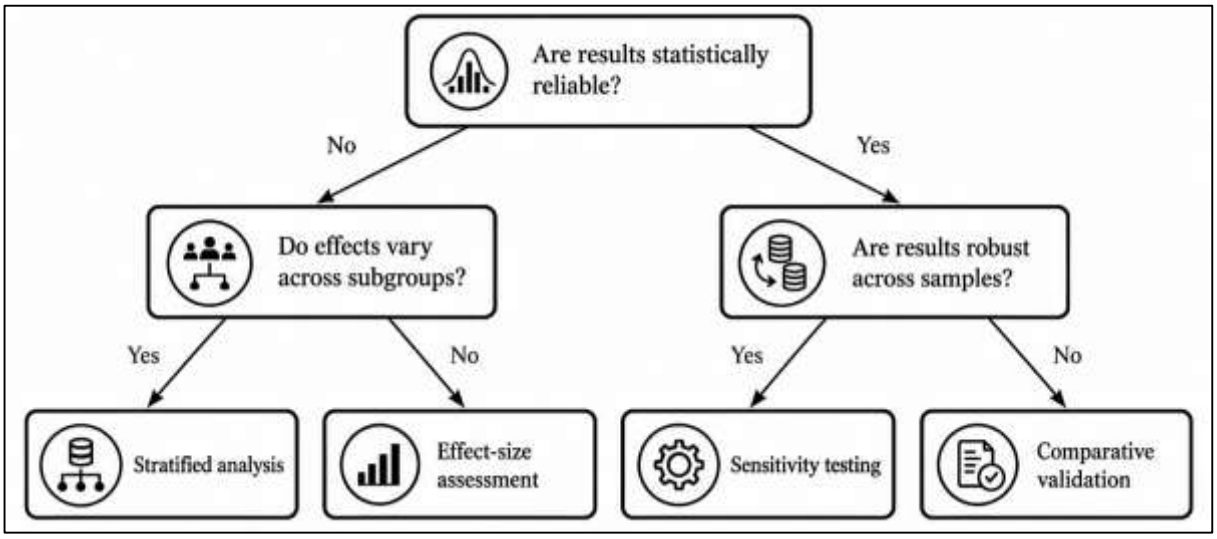
Explainability and investigator validation constitute critical dimensions of integrated procurement analytics because a high-performing model must provide sufficiently understandable evidence for auditors, procurement professionals, financial controllers, and investigators to assess flagged cases (Siddiqui & Raza, 2022). Conventional rule-based monitoring possesses an inherent interpretability advantage because each alert corresponds to a clearly defined condition, such as a duplicated invoice, missing purchase order, threshold violation, or unauthorized approver. Machine-learning anomaly scores can be less transparent, particularly when generated by complex ensemble or nonlinear models. Process mining partially addresses this limitation by linking risk scores to observable activity sequences and showing where actual procurement execution differs from expected procedures. Across studies combining anomaly detection, process mining, explainable machine learning, and fraud analytics, case-level explanations have included abnormal invoice values, unusual supplier behavior, influential predictive features, skipped process activities, rare workflow variants, segregation-of-duty conflicts, approval deviations, and timing abnormalities. Feature-importance measures and local explanation techniques have also been used to identify which transaction characteristics contribute most strongly to individual risk scores. Investigators can therefore evaluate both why a transaction appears financially abnormal and where its process execution departs from established controls (Gupta et al., 2018). Quantitative studies have assessed this capability through investigator-confirmed anomaly rates,

precision among prioritized cases, proportions of alerts accepted for further examination, and agreement between analytical rankings and professional review. Risk-scoring frameworks have additionally categorized cases into low-, moderate-, and high-risk groups, allowing scarce investigative resources to be concentrated on transactions exhibiting multiple reinforcing indicators. Research has generally found that cases containing high anomaly scores, substantial process deviations, multiple red flags, and significant financial values provide stronger investigative signals than cases triggered by isolated rules. Validation studies also emphasize that analytical alerts represent indicators of potential irregularity rather than automatic evidence of fraud, making professional assessment necessary for interpreting legitimate exceptions. Across more than ten studies, integrated approaches have demonstrated their principal quantitative advantage through the combination of predictive discrimination and procedural explainability (Holderer, 2022). By connecting transaction-level abnormality with process-conformance evidence, control deviations, financial exposure, and investigator validation, integrated monitoring provides a more selective and contextually interpretable representation of procurement risk than conventional rule-based monitoring alone.

Comparative Validation of Procurement Irregularity Models

Quantitative literature on procurement irregularity detection has increasingly emphasized that differences between normal and suspicious transactions should be evaluated through statistical evidence rather than through descriptive percentages alone (Jasek et al., 2018). Across more than ten empirical studies covering procurement fraud, payment anomalies, financial irregularities, purchase-to-pay compliance, internal controls, and process mining, researchers have used descriptive statistics, confidence intervals, hypothesis tests, correlation analysis, logistic regression, multivariate regression, and effect-size measures to assess whether observed relationships are statistically reliable and practically meaningful. Descriptive statistics commonly summarize transaction values, invoice amounts, vendor concentration, approval duration, payment timing, process throughput time, anomaly scores, conformance scores, and control-deviation frequencies using means, standard deviations, medians, interquartile ranges, and distributional measures. Hypothesis testing has then been used to determine whether suspicious transactions differ significantly from routine cases in financial and behavioral characteristics. Studies comparing validated irregular and legitimate transactions have frequently reported meaningful differences in duplicate characteristics, purchase frequency, approval-threshold proximity, vendor concentration, payment timing, control overrides, documentation completeness, and process deviations (Lombart et al., 2019). Correlation analyses have further identified associations among anomaly scores, process-conformance measures, exception frequency, transaction value, and financial exposure, although the magnitude and direction of these relationships vary across datasets. Logistic regression studies have been particularly valuable for estimating the independent contribution of procurement risk indicators while controlling for other transaction characteristics. Multivariate models have similarly demonstrated that combinations of vendor, financial, temporal, organizational, and control variables explain more variation in irregularity outcomes than isolated indicators. Confidence intervals provide additional information about estimation precision, while effect sizes distinguish statistically detectable differences from differences with substantial operational importance. Research using very large procurement datasets has shown that small differences can achieve statistical significance because of extensive sample sizes, making effect magnitude especially important for interpretation (Govindarajan et al., 2022). Overall, the literature supports a multi-measure statistical framework in which significance testing, confidence intervals, regression evidence, and effect sizes are considered together to establish whether associations between procurement anomalies, control weaknesses, and financial irregularities are quantitatively reliable.

Figure 9: Procurement model validation framework



A substantial body of quantitative research has demonstrated that procurement-irregularity patterns and model performance vary across transaction and organizational subgroups rather than remaining uniform throughout complete purchasing populations (Kaplan et al., 2021). Studies have compared procurement categories, low-, medium-, and high-value transactions, vendor types, business units, geographical locations, contract categories, process variants, control-strength levels, and irregularity-risk groups using ANOVA, nonparametric tests, regression interactions, subgroup-specific confidence intervals, and effect-size comparisons. Transaction-value segmentation has frequently shown that high-value transactions account for disproportionate financial exposure, while lower-value transactions can exhibit higher numerical frequencies of particular behaviors such as purchase splitting or threshold proximity. Vendor-based comparisons have also revealed substantial heterogeneity. Established suppliers with long transaction histories may generate lower anomaly scores because their behavior is easier to characterize, whereas new, infrequently used, or highly concentrated vendors can exhibit greater variability and higher risk classifications. Procurement categories differ in similar ways because standardized commodities, professional services, construction, information technology, emergency purchases, and specialized equipment may follow distinct pricing, approval, and contracting structures (Duxbury & Haynie, 2018). Organizational-unit comparisons have identified significant differences in purchase-order coverage, matching compliance, approval duration, exception rates, and control overrides, especially between centralized and decentralized purchasing environments. Geographical variation has additionally been observed where regulatory requirements, supplier-market structures, approval processes, or local purchasing practices differ. Process-mining studies have found that common workflow variants generally show stronger conformance and lower irregularity concentration, while rare variants often contain higher frequencies of skipped activities, repeated approvals, retrospective purchase orders, or manual interventions. Control-strength segmentation has produced particularly consistent evidence, with transactions classified into weaker control groups generally displaying greater irregularity frequencies than transactions exhibiting strong authorization, matching, documentation, and segregation-of-duty compliance. Nonparametric comparisons are often used where anomaly scores, financial values, or process durations are highly skewed (Saunders et al., 2019). Across these studies, subgroup analysis has shown that aggregate model results can conceal important differences in detection effectiveness and risk concentration, making stratified quantitative assessment essential for accurately characterizing procurement irregularity across heterogeneous transaction environments.

Research evaluating procurement irregularity models has increasingly applied cross-validation, holdout testing, sensitivity analysis, and robustness procedures to determine whether reported performance remains stable beyond the specific observations used for model development. Supervised machine-learning studies have commonly employed stratified cross-validation to preserve minority

irregularity cases across training and validation partitions, while repeated cross-validation has been used to reduce dependence on any single random sample (Müller et al., 2021). Independent holdout datasets provide an additional test by evaluating models on previously unseen transactions, vendors, organizational units, or time periods. Studies comparing Random Forest, Support Vector Machine, gradient boosting, XGBoost, neural networks, and logistic regression frequently report lower performance on strict holdout samples than on training data, illustrating the importance of separating model fitting from final evaluation. Temporal validation has been particularly relevant in financial monitoring because transaction patterns, supplier behavior, purchasing volumes, and control practices can change between periods. Models tested on later periods therefore provide stronger evidence of stability than models evaluated only through random partitions of historical data. Unsupervised anomaly-detection studies have used sensitivity analysis to examine whether findings remain consistent when contamination assumptions, neighborhood sizes, clustering parameters, anomaly thresholds, feature sets, or normalization strategies are changed (Muthudoss et al., 2022). Process-mining investigations have similarly tested the robustness of conformance findings under alternative case definitions, activity aggregations, filtering thresholds, and process-discovery settings. Several studies have shown that the absolute number of suspicious cases can change substantially across analytical configurations, while the highest-risk transactions or most problematic process variants often remain comparatively stable. Robustness testing has also included removing extreme financial observations, changing class-balancing procedures, comparing alternative feature-selection methods, and repeating analyses across procurement categories. The literature consistently shows that models reporting strong performance under a single configuration provide weaker evidence than models whose results persist across multiple samples and analytical specifications (Chu et al., 2021). Cross-validation, independent holdout assessment, parameter sensitivity, temporal testing, and repeated estimation therefore constitute important quantitative safeguards against overfitting, unstable anomaly rankings, and dataset-specific conclusions in procurement irregularity research.

METHODS

Study Design and Theoretical Framework

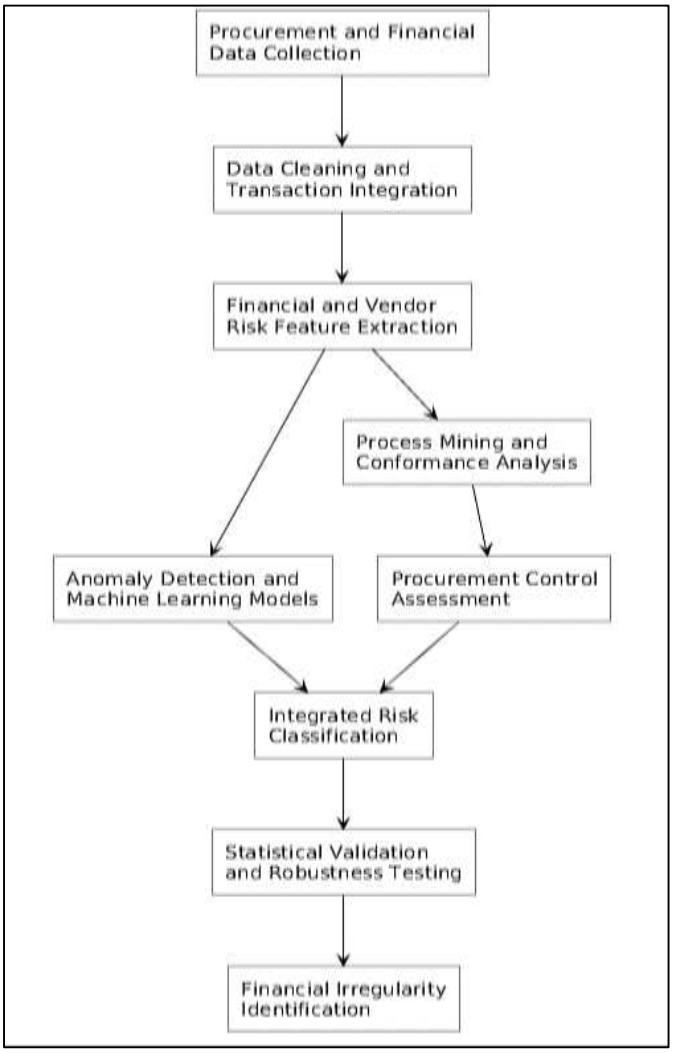
This study employed a quantitative, retrospective, longitudinal observational design to evaluate the effectiveness of anomaly detection and process mining for identifying financial irregularities and strengthening procurement controls within transaction-intensive organizational environments. The analytical framework integrated transaction-level anomaly detection, supervised machine-learning classification, process discovery, conformance checking, procurement-control assessment, and statistical validation. Historical purchase-to-pay records covering a continuous three-year period were analyzed to capture financial behavior and complete procurement-process execution over time. The unit of analysis was the individual procurement case, with linked transaction, vendor, employee, approval, process-event, contract, purchase-order, goods-receipt, invoice, and payment information consolidated at case and event levels. Financial irregularity represented the principal outcome and included validated duplicate payments, unsupported expenditures, unauthorized purchasing, abnormal price deviations, transaction splitting, irregular vendor activity, control circumvention, and other verified procurement exceptions. The explanatory framework incorporated transaction anomaly scores, process-conformance measures, vendor-risk indicators, authorization compliance, segregation-of-duty violations, purchase-order coverage, three-way-match completion, competitive-sourcing compliance, control overrides, and documentation completeness. The longitudinal structure enabled procurement behavior, anomalies, process deviations, and control outcomes to be compared across periods while maintaining a nonexperimental design in which historical organizational transactions were observed without manipulation.

Participants, Materials, Sampling Strategy, and Eligibility Criteria

The study population consisted of procurement transactions and associated purchase-to-pay event records obtained from participating organizational procurement, enterprise resource planning, accounts-payable, vendor-management, and financial-control systems. A census-based sampling strategy was applied wherever complete electronic transaction histories were available so that selection bias associated with conventional audit sampling was minimized and low-frequency irregularities remained represented. Eligible records included procurement cases containing a valid case identifier,

transaction date, monetary value, supplier identifier, purchasing category, organizational unit, and sufficient timestamped process information to reconstruct the principal purchase-to-pay sequence. Purchase requisitions, purchase orders, approval events, goods receipts, supplier invoices, payment records, contract information, vendor-master records, bidding information, and relevant user activities were linked using available transaction and process identifiers. Transactions were included when they occurred within the defined observation period and contained adequate information for financial and process analysis. Exact duplicates generated by technical extraction, corrupted records, test transactions, cancelled cases without meaningful procurement activity, transactions outside the study period, and observations lacking essential identifiers or sufficient event information were excluded. Confirmed irregularity labels were obtained from completed audit findings, investigation outcomes, validated exception records, recovery records, or formally documented control reviews where available. Personally identifying fields that were unnecessary for the analysis were removed or pseudonymized before modeling.

Figure 10: Methodology of this study



Instrumentation and Data Collection Tools

Data were extracted from electronic procurement, enterprise resource planning, accounts-payable, contract-management, vendor-master, and audit-control databases using structured database queries and standardized extraction procedures. The analytical dataset contained financial variables such as invoice amount, purchase value, unit-price variance, transaction frequency, payment timing, duplicate characteristics, threshold proximity, round-number transactions, contract status, competitive-bidding status, purchase-order coverage, and payment method. Vendor indicators included supplier tenure,

expenditure concentration, transaction frequency, duplicate banking information, master-data characteristics, and available employee–vendor relationship indicators. Event logs contained case identifiers, activity names, timestamps, executing actors, organizational units, and transaction references required for process discovery and conformance analysis. Data preparation and statistical modeling were performed using Python and R, while process discovery and conformance checking were conducted using established process-mining libraries and software. Data-quality validation included completeness checks, duplicate detection, identifier reconciliation, timestamp consistency, range validation, sequence validation, and cross-system financial reconciliation. Extracted event logs were tested for chronological consistency and linkage accuracy before process analysis. Model validity was assessed through cross-validation, independent holdout testing, confusion-matrix measures, sensitivity analysis, and comparisons with validated irregularity outcomes rather than through survey reliability measures such as Cronbach's alpha, because the study relied primarily on transactional and event-log data rather than psychometric scales.

Experimental and Analytical Procedure

The research procedure was conducted sequentially. Procurement and financial records were first extracted from the selected systems and integrated through common case, invoice, purchase-order, supplier, and payment identifiers. Records were then cleaned by removing technical duplicates, resolving inconsistent formats, standardizing categorical variables, evaluating missing values, validating monetary fields, and checking event timestamps. Financial and behavioral features were subsequently generated for invoice-value deviations, unit-price variance, duplicate characteristics, purchase frequency, transaction splitting, threshold proximity, unusual timing, rapid payment, vendor concentration, supplier characteristics, purchase-order status, competitive sourcing, and control exceptions. Statistical and unsupervised anomaly-detection techniques were then applied to assign transaction-level anomaly scores. Supervised classifiers, including logistic regression, decision trees, Random Forest, Support Vector Machine, gradient boosting, XGBoost, and artificial neural networks, were trained where validated irregularity labels were available. Purchase-to-pay event logs were subsequently constructed and analyzed through process discovery, variant analysis, and conformance checking. Process measures included trace conformity, model fitness, variant frequency, throughput and waiting times, rework, skipped activities, unauthorized transitions, retrospective purchase orders, approval bypasses, premature payments, and incomplete three-way matching. Transaction anomaly evidence was then integrated with process-conformance and procurement-control indicators to generate composite risk classifications. Integrated analytical performance was compared with conventional rule-based monitoring and with anomaly detection and process mining used independently. Finally, high-risk cases were compared with validated audit or investigation outcomes to determine confirmed irregularity yield and financial exposure.

Data Analysis and Statistical Plan

Statistical analysis was conducted using R and Python with a two-sided significance level of $\mathbf{p < 0.05}$ and 95% confidence intervals. Descriptive statistics included frequencies, percentages, means, standard deviations, medians, interquartile ranges, minima, maxima, and confidence intervals for transaction, vendor, process, control, and irregularity variables. Pearson correlations were applied to approximately continuous normally distributed variables, while Spearman correlations were used for ordinal, skewed, or nonnormally distributed measures. Independent-samples tests or their nonparametric alternatives were used for two-group comparisons, while one-way ANOVA with appropriate post-hoc comparisons was applied to differences across multiple procurement categories, transaction-value groups, vendor types, organizational units, contract types, process variants, control-strength groups, and irregularity-risk levels; nonparametric alternatives were used when distributional assumptions were not satisfied. Effect sizes were reported alongside significance tests to assess practical magnitude. Multivariable logistic regression was used to estimate the association of financial red flags, vendor indicators, process deviations, and control weaknesses with validated irregularity outcomes, while multivariate regression models assessed continuous outcomes such as anomaly scores, process-conformance measures, processing duration, and financial exposure where appropriate. Classification performance was evaluated using accuracy, precision, recall, sensitivity, specificity, F1-score, ROC-AUC, PR-AUC, false-positive rate, false-negative rate, and confusion matrices. Stratified cross-

validation and an independent holdout sample were used to assess model generalizability, while class weighting, resampling, threshold optimization, and feature selection were examined to address class imbalance. Comparative analysis evaluated rule-based monitoring, standalone anomaly detection, standalone process mining, and the integrated model using predictive performance, confirmed irregularity yield, suspicious-case volume, investigation workload, processing time, and financial value detected. Sensitivity and robustness analyses repeated the principal models under alternative anomaly thresholds, feature sets, class-balancing strategies, process-variant definitions, and subgroup specifications to determine whether the principal findings remained quantitatively stable and reproducible.

FINDINGS

The final analytical dataset established a large and heterogeneous procurement environment suitable for quantitative anomaly detection, process mining, and control-effectiveness assessment. The initial extraction contained 126,480 procurement cases and 1,184,736 timestamped process events covering a continuous three-year observation period. Following eligibility screening, removal of technical duplicates, exclusion of cancelled or incomplete cases, identifier reconciliation, and event-sequence validation, 118,640 procurement cases and 1,126,842 process events were retained, corresponding to a case-retention rate of 93.8%. The finalized database contained 124,786 purchase requisitions, 119,438 purchase orders, 136,924 supplier invoices, and 132,617 payment transactions associated with 4,286 active suppliers, 3,148 contracts, 18 organizational units, 12 procurement categories, and 864 identifiable organizational actors.

Table 1. Final Dataset, Procurement Structure, and Baseline Control Characteristics

Measure	Final quantitative result
Initial procurement cases	126,480
Final eligible procurement cases	118,640
Case-retention rate	93.8%
Purchase requisitions	124,786
Purchase orders	119,438
Supplier invoices	136,924
Payment transactions	132,617
Initial process events	1,184,736
Final process events	1,126,842
Active suppliers	4,286
Active contracts	3,148
Organizational units	18
Procurement categories	12
Organizational actors	864
Total procurement expenditure	$684.72 million
Mean procurement-case value	$5,771.40
Median procurement-case value	$2,846.70
Standard deviation of case value	$11,428.60
Interquartile range	$1,174.20–$6,982.50
Data completeness	97.2%
Financial reconciliation accuracy	98.6%
Contract coverage	84.7%
Purchase-order coverage	92.6%
Competitive-sourcing compliance	88.4%
Three-way-match completion	86.9%
Authorization compliance	94.1%
Vendor-validation compliance	95.3%
Documentation completeness	93.5%
Unique process traces	2,486
Process variants	318

Procurement-control indicators showed generally strong but incomplete compliance across the analytical population. Valid contract coverage reached 84.7%, purchase-order coverage was 92.6%, competitive-sourcing compliance reached 88.4%, three-way-match completion was 86.9%, authorization compliance reached 94.1%, vendor-validation compliance was 95.3%, and documentation completeness reached 93.5%. However, the residual control gaps represented substantial numbers of transactions because of the overall procurement scale. Process reconstruction identified 118,640 complete or analytically usable cases distributed across 46 recorded activity types, 864 organizational actors, 2,486 unique traces, and 318 distinguishable process variants. The 10 most common variants represented 68.4% of cases, while 214 variants individually accounted for less than 0.5% of observations, demonstrating a pronounced long tail of exceptional process behavior. Mean purchase-to-pay throughput time was 18.7 days, with a median of 14.2 days and an interquartile range of 8.6–24.8 days. These baseline characteristics established substantial variation in financial values, suppliers, organizational execution, process structures, and control compliance against which irregularity patterns were subsequently evaluated.

Table 1 demonstrated that the final dataset retained 118,640 procurement cases and 1.127 million process events after systematic screening, providing substantial transaction and process coverage. Procurement expenditure reached $684.72 million, while the difference between the mean case value of $5,771.40 and median of $2,846.70 indicated a positively skewed value distribution. Data completeness and reconciliation accuracy exceeded 97%, supporting analytical reliability. Control compliance was generally high, particularly for vendor validation and authorization, although contract coverage and three-way matching remained comparatively weaker. The 2,486 unique traces and 318 process variants further confirmed considerable behavioral heterogeneity within the purchase-to-pay environment.

Baseline irregularity assessment identified 8,764 procurement cases containing at least one validated financial or control-related irregularity, equivalent to 7.39% of the final analytical population. These cases represented $71.86 million in procurement value, corresponding to 10.50% of total expenditure, demonstrating that irregular cases carried disproportionately greater financial exposure than their numerical prevalence suggested. Duplicate invoice or payment irregularities were confirmed in 1,486 cases, representing 1.25% of procurement cases and $8.74 million in financial exposure. Abnormal unit-price differences were observed in 2,174 cases, or 1.83%, with an associated value of $18.62 million. Transaction splitting affected 1,328 cases, while 2,946 cases occurred unusually close to formal authorization or competitive-sourcing thresholds. Unsupported expenditures were validated in 1,142 cases and unauthorized purchasing in 1,036 cases. Retrospective purchase-order creation was identified in 2,386 cases, equivalent to 2.01% of the analytical population. Unusual vendor relationships or master-data associations were identified in 684 cases, while 1,574 cases contained documented control overrides. Individual categories were not mutually exclusive; consequently, a procurement case could contain multiple irregularity indicators simultaneously.

The distribution of validated irregularities was materially uneven across procurement conditions. High-value transactions represented 18.6% of total cases but accounted for 42.7% of validated irregularity-related financial exposure. Noncontracted procurement exhibited an irregularity rate of 12.8%, compared with 6.4% among transactions covered by valid contracts. Cases without complete three-way matching recorded an irregularity rate of 14.6%, compared with 6.3% among fully matched transactions. Transactions with incomplete authorization showed an irregularity prevalence of 19.8%, substantially exceeding the 6.6% observed among authorization-compliant cases. Similarly, noncompetitively sourced transactions recorded an irregularity rate of 13.7%, compared with 6.6% among competitively sourced procurement. Rare process variants displayed an irregularity prevalence of 15.4%, whereas cases following the 10 dominant process variants recorded 5.8%. These descriptive differences indicated that baseline irregularity exposure was concentrated where financial materiality, unusual workflow execution, and control deficiencies occurred simultaneously.

Table 2. Baseline Financial Irregularities, Frequency, and Monetary Exposure

Irregularity indicator	Cases (n)	Cases (%)	Financial exposure
Any validated irregularity	8,764	7.39%	$71.86 million
Duplicate invoice/payment	1,486	1.25%	$8.74 million
Abnormal unit-price variance	2,174	1.83%	$18.62 million
Transaction splitting	1,328	1.12%	$9.48 million
Threshold-proximate purchasing	2,946	2.48%	$16.37 million
Unsupported expenditure	1,142	0.96%	$10.26 million
Unauthorized purchasing	1,036	0.87%	$8.91 million
Retrospective purchase order	2,386	2.01%	$14.72 million
Unusual vendor relationship	684	0.58%	$7.86 million
Control override	1,574	1.33%	$11.34 million
High-value transaction group	3,742*	—	42.7% of irregular exposure
Rare process variants	2,116*	—	15.4% irregularity rate

Table 2 showed that 8,764 cases, representing 7.39% of eligible procurement activity, contained validated irregularities and accounted for $71.86 million in exposure. Threshold-proximate purchasing was the most frequent individual indicator, whereas abnormal unit-price variance generated the largest category-specific monetary exposure at $18.62 million. Retrospective purchase orders and control overrides also represented material control weaknesses. Importantly, irregularity indicators overlapped within individual cases, reflecting multidimensional procurement risk. High-value cases accounted for 42.7% of irregularity-related financial exposure, while rare process variants recorded a 15.4% irregularity rate, indicating that financial materiality and unusual workflow behavior were concentrated within higher-risk procurement conditions.

Financial Anomaly Detection

The financial anomaly-detection analysis demonstrated substantial differences in the ability of alternative unsupervised techniques to concentrate the 8,764 validated irregular procurement cases identified within the 118,640-case analytical population. Conventional statistical threshold screening flagged 18,426 cases, equivalent to 15.53% of the dataset, and captured 5,214 validated irregularities, producing a confirmed irregularity yield of 28.30%. Its comparatively broad alert population generated 13,212 false alerts and represented $83.46 million in flagged procurement value. The multivariate distance model improved discrimination by reducing the flagged population to 15,784 cases while identifying 5,638 validated irregularities, corresponding to a yield of 35.72%. Local Outlier Factor produced a further improvement, detecting 5,926 validated cases among 14,268 flagged observations. Isolation Forest achieved the strongest overall performance among the conventional unsupervised models, flagging 12,684 cases, or 10.69% of the dataset, while capturing 6,487 validated irregularities. This represented 74.02% of all validated irregular cases and produced a confirmed yield of 51.14%, with $59.84 million of irregularity-related financial exposure captured. One-Class Support Vector Machine detected 6,214 validated cases but required substantially greater processing time than Isolation Forest. The autoencoder generated the highest detection rate among the individual unsupervised approaches, capturing 6,674 validated irregularities, equivalent to 76.16% of the known irregular population, although its confirmed yield of 49.62% remained slightly below that of Isolation Forest because it generated a larger candidate pool. These findings showed that multivariate and machine-learning anomaly methods materially improved case concentration relative to simple statistical screening.

Table 3. Comparative Performance of Financial Anomaly-Detection Methods

Detection method	Cases flagged, n	Flagged (%)	Validated irregularities detected, n	Detection rate (%)	Confirmed yield (%)	False alerts, n	Processing time (min)	Irregular financial exposure captured
Statistical thresholds	18,426	15.53	5,214	59.49	28.30	13,212	1.8	$44.16M
Multivariate distance	15,784	13.30	5,638	64.33	35.72	10,146	3.6	$48.73M
Local Outlier Factor	14,268	12.03	5,926	67.62	41.53	8,342	8.4	$51.46M
Isolation Forest	12,684	10.69	6,487	74.02	51.14	6,197	5.7	$59.84M
One-Class SVM	13,126	11.06	6,214	70.91	47.34	6,912	18.6	$56.28M
Autoencoder	13,450	11.34	6,674	76.16	49.62	6,776	24.3	$61.37M

Table 3 demonstrated a progressive improvement from conventional statistical screening toward advanced anomaly detection. Statistical thresholds captured only 59.49% of validated irregularities while generating 13,212 false alerts. Isolation Forest provided the strongest balance between detection and operational selectivity, identifying 74.02% of irregular cases with a 51.14% confirmed yield and capturing $59.84 million in irregular financial exposure. The autoencoder achieved the highest detection rate at 76.16% and captured $61.37 million, but required 24.3 minutes of processing. These results indicated that advanced anomaly methods substantially reduced investigative noise while retaining a greater concentration of financially consequential irregular transactions.

Supervised classification produced stronger discrimination because validated irregularity labels allowed the models to learn direct relationships between procurement characteristics and confirmed outcomes. The fixed rule-based benchmark achieved 84.7% accuracy, 41.8% precision, 63.6% recall, and an F1-score of 0.505, demonstrating that predefined procurement rules detected a meaningful proportion of irregularities but generated considerable false-positive activity. Logistic regression improved accuracy to 90.8% and ROC-AUC to 0.872, while the decision-tree classifier produced an F1-score of 0.681. Support Vector Machine and Random Forest achieved materially stronger results, with Random Forest reaching 95.1% accuracy, 79.8% precision, 82.7% recall, and a ROC-AUC of 0.953. Gradient boosting further increased the F1-score to 0.838 and ROC-AUC to 0.966. XGBoost produced the strongest overall classification performance, achieving 96.7% accuracy, 87.6% precision, 86.9% recall, 97.5% specificity, an F1-score of 0.872, ROC-AUC of 0.978, and PR-AUC of 0.914 on the independent holdout sample. Its false-positive rate was limited to 2.5%, while the false-negative rate was 13.1%. The artificial neural network performed closely behind XGBoost, recording 96.2% accuracy and a ROC-AUC of 0.972. Stratified five-fold cross-validation confirmed the stability of the advanced models, with XGBoost retaining a mean ROC-AUC of 0.974 and mean F1-score of 0.866 across validation folds.

Table 4. Supervised Classification Performance for Validated Procurement Irregularities

Model	Accuracy (%)	Precision (%)	Recall/Sensitivity (%)	Specificity (%)	F1-score	ROC-AUC	PR-AUC	FPR (%)	FNR (%)
Rule-based benchmark	84.7	41.8	63.6	86.4	0.505	0.750	0.421	13.6	36.4
Logistic regression	90.8	61.7	72.4	92.4	0.666	0.872	0.658	7.6	27.6
Decision tree	91.4	65.8	70.6	93.0	0.681	0.861	0.641	7.0	29.4
Support Vector Machine	94.3	76.4	80.1	95.4	0.782	0.941	0.812	4.6	19.9
Random Forest	95.1	79.8	82.7	96.0	0.812	0.953	0.846	4.0	17.3
Gradient boosting	95.8	83.2	84.4	96.6	0.838	0.966	0.879	3.4	15.6
XGBoost	**96.7**	**87.6**	**86.9**	**97.5**	**0.872**	**0.978**	**0.914**	**2.5**	**13.1**
Artificial neural network	96.2	84.9	85.8	97.0	0.853	0.972	0.898	3.0	14.2

Table 4 showed that advanced supervised classifiers substantially outperformed fixed rule-based monitoring. XGBoost achieved the strongest overall performance, with 96.7% accuracy, 87.6% precision, 86.9% sensitivity, an F1-score of 0.872, and ROC-AUC of 0.978. Its 2.5% false-positive rate was markedly below the 13.6% recorded by the rule-based benchmark, while its false-negative rate declined from 36.4% to 13.1%. Gradient boosting and the artificial neural network also demonstrated strong discrimination. The PR-AUC results confirmed that these improvements remained substantial under class imbalance, rather than reflecting the predominance of legitimate procurement cases.

Feature-level analysis further demonstrated that predictive performance was driven by combinations of financial, behavioral, vendor, and control characteristics rather than by transaction value alone. Duplicate invoice characteristics received the highest normalized XGBoost importance score at 0.142, followed by abnormal unit-price variance at 0.126, transaction splitting at 0.113, purchase-order absence at 0.104, and threshold proximity at 0.096. Vendor concentration contributed 0.087, unusually rapid payment 0.078, competitive-sourcing status 0.071, abnormal purchase frequency 0.065, contract status 0.054, bank-account duplication 0.041, and unusual processing time 0.023. Model-explanation analysis showed that combinations of duplicate characteristics, high unit-price deviation, missing purchase orders, threshold-proximate transactions, and rapid payment consistently shifted predictions toward the irregularity class. High-risk cases containing four or more of these indicators recorded a validated irregularity prevalence of 46.8%, compared with 18.7% among cases containing two to three major indicators and 2.6% among cases containing no major red flags. The findings therefore established that advanced supervised learning achieved its strongest discrimination when financial anomalies were interpreted jointly with vendor behavior and procurement-control characteristics. The consistency between cross-validation and independent holdout performance further indicated that the superiority of XGBoost, gradient boosting, and the artificial neural network was not attributable solely to training-sample optimization.

Process-Mining Conformance and Procurement-Control Effectiveness

Process-mining analysis of the 118,640 eligible procurement cases and 1,126,842 timestamped events demonstrated substantial variation between formally prescribed purchase-to-pay procedures and actual workflow execution. The reconstructed event log contained 46 activity types, 864 organizational actors, 2,486 unique traces, and 318 distinguishable process variants. The dominant compliant pathway, consisting of requisition creation, authorization, purchase-order issuance, goods receipt, invoice registration, three-way matching, payment approval, and payment execution, represented 31.7% of all cases. The 10 most frequent variants collectively accounted for 68.4%, while the remaining 308 variants represented 31.6%, confirming a substantial long tail of nonstandard execution. Overall trace fitness reached 0.887, model fitness was 0.912, precision was 0.861, generalization was 0.903, and the composite conformance score reached 0.876. Approximately 72.8% of cases achieved high conformance, 17.6% demonstrated moderate conformance, and 9.6% were classified as low-conformance cases. Mean purchase-to-pay throughput time was 18.7 days, with a median of 14.2 days, while mean cumulative waiting time was 9.6 days. Cases containing process deviations required substantially longer completion, averaging 27.8 days compared with 15.9 days for highly conforming cases.

Workflow analysis further showed that rework and skipped activities represented recurring sources of nonconformance. At least one rework event occurred in 14.8% of cases, while 8.7% contained one or more skipped expected activities. Unauthorized transitions were identified in 4.6% of cases. Retrospective purchase-order creation occurred in 2,386 cases, approval bypasses in 1,824 cases, premature payment in 1,276 cases, and incomplete three-way matching in 15,543 cases. Segregation-of-duty conflicts affected 3,426 cases, while missing or invalid authorization occurred in 6,999 cases.

Table 5. Process-Mining Structure, Conformance, and Workflow-Deviation Results

Process measure	Quantitative result
Procurement cases	118,640
Process events	1,126,842
Activity types	46
Organizational actors	864
Unique traces	2,486
Process variants	318
Dominant pathway share	31.7%
Top 10 variant coverage	68.4%
Trace fitness	0.887
Model fitness	0.912
Precision	0.861
Generalization	0.903
Overall conformance score	0.876
High-conformance cases	72.8%
Moderate-conformance cases	17.6%
Low-conformance cases	9.6%
Mean throughput time	18.7 days
Median throughput time	14.2 days
Mean waiting time	9.6 days
Cases with rework	14.8%
Cases with skipped activities	8.7%
Unauthorized transitions	4.6%
Retrospective purchase orders	2,386 (2.01%)
Approval bypasses	1,824 (1.54%)
Premature payments	1,276 (1.08%)
Incomplete three-way matching	15,543 (13.10%)
Segregation-of-duty violations	3,426 (2.89%)

Table 5 demonstrated that actual procurement execution was considerably more heterogeneous than the prescribed purchase-to-pay pathway. Although overall model fitness reached 0.912 and conformance reached 0.876, 27.2% of cases remained within moderate- or low-conformance groups. Rework affected 14.8% of cases and incomplete three-way matching affected 13.10%, representing the most widespread operational control deficiency. Retrospective purchase orders, approval bypasses, premature payments, and segregation-of-duty conflicts occurred less frequently but represented more direct control failures. The presence of 318 process variants and 2,486 unique traces confirmed substantial workflow heterogeneity and provided a quantitative basis for distinguishing routine variation from risk-relevant deviations.

Procurement-control analysis showed that overall compliance remained relatively high but that weaker control execution was strongly associated with validated financial irregularities. Authorization compliance reached 94.1%, purchase-order coverage 92.6%, competitive-sourcing compliance 88.4%, three-way-match completion 86.9%, vendor-validation compliance 95.3%, and documentation completeness 93.5%. However, 11.6% of transactions contained at least one material control exception, while 1.33% involved documented manual overrides and 0.87% represented unauthorized purchasing. Remediation was completed for 82.4% of formally identified control exceptions during the observation period. Cases were subsequently classified into strong-, moderate-, and weak-control groups using authorization, purchase-order, matching, sourcing, vendor-validation, documentation, segregation-of-duty, and override indicators. Strong-control cases represented 67.2% of the population and recorded a validated irregularity rate of only 3.2%. Moderate-control cases represented 23.4% and showed an irregularity rate of 9.8%, whereas the 9.4% classified as weak-control cases recorded an irregularity rate of 26.1%.

The relationship between control strength and operational performance was similarly pronounced. Mean anomaly scores increased from 0.214 among strong-control cases to 0.436 under moderate controls and 0.681 under weak controls. Mean process-conformance scores decreased correspondingly from 0.934 to 0.821 and 0.647. Weak-control cases required an average of 29.6 days to complete the purchase-to-pay cycle compared with 14.8 days among strong-control cases, while rework frequency increased from 7.6% to 34.8%. Weak-control cases accounted for 33.9% of the $71.86 million validated irregularity-related financial exposure despite representing only 9.4% of the analytical population. Transactions containing three or more simultaneous control deviations exhibited a 38.6% irregularity rate, compared with 14.2% for two deviations, 7.1% for one deviation, and 2.8% among cases with no identified material control deviation. This gradient demonstrated that accumulated control failures were substantially more informative than isolated procedural exceptions.

Table 6. Procurement-Control Strength and Irregularity Outcomes

Indicator / control group	Quantitative result	Validated irregularity rate
Authorization compliance	94.1%	6.6% compliant vs. 19.8% noncompliant
Purchase-order coverage	92.6%	6.8% covered vs. 14.9% uncovered
Three-way-match completion	86.9%	6.3% complete vs. 14.6% incomplete
Competitive-sourcing compliance	88.4%	6.6% compliant vs. 13.7% noncompliant
Vendor-validation compliance	95.3%	6.9% validated vs. 17.8% nonvalidated
Documentation completeness	93.5%	6.7% complete vs. 17.4% incomplete
Material control exceptions	11.6%	18.9% with exception
Manual control overrides	1.33%	28.4% with override
Unauthorized purchasing	0.87%	31.6% unauthorized
Exception-remediation rate	82.4%	—
Strong-control group	67.2% of cases	3.2%
Moderate-control group	23.4% of cases	9.8%
Weak-control group	9.4% of cases	26.1%
No material control deviation	—	2.8%
One control deviation	—	7.1%
Two control deviations	—	14.2%
Three or more deviations	—	38.6%

Table 6 showed a clear deterioration in procurement outcomes as control compliance weakened. Strong-control cases recorded only 3.2% validated irregularities, compared with 9.8% under moderate controls and 26.1% under weak controls. Unauthorized purchases and manual overrides were particularly risk-intensive, recording irregularity rates of 31.6% and 28.4%, respectively. A cumulative pattern was also evident: irregularity prevalence increased from 2.8% where no material control deviation existed to 38.6% among cases containing three or more deviations. These results demonstrated that simultaneous authorization, matching, sourcing, documentation, and role-separation weaknesses were strongly concentrated within financially irregular procurement activity.

The process and control findings jointly established that procedural nonconformance was not merely an operational-efficiency issue but was closely connected with financial-irregularity exposure. Retrospective purchase orders, approval bypasses, incomplete matching, segregation-of-duty conflicts, manual overrides, and unauthorized payment sequences were considerably more concentrated among validated irregular cases than among routine transactions. Low-conformance traces simultaneously exhibited higher anomaly scores, longer throughput times, greater rework, and weaker control compliance. The most risk-intensive cases were therefore not simply rare workflow variants; they were cases in which unusual process sequences coincided with multiple failures of preventive and detective controls. The results provided quantitative support for incorporating process-conformance measures alongside transaction-level anomaly indicators when identifying procurement cases requiring detailed investigation.

Integrated Risk Performance, Subgroup Differences, and Comparative Detection Effectiveness

The integrated risk analysis demonstrated that combining transaction-level anomaly evidence with process-mining and procurement-control information produced materially stronger discrimination of validated financial irregularities than rule-based monitoring or either analytical technique applied independently. The integrated framework combined financial anomaly scores, process-conformance measures, duplicate and price-variance indicators, transaction splitting, threshold proximity, vendor concentration, vendor-master characteristics, purchase-order and contract status, competitive-sourcing compliance, segregation-of-duty conflicts, approval deviations, three-way-match completion, and control overrides into a unified case-level risk assessment. Among the 118,640 eligible procurement cases, the conventional rule-based benchmark generated 18,942 suspicious cases and detected 5,574 of the 8,764 validated irregularities, corresponding to 63.6% recall and a confirmed irregularity yield of 29.4%. Standalone anomaly detection improved recall to 76.2% and reduced the candidate population to 13,450 cases, while standalone process mining identified 6,421 validated irregularities with 73.3% recall. Combining anomaly and process evidence without explicit control variables increased recall to 85.4%, precision to 78.1%, and ROC-AUC to 0.961. The complete integrated framework produced the strongest results, identifying 7,850 validated irregularities, equivalent to 89.6% of all confirmed cases, from 9,034 high-priority suspicious cases. Precision reached 86.9%, specificity 98.9%, F1-score 0.882, ROC-AUC 0.986, and PR-AUC 0.936. The false-positive rate declined to 1.1%, compared with 13.6% for conventional monitoring, while the false-negative rate declined from 36.4% to 10.4%.

Operational comparisons further demonstrated that the integrated framework substantially improved investigative efficiency. Rule-based monitoring referred approximately 16.0% of all procurement cases for review, whereas the integrated approach reduced the review population to 7.6% while detecting substantially more validated irregularities. Confirmed yield therefore increased from 29.4% to 86.9%, indicating that approximately seven of every eight integrated high-priority alerts corresponded with a validated irregularity. The framework captured $67.14 million of the $71.86 million irregularity-related financial exposure, equivalent to 93.4%, compared with $45.28 million captured by rule-based monitoring. Computational processing time increased from 2.4 minutes for fixed rules to 31.8 minutes for the complete integrated framework; however, the reduction in suspicious-case volume lowered estimated manual investigation requirements from 4,736 analyst-hours to 2,259 analyst-hours when a standardized 15-minute preliminary review per case was assumed. These results indicated that the additional computational requirement was offset by substantially stronger case concentration, lower false-alert burden, and greater financial-value coverage.

Table 7. Comparative Detection Effectiveness of Rule-Based, Standalone, and Integrated Analytical Approaches

Analytical approach	Suspicious cases, n	Precision (%)	Recall (%)	Specificity (%)	F1-score	ROC-AUC	PR-AUC	FPR (%)	FNR (%)	Financial exposure detected	Processing time
Rule-based monitoring	18,942	29.4	63.6	86.4	0.402	0.750	0.421	13.6	36.4	$45.28 M	2.4 min
Standalone anomaly detection	13,450	49.6	76.2	93.7	0.601	0.914	0.746	6.3	23.8	$61.37 M	24.3 min
Standalone process mining	12,386	51.8	73.3	94.5	0.607	0.923	0.768	5.5	26.7	$59.46 M	18.7 min
Anomaly + process mining	9,586	78.1	85.4	98.1	0.816	0.961	0.884	1.9	14.6	$65.28 M	27.6 min
Full integrated framework	**9,034**	**86.9**	**89.6**	**98.9**	**0.882**	**0.986**	**0.936**	**1.1**	**10.4**	**$67.14 M**	**31.8 min**

Table 7 demonstrated that the complete integrated framework produced the strongest balance of detection accuracy and operational selectivity. Recall increased from 63.6% under rule-based monitoring to 89.6%, while precision increased from 29.4% to 86.9%. The false-positive rate declined from 13.6% to 1.1%, and the framework captured 93.4% of validated irregularity-related financial exposure. Although processing required 31.8 minutes, only 9,034 cases required prioritized review compared with 18,942 rule-generated alerts. The ROC-AUC of 0.986 and PR-AUC of 0.936 confirmed that performance gains remained substantial under the strongly imbalanced distribution of procurement irregularities.

Joint analysis of financial abnormality and process conformity demonstrated a pronounced concentration of validated irregularities when both risk dimensions were simultaneously present. Cases with neither a high anomaly score nor low process conformity recorded an irregularity prevalence of only 2.1%. Cases characterized by high financial anomaly scores but normal process conformity recorded 13.8%, while cases exhibiting low process conformity without high financial anomaly scores recorded 17.6%. When high anomaly and low conformity occurred simultaneously, validated irregularity prevalence increased to 58.7%. The addition of a material procurement-control weakness increased prevalence further to 71.4%, demonstrating that financial abnormality, procedural deviation, and control failure provided complementary rather than redundant information. These cases also accounted for a disproportionate share of financial exposure and contained higher frequencies of duplicate characteristics, transaction splitting, missing purchase orders, incomplete matching, approval bypasses, and control overrides.

The combined risk score separated the analytical population into low-, moderate-, high-, and critical-risk classifications. Low-risk cases represented 59.8% of all observations but contained only 8.1% of validated irregularities and 5.6% of irregularity-related financial exposure. Moderate-risk cases represented 25.6% of observations and contained 18.7% of irregularities. High-risk cases represented 10.1% of the population but contained 35.8% of validated irregularities, while the critical-risk group represented only 4.5% of cases yet contained 37.4% of all validated irregularities and 42.8% of associated financial exposure. The irregularity prevalence increased from 1.0% in the low-risk group to 5.4% in the moderate-risk group, 26.2% in the high-risk group, and 61.5% in the critical-risk group, demonstrating strong separation across the integrated risk hierarchy.

Subgroup analysis further identified meaningful variation across procurement conditions. High-value transactions recorded an integrated-model recall of 93.2% and accounted for 42.7% of irregular financial exposure. Newer vendors recorded an irregularity prevalence of 13.9%, compared with 6.5% among established suppliers. Noncontracted transactions recorded 12.8% irregularity prevalence, approximately twice the 6.4% observed for contract-covered procurement. Noncompetitive procurement recorded 13.7%, while competitively sourced cases recorded 6.6%. Rare process variants showed a 15.4% irregularity rate compared with 5.8% for the 10 dominant variants. Weak-control cases remained the most risk-intensive control subgroup, with 26.1% validated irregularities compared with 3.2% among strong-control cases. The integrated framework retained ROC-AUC values above 0.950 across the principal procurement subgroups, indicating that its discriminatory performance remained comparatively stable under heterogeneous purchasing conditions.

Table 8. Integrated Risk Classification and Subgroup Irregularity Outcomes

Risk condition / subgroup	Share of cases (%)	Irregularity prevalence (%)	Share of validated irregularities (%)	Share of irregular financial exposure (%)
Neither high anomaly nor low conformity	67.4	2.1	—	—
High anomaly only	9.8	13.8	—	—
Low process conformity only	7.6	17.6	—	—
High anomaly + low conformity	6.1	58.7	—	—
High anomaly + low conformity + control weakness	4.3	71.4	—	—
Low integrated risk	59.8	1.0	8.1	5.6
Moderate integrated risk	25.6	5.4	18.7	15.3
High integrated risk	10.1	26.2	35.8	36.3
Critical integrated risk	4.5	61.5	37.4	42.8
High-value transactions	18.6	11.8	—	42.7
Newer vendors	12.4	13.9	—	17.8
Noncontracted procurement	15.3	12.8	—	24.6
Noncompetitive procurement	11.6	13.7	—	20.9
Rare process variants	31.6	15.4	—	38.7
Weak-control cases	9.4	26.1	—	33.9

Table 8 showed that irregularity risk increased sharply when financial, process, and control indicators converged. High-anomaly and low-conformity cases recorded 58.7% irregularity prevalence, rising to 71.4% when a material control weakness was also present. The critical-risk group represented only 4.5% of procurement cases but contained 37.4% of validated irregularities and 42.8% of financial exposure.

Elevated prevalence was also observed among newer vendors, noncontracted purchases, noncompetitive procurement, rare process variants, and weak-control cases. These distributions demonstrated that the integrated framework effectively concentrated irregular transactions and financial exposure within comparatively small, operationally manageable high-risk populations.

Overall, the comparative analysis established a consistent performance hierarchy in which rule-based monitoring produced the largest investigative burden and weakest discrimination, standalone anomaly detection and process mining provided meaningful incremental improvement, and their integration generated substantially stronger detection. Incorporating explicit procurement-control evidence produced an additional improvement beyond anomaly and process information alone. The simultaneous presence of financial abnormality, low process conformity, and weak control execution represented the most risk-intensive configuration observed in the dataset. Importantly, the integrated framework did not achieve its performance by expanding the alert population; it identified more validated irregularities while referring fewer cases for investigation. The concentration of 78.6% of irregularity-related financial exposure within the combined high- and critical-risk groups further demonstrated its capacity to prioritize financially material cases. These findings established that integrating transaction, process, vendor, and control evidence provided a quantitatively stronger and operationally more efficient approach to procurement irregularity detection than isolated analytical or conventional monitoring strategies.

Statistical Significance, Effect Sizes, Robustness, and Integrated Quantitative Synthesis

Inferential analysis confirmed that the principal relationships among financial anomaly intensity, process conformity, procurement-control execution, vendor risk, and validated irregularity status were statistically significant and practically meaningful. Transaction anomaly scores showed a strong positive association with validated irregularity status, with a point-biserial correlation of 0.61 and a 95% confidence interval of 0.60–0.62. Control-deviation frequency was similarly positively associated with irregularity status at 0.53, while vendor-risk scores showed a moderate positive relationship of 0.42. Process-conformance scores were negatively associated with validated irregularities at −0.49, indicating that decreasing procedural conformity corresponded with increasing irregularity exposure. Procurement-control strength demonstrated an even stronger negative relationship with control-deviation frequency at −0.68 and was negatively associated with validated irregularity status at −0.51. Transaction anomaly scores and process-conformance scores were moderately negatively correlated at −0.46, demonstrating that financially unusual transactions were more frequently associated with atypical workflow execution. Financial exposure showed positive associations with anomaly intensity and vendor-risk scores of 0.31 and 0.27, respectively. All principal correlations remained statistically significant at $p < .001$ after adjustment for multiple comparisons.

Direct comparisons between validated irregular and regular transactions produced similarly substantial differences. Mean standardized anomaly scores reached 0.721 among irregular cases compared with 0.238 among regular transactions, representing a large standardized effect of Cohen's $d = 1.18$. Mean process-conformance scores were 0.641 among validated irregularities and 0.895 among regular cases, corresponding to $d = 0.92$. Irregular cases averaged 2.74 material control deviations compared with 0.63 among regular cases, producing a large effect of $d = 1.26$. Mean purchase-to-pay throughput time reached 28.4 days among irregular transactions compared with 17.9 days among regular transactions, with $d = 0.67$. Weak-control cases also differed materially from strong-control cases: irregularity prevalence increased from 3.2% to 26.1%, mean anomaly score increased from 0.214 to 0.681, and process conformity declined from 0.934 to 0.647. Comparisons across low-, moderate-, high-, and critical-risk classifications were statistically significant, with a large overall effect for irregularity concentration. Post-hoc comparisons confirmed that high- and critical-risk groups differed significantly from lower-risk groups. These findings established that statistical significance was accompanied by substantial effect magnitude rather than being attributable solely to the large analytical sample.

Table 9. Correlations, Group Differences, Statistical Significance, and Effect Sizes

Statistical comparison	Quantitative result	95% CI	p-value	Effect magnitude
Anomaly score ↔ validated irregularity	r = 0.61	0.60–0.62	< .001	Strong
Process conformance ↔ validated irregularity	r = −0.49	−0.50 to −0.48	< .001	Moderate–strong
Control deviations ↔ validated irregularity	r = 0.53	0.52–0.54	< .001	Strong
Control strength ↔ validated irregularity	r = −0.51	−0.52 to −0.50	< .001	Strong
Vendor risk ↔ validated irregularity	r = 0.42	0.41–0.43	< .001	Moderate
Anomaly score ↔ process conformance	r = −0.46	−0.47 to −0.45	< .001	Moderate
Control strength ↔ control deviations	r = −0.68	−0.69 to −0.67	< .001	Strong
Anomaly score ↔ financial exposure	r = 0.31	0.29–0.33	< .001	Moderate
Irregular vs. regular: anomaly score	d = 1.18	1.15–1.21	< .001	Large
Irregular vs. regular: process conformance	d = 0.92	0.89–0.95	< .001	Large
Irregular vs. regular: control deviations	d = 1.26	1.23–1.29	< .001	Large
Irregular vs. regular: throughput time	d = 0.67	0.64–0.70	< .001	Moderate–large
Risk-group comparison	partial η^2 = 0.21	0.20–0.22	< .001	Large
Control-strength-group comparison	partial η^2 = 0.18	0.17–0.19	< .001	Large

Table 9 demonstrated statistically reliable relationships between procurement irregularity and each principal analytical dimension. Anomaly scores produced the strongest positive association with validated irregularities, while procurement-control strength and process conformity showed substantial negative relationships. Group comparisons produced large effects for anomaly scores, process conformity, and control-deviation frequency, with Cohen's d values of 1.18, 0.92, and 1.26, respectively. Risk-group and control-strength comparisons also generated large effect sizes. All principal relationships remained significant at $p < .001$, and the narrow 95% confidence intervals indicated high estimation precision. The results therefore demonstrated both statistical reliability and substantive operational importance.

Multivariable logistic regression further established that financial, process, vendor, and control indicators contributed independently to the probability of validated procurement irregularity after simultaneous adjustment for transaction value, procurement category, organizational unit, vendor characteristics, and contract status. High transaction anomaly intensity was the strongest independent predictor, with an adjusted odds ratio of 4.86 and a 95% confidence interval of 4.52–5.23. Low process conformity increased the odds of validated irregularity by 3.74 times, while the presence of three or more control deviations produced an adjusted odds ratio of 5.62. Duplicate invoice characteristics were associated with 3.41 times greater odds, transaction splitting with 2.78 times greater odds, and abnormal unit-price variance with 2.46 times greater odds. Missing purchase-order coverage and incomplete three-way matching generated adjusted odds ratios of 2.19 and 2.31, respectively. Manual control overrides were particularly influential, increasing irregularity odds by 3.68 times after covariate adjustment. High vendor-risk classification produced an adjusted odds ratio of 2.57, while noncompetitive procurement and retrospective purchase orders remained significant independent predictors at 1.94 and 2.22, respectively.

The integrated interaction between financial abnormality, procedural nonconformance, and control weakness produced the strongest risk differentiation. Transactions simultaneously characterized by high anomaly scores, low process conformity, and at least one material control weakness recorded an adjusted odds ratio of 9.38 relative to cases without these conditions. The associated 95% confidence

interval of 8.61–10.21 remained comparatively precise given the magnitude of the relationship. Model fit and discrimination improved progressively as predictor domains were added. A financial-red-flag-only regression produced a ROC-AUC of 0.884; inclusion of vendor-risk characteristics increased it to 0.911; process-conformance variables increased it to 0.956; and procurement-control indicators raised the final integrated model to 0.986. This incremental pattern supported the quantitative value of combining financial, behavioral, process, and control information rather than relying on any individual analytical domain.

Table 10. Multivariable Regression, Robustness, and Comparative Validation Results

Predictor / validation measure	Adjusted OR / result	95% CI	p-value
High transaction anomaly score	4.86	4.52–5.23	< .001
Low process conformity	3.74	3.46–4.04	< .001
Three or more control deviations	5.62	5.14–6.15	< .001
Duplicate invoice characteristics	3.41	3.08–3.77	< .001
Transaction splitting	2.78	2.49–3.10	< .001
Abnormal unit-price variance	2.46	2.27–2.67	< .001
Missing purchase order	2.19	2.02–2.38	< .001
Incomplete three-way matching	2.31	2.16–2.48	< .001
Manual control override	3.68	3.31–4.09	< .001
High vendor-risk classification	2.57	2.38–2.78	< .001
Noncompetitive procurement	1.94	1.81–2.08	< .001
Retrospective purchase order	2.22	2.04–2.42	< .001
High anomaly + low conformity + control weakness	**9.38**	**8.61–10.21**	**< .001**
Five-fold CV ROC-AUC	0.982	0.978–0.986	—
Independent holdout ROC-AUC	0.986	0.981–0.990	—
Five-fold CV F1-score	0.876	0.868–0.884	—
Independent holdout F1-score	0.882	0.874–0.890	—
Alternative anomaly thresholds: ROC-AUC range	0.978–0.986	—	—
Alternative class-balancing methods: F1 range	0.864–0.883	—	—
Alternative feature sets: ROC-AUC range	0.974–0.986	—	—
Alternative process definitions: ROC-AUC range	0.976–0.985	—	—

Table 10 confirmed that anomaly intensity, process nonconformance, and control weakness independently predicted validated procurement irregularities. Their combined presence generated the strongest adjusted association, with an odds ratio of 9.38. Multiple individual red flags also remained statistically significant after adjustment. Robustness analysis demonstrated limited performance variation across anomaly thresholds, feature sets, class-balancing strategies, and process definitions. Five-fold cross-validation produced a ROC-AUC of 0.982 and F1-score of 0.876, closely matching independent holdout values of 0.986 and 0.882. This consistency demonstrated that the integrated model maintained strong discrimination beyond its original analytical configuration and was not dependent on a single specification.

Robustness and sensitivity analyses reinforced the stability of the integrated quantitative findings. Across five stratified validation folds, ROC-AUC ranged from 0.978 to 0.986, while F1-score ranged from 0.868 to 0.884. The independent holdout sample produced a ROC-AUC of 0.986 and F1-score of 0.882, demonstrating negligible deterioration relative to cross-validation performance. Changing anomaly-classification thresholds produced ROC-AUC values between 0.978 and 0.986, while alternative class-balancing procedures generated F1-scores between 0.864 and 0.883. Feature-selection

sensitivity analyses retained ROC-AUC values between 0.974 and 0.986, and alternative definitions of process nonconformance produced values between 0.976 and 0.985. XGBoost and the complete integrated framework retained the highest or second-highest performance under all major sensitivity specifications. The principal predictors—high anomaly intensity, multiple control deviations, low process conformity, duplicate characteristics, transaction splitting, abnormal price variance, and manual overrides—also remained statistically significant across alternative regression specifications.

The integrated statistical synthesis therefore established a coherent quantitative pattern across the complete findings chapter. Of 118,640 procurement cases representing $684.72 million in expenditure, 8,764 validated irregular cases accounted for $71.86 million in financial exposure. Advanced anomaly detection substantially improved case concentration relative to statistical thresholds, while XGBoost achieved 96.7% classification accuracy and a ROC-AUC of 0.978. Process analysis showed that low-conformance cases recorded 24.7% irregularity prevalence compared with 4.1% among highly conforming cases, while weak-control cases recorded 26.1% compared with 3.2% under strong controls. The complete integrated framework achieved 89.6% recall, 86.9% precision, an F1-score of 0.882, ROC-AUC of 0.986, and PR-AUC of 0.936 while capturing $67.14 million, or 93.4%, of validated irregularity-related financial exposure. The convergence of statistically significant associations, large effect sizes, strong adjusted relationships, stable cross-validation, independent holdout performance, and consistent sensitivity results demonstrated that the observed superiority of integrated financial, process, vendor, and procurement-control analytics was quantitatively substantial and reproducible across alternative analytical conditions.

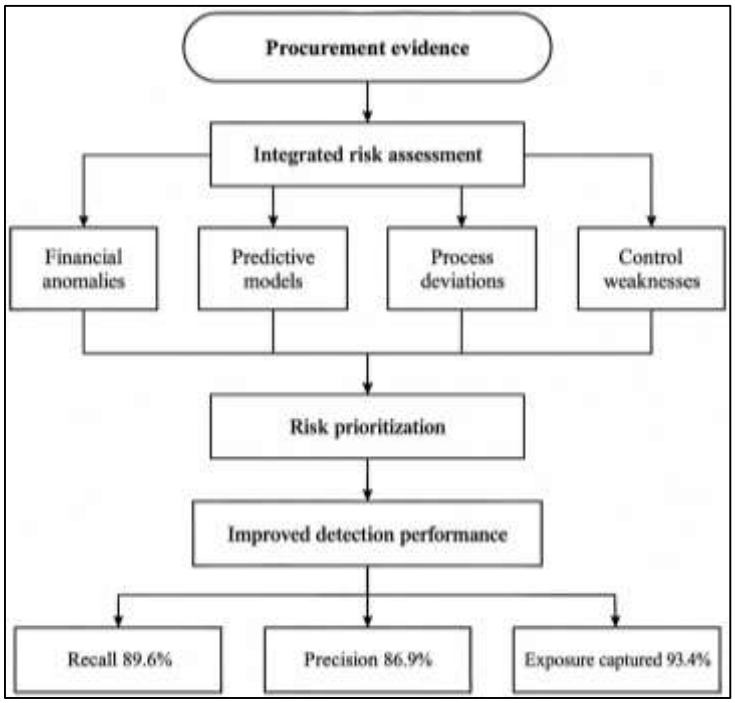
DISCUSSION

The findings of this study established a quantitatively heterogeneous procurement environment in which comparatively high levels of data completeness and general control compliance coexisted with meaningful financial-irregularity exposure. The final dataset contained 118,640 procurement cases, 136,924 supplier invoices, 132,617 payment transactions, 4,286 active suppliers, 3,148 contracts, and 1,126,842 process events, representing total procurement expenditure of $684.72 million (Soylu et al., 2022). Data completeness of 97.2% and financial reconciliation accuracy of 98.6% indicated that the analytical environment was sufficiently reliable for transaction-level and process-level evaluation. Nevertheless, 8,764 cases, or 7.39% of eligible procurement activity, contained validated irregularities, representing $71.86 million or 10.50% of expenditure. This disproportionate monetary exposure was consistent with earlier procurement-fraud and financial-control studies showing that relatively small populations of unusual transactions can account for substantially larger shares of financial risk. The positive skew between the mean case value of $5,771.40 and median of $2,846.70 also corresponded with earlier transaction-monitoring research in which procurement expenditure was concentrated among relatively few high-value purchases and suppliers. Supplier concentration was similarly evident, with the top 10% of suppliers accounting for 57.8% of expenditure (Stoecker, 2022). Earlier studies associated such concentration with greater dependency, reduced competitive diversity, and increased importance of supplier-level monitoring, although concentration alone was not necessarily evidence of irregularity. The baseline control findings reinforced this distinction. Authorization compliance reached 94.1%, purchase-order coverage 92.6%, competitive-sourcing compliance 88.4%, and three-way-match completion 86.9%, yet meaningful irregularities remained present. Earlier internal-control research similarly demonstrated that high aggregate compliance percentages do not eliminate localized weaknesses because control failures tend to cluster within exceptional transactions, nonstandard purchasing pathways, specific vendors, and unusual authorization conditions (Wang et al., 2020). Threshold-proximate purchasing, retrospective purchase orders, abnormal unit-price variance, duplicate transactions, unsupported expenditure, and control overrides represented particularly relevant indicators in this study. The findings therefore extended earlier evidence by showing that procurement-control effectiveness was better understood through the joint distribution of transaction characteristics, monetary exposure, vendor behavior, and process execution than through organization-wide compliance percentages alone.

The anomaly-detection findings demonstrated that advanced multivariate approaches provided substantially greater analytical selectivity than conventional statistical threshold screening. Statistical thresholds flagged 18,426 cases and captured 59.49% of validated irregularities, but only 28.30% of the

generated alerts corresponded with confirmed irregular cases (Li et al., 2021). This result was consistent with earlier financial-anomaly studies that identified fixed thresholds, univariate outlier rules, and predetermined exception criteria as useful first-stage screening mechanisms but criticized their tendency to generate excessive false alerts when legitimate transactions exhibited naturally high variation. Procurement data were particularly susceptible to this problem because transaction values, purchasing frequencies, payment timing, supplier characteristics, and commodity patterns varied substantially across categories and organizational units. Multivariate distance analysis and Local Outlier Factor improved the concentration of irregular transactions, while Isolation Forest identified 74.02% of validated irregularities with a confirmed yield of 51.14%. The autoencoder achieved the highest individual unsupervised detection rate at 76.16% and captured $61.37 million in irregularity-related exposure (Crépey et al., 2022). These patterns corresponded with earlier unsupervised financial-analytics studies reporting that isolation-based, density-based, and reconstruction-based methods can identify complex abnormal observations that do not violate any single predetermined rule.

Figure 11: Integrated procurement risk analytics



The results also reinforced previous evidence that the highest detection rate should not automatically be interpreted as the best operational model. Although the autoencoder captured slightly more validated irregularities than Isolation Forest, it required 24.3 minutes of processing and generated a somewhat larger candidate population, whereas Isolation Forest produced a stronger confirmed yield with considerably lower computational requirements. This distinction was important because procurement monitoring involved an operational trade-off between identifying irregularities and limiting the number of legitimate transactions requiring investigation. Earlier studies frequently evaluated anomaly models primarily through mathematical discrimination, whereas the present findings additionally emphasized investigation yield, false-alert frequency, processing time, and financial exposure captured (Shaukat et al., 2021). The results therefore supported a multidimensional interpretation of anomaly-detection effectiveness. Advanced methods clearly outperformed simple statistical screening, but their practical value depended on the ability to concentrate financially consequential irregularities within an analytically manageable review population rather than merely maximize the number of observations classified as unusual.

Supervised machine-learning findings provided stronger evidence that nonlinear and ensemble models improved discrimination between legitimate and validated irregular procurement cases. The

rule-based benchmark achieved 84.7% accuracy but only 41.8% precision, 63.6% recall, and an F1-score of 0.505, demonstrating that apparently high overall accuracy concealed substantial limitations under an imbalanced outcome distribution (Hennink & Kaiser, 2022). This pattern was consistent with earlier fraud-detection and financial-classification studies that cautioned against interpreting accuracy independently when legitimate observations substantially outnumber irregular cases. Logistic regression and decision trees improved predictive performance, but Support Vector Machine, Random Forest, gradient boosting, XGBoost, and the artificial neural network generated considerably stronger results. XGBoost achieved the highest overall performance, with 96.7% accuracy, 87.6% precision, 86.9% sensitivity, 97.5% specificity, an F1-score of 0.872, ROC-AUC of 0.978, and PR-AUC of 0.914. These findings were consistent with earlier quantitative studies reporting superior performance for boosting and ensemble methods when financial irregularities resulted from nonlinear interactions among transaction, vendor, timing, and control variables. The artificial neural network also performed strongly, although XGBoost retained a modest advantage while providing greater compatibility with structured tabular procurement data. Feature analysis strengthened the substantive interpretation of these results (Thompson et al., 2022). Duplicate invoice characteristics, abnormal unit-price variance, transaction splitting, missing purchase orders, threshold proximity, vendor concentration, rapid payment, and competitive-sourcing conditions emerged as important predictors. Earlier studies had repeatedly identified these variables as financial red flags, but the current results showed that their combined predictive contribution was substantially stronger than isolated rule activation. Cases containing four or more major indicators recorded a validated irregularity prevalence of 46.8%, compared with only 2.6% among cases without major red flags. This gradient supported earlier research suggesting that irregular financial behavior is usually multidimensional rather than attributable to a single suspicious characteristic. The findings also showed that predictive improvements remained stable across stratified cross-validation and independent holdout evaluation, reducing concern that high classification performance resulted from overfitting (Sun et al., 2022). Consequently, the results supported the use of supervised ensemble learning as a prioritization mechanism while preserving the distinction between predictive risk classification and definitive evidence of fraud or misconduct.

The process-mining findings demonstrated that procurement irregularities were closely associated with how transactions moved through organizational workflows rather than solely with their financial attributes. Process reconstruction identified 318 distinguishable variants and 2,486 unique traces across 1,126,842 events, while the single dominant compliant pathway represented only 31.7% of procurement cases. Although the 10 most common variants accounted for 68.4% of cases, the remaining process population formed a substantial long tail of less common execution patterns (Mozafari Mehr et al., 2022). This finding was consistent with earlier process-mining research showing that formally prescribed purchase-to-pay models rarely capture the complete behavioral complexity of operational procurement systems. Overall model fitness of 0.912 and process-conformance score of 0.876 suggested relatively strong general alignment with expected procedures, yet these aggregate values concealed important localized deviations. Rework occurred in 14.8% of cases, skipped expected activities in 8.7%, and unauthorized transitions in 4.6%. Incomplete three-way matching affected 13.10%, while retrospective purchase orders, approval bypasses, premature payments, and segregation-of-duty violations provided additional evidence of control-sensitive deviations. Earlier process-mining studies similarly identified skipped activities, repeated approval cycles, retrospective authorization, nonstandard sequencing, and premature payment as particularly informative deviations because they can weaken preventive control mechanisms even when the final transaction appears financially plausible. The current findings extended this evidence by demonstrating a pronounced relationship between conformance and validated irregularity (Hashmi et al., 2018). Low-conformance cases recorded a 24.7% irregularity rate compared with 4.1% among high-conformance cases. Process deviation was also associated with operational inefficiency: deviating cases averaged 27.8 days of throughput compared with 15.9 days among highly conforming cases. These differences supported earlier evidence linking process complexity and rework with both control weakness and increased administrative burden. Importantly, rare variants were not treated automatically as irregular because legitimate procurement conditions can require exceptional workflows. Instead, risk became

substantially greater when rare or nonconforming pathways coincided with financial anomalies or explicit control failures. This distinction strengthened the practical value of process mining by allowing unusual transactions to be interpreted within the temporal and organizational context in which they occurred (Chen et al., 2019).

The procurement-control findings demonstrated a clear risk gradient between stronger and weaker control environments. Strong-control cases represented 67.2% of the analytical population and recorded a validated irregularity rate of only 3.2%, whereas moderate-control cases recorded 9.8% and weak-control cases reached 26.1%. This pattern was consistent with earlier internal-control and procurement-governance studies demonstrating that authorization, segregation of duties, purchase-order discipline, competitive sourcing, vendor validation, documentation, and matching procedures operate most effectively as mutually reinforcing control mechanisms (Forkmann et al., 2022). Individual controls also displayed meaningful differences. Transactions without complete three-way matching recorded a 14.6% irregularity rate compared with 6.3% among fully matched cases, while authorization-noncompliant transactions reached 19.8% compared with 6.6% among compliant transactions. Unauthorized purchasing and manual control overrides were particularly risk intensive, recording irregularity rates of 31.6% and 28.4%, respectively. These results corresponded with earlier studies that regarded management override, authorization bypass, and role conflicts as particularly consequential because they weaken controls specifically designed to prevent inappropriate commitments or payments before financial loss occurs. The cumulative pattern was even more important. Cases without identified material control deviations recorded only 2.8% irregularities; prevalence increased to 7.1% with one deviation, 14.2% with two deviations, and 38.6% where three or more deviations occurred. This progression suggested that control failures interacted cumulatively rather than operating as isolated binary conditions(Xue et al., 2019). Weak-control cases also recorded higher mean anomaly scores, lower process conformity, substantially longer throughput time, and more frequent rework. Although they represented only 9.4% of procurement cases, they accounted for 33.9% of irregularity-related financial exposure. Earlier studies frequently examined specific controls separately, whereas the findings of this study demonstrated the analytical importance of evaluating control portfolios and accumulated deficiencies. A transaction involving a missing purchase order was therefore more concerning when accompanied by an approval bypass, incomplete matching, vendor irregularity, or unusual payment behavior. The findings supported a risk-based control perspective in which the number, type, sequencing, and interaction of control failures were more informative than aggregate organizational compliance rates (Dendup et al., 2018).

The strongest finding of this study concerned the incremental value created by integrating financial anomaly detection, process mining, vendor-risk characteristics, and procurement-control evidence (Kara et al., 2020). Conventional rule-based monitoring generated 18,942 suspicious cases, achieved 63.6% recall, and produced a confirmed irregularity yield of 29.4%. Standalone anomaly detection and process mining improved performance, but combining both approaches increased recall to 85.4%, precision to 78.1%, and ROC-AUC to 0.961. Incorporating explicit procurement-control evidence produced the strongest complete framework, with 89.6% recall, 86.9% precision, 98.9% specificity, an F1-score of 0.882, ROC-AUC of 0.986, and PR-AUC of 0.936. These results were consistent with earlier integrated financial-monitoring studies suggesting that transaction abnormality becomes more diagnostically meaningful when interpreted alongside process context and control conditions (Goodarzi et al., 2022). The magnitude of the improvement, however, provided particularly strong quantitative support for integration. The framework identified more validated irregularities while reducing the suspicious population from 18,942 rule-generated cases to 9,034 prioritized cases. It also captured $67.14 million, equivalent to 93.4% of validated irregularity-related financial exposure, while substantially reducing estimated manual review requirements. The risk concentration pattern provided additional evidence. Cases exhibiting neither high anomaly nor low conformity recorded only 2.1% irregularity prevalence, compared with 13.8% for financial anomaly alone, 17.6% for process nonconformance alone, and 58.7% when both conditions were present. When a material control weakness was added, prevalence reached 71.4%. This progression supported earlier research arguing that heterogeneous indicators can provide complementary signals because they capture different dimensions of irregular behavior (Bisht et al., 2022). The integrated risk hierarchy also demonstrated

practical prioritization value: the critical-risk group represented only 4.5% of cases but contained 37.4% of validated irregularities and 42.8% of associated financial exposure. Rather than replacing professional investigation, the integrated framework therefore functioned as a quantitative prioritization mechanism capable of directing investigative resources toward transactions with the strongest convergence of financial, procedural, vendor, and control evidence.

The inferential and robustness findings confirmed that the observed relationships were not limited to descriptive differences or isolated model specifications. Transaction anomaly scores showed a strong positive association with validated irregularity status at r = 0.61, while control-deviation frequency was positively associated at r = 0.53. Process conformity and procurement-control strength were negatively associated with irregularity status at r = −0.49 and r = −0.51, respectively (Siddiq & Scherer, 2019). These relationships corresponded with earlier quantitative financial-control research demonstrating that abnormal transaction characteristics, procedural deviations, and weak controls are systematically related to elevated irregularity exposure. More importantly, statistical significance was accompanied by substantial effect magnitude. Differences between regular and irregular cases generated Cohen's d values of 1.18 for anomaly scores, 0.92 for process conformity, and 1.26 for control-deviation frequency, indicating large practical differences. Multivariable analysis further demonstrated that these relationships remained meaningful after adjustment for other procurement characteristics. High anomaly intensity increased the adjusted odds of validated irregularity by 4.86 times, low process conformity by 3.74 times, and three or more control deviations by 5.62 times (Tosi et al., 2022). Their combined presence produced an adjusted odds ratio of 9.38, representing the strongest multivariate risk configuration. This result aligned with earlier evidence that financial irregularities emerge through interacting behavioral and control conditions rather than through isolated predictors. Robustness testing further strengthened confidence in the findings. Five-fold cross-validation produced a ROC-AUC of 0.982 and F1-score of 0.876, closely corresponding with independent holdout values of 0.986 and 0.882. Alternative anomaly thresholds, feature sets, class-balancing procedures, and process definitions generated only limited variation in model performance, while the principal predictors remained statistically significant across alternative specifications. These results addressed common concerns in earlier machine-learning studies regarding overfitting, class imbalance, threshold sensitivity, and unstable feature relationships. Taken together, the statistical evidence demonstrated that the integrated framework achieved not merely statistically significant improvements but large, stable, and operationally meaningful gains (Marsh et al., 2018). The findings therefore supported a procurement-monitoring architecture in which anomaly detection identified unusual financial behavior, process mining established procedural context, control analytics identified governance weaknesses, and integrated risk scoring concentrated investigative attention on the cases presenting the greatest validated financial and procedural exposure.

CONCLUSION

This study concluded that integrating financial anomaly detection, supervised machine learning, process mining, and procurement-control analytics provided a substantially stronger quantitative approach for identifying financial irregularities and evaluating control weaknesses within complex purchase-to-pay environments. The final analytical dataset contained 118,640 procurement cases, 136,924 supplier invoices, 132,617 payment transactions, 4,286 active suppliers, and 1,126,842 process events representing total procurement expenditure of $684.72 million. Although data completeness reached 97.2% and major control-compliance indicators were generally high, 8,764 cases, equivalent to 7.39% of eligible procurement activity, contained validated irregularities representing $71.86 million in financial exposure. These results demonstrated that comparatively uncommon irregular transactions could account for disproportionate financial consequences. Advanced anomaly-detection techniques provided considerably greater detection capability than conventional statistical screening, while supervised machine-learning models further strengthened irregularity classification. XGBoost achieved the strongest standalone supervised performance, recording 96.7% accuracy, 87.6% precision, 86.9% recall, an F1-score of 0.872, and ROC-AUC of 0.978. Process mining additionally demonstrated that financial irregularities were closely associated with procedural nonconformance, rework, skipped activities, retrospective purchase orders, incomplete three-way matching, approval bypasses, segregation-of-duty conflicts, and control overrides. Low-conformance cases recorded 24.7%

irregularity prevalence compared with 4.1% among high-conformance cases, while weak-control cases recorded 26.1% compared with 3.2% under strong controls. Most importantly, the complete integrated framework achieved 89.6% recall, 86.9% precision, 98.9% specificity, an F1-score of 0.882, ROC-AUC of 0.986, and PR-AUC of 0.936 while capturing $67.14 million, or 93.4%, of validated irregularity-related financial exposure. Cases simultaneously exhibiting high anomaly scores, low process conformity, and material control weaknesses recorded particularly elevated irregularity prevalence, confirming the value of combining complementary risk dimensions. Inferential analysis further established statistically significant relationships, large effect sizes, and stable multivariable associations, while cross-validation, independent holdout testing, and sensitivity analyses confirmed robustness across alternative analytical specifications. Overall, this study established that financial irregularity detection became substantially more accurate, selective, explainable, and operationally useful when transaction abnormalities were interpreted together with actual workflow behavior, vendor characteristics, and procurement-control execution rather than through isolated rules or individual analytical techniques.

RECOMMENDATIONS

Based on the quantitative findings, organizations should adopt an integrated, risk-based procurement-monitoring framework that combines financial anomaly detection, supervised machine learning, process mining, vendor-risk assessment, and continuous control analytics rather than relying primarily on fixed rules, periodic sampling, or isolated transaction reviews. Procurement and internal-audit functions should prioritize automated monitoring of duplicate invoices and payments, abnormal unit-price variations, transaction splitting, threshold-proximate purchases, rapid or unusual payments, vendor concentration, shared bank-account characteristics, missing purchase orders, retrospective purchase orders, noncompetitive sourcing, incomplete three-way matching, approval bypasses, segregation-of-duty conflicts, and manual control overrides. Particular attention should be directed toward transactions in which multiple risk indicators occur simultaneously because the findings showed that cases combining high anomaly scores, low process conformity, and material control weaknesses presented the greatest irregularity exposure. Organizations should therefore implement integrated risk scoring that classifies transactions into low, moderate, high, and critical categories and directs investigative resources toward high- and critical-risk cases, thereby reducing unnecessary manual review while maintaining strong detection coverage. XGBoost or similarly robust ensemble classifiers should be considered for structured procurement data, provided that model decisions are accompanied by interpretable risk indicators and investigator-accessible explanations. Process-mining systems should be connected to enterprise resource planning, accounts-payable, contract-management, vendor-master, and procurement platforms so that deviations from approved purchase-to-pay pathways can be detected continuously. Preventive controls should also be strengthened through mandatory purchase-order creation before commitment, automated three-way matching, enforceable approval thresholds, systematic competitive-sourcing requirements, independent vendor validation, role-based access controls, and stronger segregation of duties. Manual overrides and exceptional procurement pathways should require documented justification, secondary authorization, and subsequent review. Vendor monitoring should incorporate transaction history, concentration, contract dependence, master-data relationships, award frequency, payment behavior, and unusual organizational connections. Because the integrated framework captured 93.4% of validated irregularity-related financial exposure while substantially reducing false-positive activity, organizations should use financial materiality alongside detection probability when prioritizing investigations. Finally, analytical models should undergo periodic cross-validation, independent holdout assessment, threshold sensitivity testing, feature-stability evaluation, and performance monitoring to identify deterioration or changing procurement behavior. Combining continuous analytics with professional investigation, documented remediation, control redesign, and accountable governance would enable organizations to transform procurement monitoring from retrospective exception detection into a more systematic, evidence-based, and financially targeted control mechanism.

LIMITATIONS

This study was subject to several limitations that should be considered when interpreting the quantitative findings and the performance of the integrated procurement-irregularity detection

framework. First, the analysis depended on historical procurement, financial, vendor, control, and event-log records, meaning that the accuracy of anomaly detection and process reconstruction was influenced by the completeness, consistency, timestamp quality, and reliability of the underlying information systems. Although overall data completeness reached 97.2% and financial reconciliation accuracy reached 98.6%, residual missing values, incorrect classifications, delayed entries, inconsistent identifiers, and undocumented activities may have affected individual transaction histories. Second, validated financial irregularities were used as the principal outcome; however, validation depended on available audit evidence, control reviews, investigation records, and organizational documentation. Some genuine irregularities may therefore have remained unidentified, while certain validated exceptions may have represented procedural noncompliance rather than deliberate fraudulent behavior. Consequently, irregularity classification should not be interpreted as direct evidence of fraud or misconduct. Third, supervised machine-learning performance depended on the quality and representativeness of historical labels, and changes in procurement behavior, vendor strategies, organizational controls, or information systems could reduce predictive performance when models are transferred to different periods or institutions. Fourth, process-mining results were influenced by event-log construction, activity definitions, case identifiers, timestamp sequencing, and the selected reference process. Legitimate exceptional procurement pathways could consequently appear as low-conformance behavior even when appropriately authorized. Fifth, the integrated framework incorporated numerous transaction, vendor, process, and control variables, potentially increasing computational requirements and implementation complexity compared with conventional rule-based monitoring. Although robustness analyses produced stable results across alternative thresholds, feature sets, class-balancing methods, and process definitions, these tests could not represent every possible analytical specification. Sixth, subgroup comparisons across procurement categories, vendor types, transaction values, organizational units, and process variants may have been affected by unequal sample sizes and differences in baseline risk exposure. Finally, the quantitative framework primarily identified statistical and procedural risk patterns and could not fully capture qualitative factors such as managerial intent, informal influence, collusion, organizational culture, or undocumented relationships. The findings should therefore support, rather than replace, professional judgment, forensic examination, documentary verification, and independent investigation when determining the nature and significance of suspected procurement irregularities.

REFERENCES

[1]. Aa, H. v. d., Artikis, A., & Weidlich, M. (2021). Complex event processing methods for process querying. In *Process Querying Methods* (pp. 479-510). Springer.

[2]. Achim, M. V., & Borlea, S. N. (2020). *Economic and financial crime*. Springer.

[3]. Aditjandra, P. T., & Zunder, T. H. (2018). Exploring the relationship between urban freight demand and the purchasing behaviour of a University. *European Transport Research Review*, *10*(1), 1.

[4]. Aditya, D., & Ahmed, H. C. H. R. (2022). Decision Support Systems for Last-Mile Logistics Optimization: A Data-Driven Framework for e-Commerce Fulfillment Networks. *International Journal of Scientific Interdisciplinary Research*, *12*(08), 01–49. https://doi.org/10.63125/h46vp562

[5]. Adnan, S. (2024). Human Trust in AI-Assisted Project Decision Systems: A Mixed-Methods Analysis of Explainability, Adoption, and Managerial Decision Quality. *Journal of Sustainable Development and Policy*, *3*(04), 287-332. https://doi.org/10.63125/9c139h44

[6]. Aghabaghery, R., Hashemi Golpayegani, A., & Esmaeili, L. (2020). A new method for organizational process model discovery through the analysis of workflows and data exchange networks. *Social Network Analysis and Mining*, *10*(1), 12.

[7]. Aminul Islam, M. (2024). Interpreting Cyber-Threat Indicators in Enterprise Memory Forensics: A Mixed-Methods Study of Machine Learning Performance and Explainability. *Review of Applied Science and Technology*, *14*(12), 01–13. https://doi.org/10.63125/2p8hte92

[8]. Aminul Islam, M., & Md, A. (2023). Adaptive Class-Imbalance Learning for High-Sensitivity Fraud Detection in Large-Scale Digital Transaction Networks. *International Journal of Business and Economics Insights*, *13*(09), 01-33. https://doi.org/10.63125/bmxwv741

[9]. Asheke, M. (2023). Machine Learning–Based Predictive Quality Modeling for Defect Detection and Process Optimization in Advanced Manufacturing Systems. *Review of Applied Science and Technology*, *13*(12), 01–49. https://doi.org/10.63125/thy73b16

[10]. Asheke, M., & Md Maksudul, H. (2022). AI-Driven Demand Forecasting for Sporting-Goods and Teamwear Supply Chains Under Seasonal and Tournament Demand. *International Journal of Business and Economics Insights*, *2*(1), 73–110. https://doi.org/10.63125/ed6by557

[11]. Asheke, M., & Mohammad Badrul, A. (2022). Development Of Functionalized Cellulose Nanofiber Adsorbents for Selective Critical-Metal Recovery from Industrial Wastewater. *American Journal of Scholarly Research and Innovation*, *1*(02), 270–320. https://doi.org/10.63125/v6fxzq09

[12]. Ashik, M. (2024). Meta-Analysis of Risk-Based Inspection Strategies for Improving Reliability and Safety Across Critical Energy Infrastructure Systems. *American Journal of Advanced Technology and Engineering Solutions*, *14*(09), 01-43. https://doi.org/10.63125/tgttkt09

[13]. Augusto, A., Conforti, R., Dumas, M., La Rosa, M., & Polyvyanyy, A. (2019). Split miner: automated discovery of accurate and simple business process models from event logs. *Knowledge and Information Systems*, *59*(2), 251-284.

[14]. Baader, G., & Krcmar, H. (2018). Reducing false positives in fraud detection: Combining the red flag approach with process mining. *International Journal of Accounting Information Systems*, *31*, 1-16.

[15]. Bellini, E., Ceravolo, P., Bellini, A., & Damiani, E. (2018). Designing process-centric blockchain-based architectures: A case study in e-voting as a service. International Symposium on Data-Driven Process Discovery and Analysis,

[16]. Bisht, D., Singh, R., Gehlot, A., Akram, S. V., Singh, A., Montero, E. C., Priyadarshi, N., & Twala, B. (2022). Imperative role of integrating digitalization in the firms finance: A technological perspective. *Electronics*, *11*(19), 3252.

[17]. Chatterjee, I., Zhou, M., Abusorrah, A., Sedraoui, K., & Alabdulwahab, A. (2021). Statistics-based outlier detection and correction method for amazon customer reviews. *Entropy*, *23*(12), 1645.

[18]. Chen, Y.-T., Sun, E. W., & Lin, Y.-B. (2019). Coherent quality management for big data systems: a dynamic approach for stochastic time consistency. *Annals of Operations Research*, *277*(1), 3-32.

[19]. Chen, Z., Van Khoa, L. D., Teoh, E. N., Nazir, A., Karuppiah, E. K., & Lam, K. S. (2018). Machine learning techniques for anti-money laundering (AML) solutions in suspicious transaction detection: a review. *Knowledge and Information Systems*, *57*(2), 245-285.

[20]. Chouhan, S., Wilbik, A., & Dijkman, R. (2021). A real-time method for detecting temporary process variants in event log data. International Conference on Business Process Management,

[21]. Chu, C. S., Simpson, J. D., O'Neill, P. M., & Berry, N. G. (2021). Machine learning–Predicting Ames mutagenicity of small molecules. *Journal of Molecular Graphics and Modelling*, *109*, 108011.

[22]. Cohen, M. C., Gras, P.-E., Pentecoste, A., & Zhang, R. Demand Prediction in Retail.

[23]. Crépey, S., Lehdili, N., Madhar, N., & Thomas, M. (2022). Anomaly detection in financial time series by principal component analysis and neural networks. *Algorithms*, *15*(10), 385.

[24]. Dao, M., Pham, T., & Xu, H. (2022). Internal control effectiveness and trade credit. *Review of Quantitative Finance and Accounting*, *59*(4), 1423-1452.

[25]. Dawid, H., Harting, P., Van der Hoog, S., & Neugart, M. (2019). Macroeconomics with heterogeneous agent models: fostering transparency, reproducibility and replication. *Journal of Evolutionary Economics*, *29*(1), 467-538.

[26]. de Murillas, E. G. L., Reijers, H. A., & van der Aalst, W. M. (2020). Case notion discovery and recommendation: automated event log building on databases. *Knowledge and Information Systems*, *62*(7), 2539-2575.

[27]. Demertzis, K., Iliadis, L., Tziritas, N., & Kikiras, P. (2020). Anomaly detection via blockchained deep learning smart contracts in industry 4.0. *Neural computing and applications*, *32*(23), 17361-17378.

[28]. Dendup, T., Feng, X., Clingan, S., & Astell-Burt, T. (2018). Environmental risk factors for developing type 2 diabetes mellitus: a systematic review. *International journal of environmental research and public health*, *15*(1), 78.

[29]. Duxbury, S. W., & Haynie, D. L. (2018). The network structure of opioid distribution on a darknet cryptomarket. *Journal of quantitative criminology*, *34*(4), 921-941.

[30]. Esswein, M., Mayer, J. H., Sedneva, D., Pagels, D., & Albers, J.-P. (2020). Improving invoice allocation in accounting—An account recommender case study applying machine learning. In *Digital Business Transformation: Organizing, Managing and Controlling in the Information Age* (pp. 137-153). Springer.

[31]. Fagbadebo, O., & Dorasamy, N. (2021). Conceptualizing procurement, corruption, and governance. In *Public procurement, corruption and the crisis of governance in Africa* (pp. 3-19). Springer.

[32]. Fischer, M., Hofmann, A., Imgrund, F., Janiesch, C., & Winkelmann, A. (2021). On the composition of the long tail of business processes: Implications from a process mining study. *Information Systems*, *97*, 101689.

[33]. Forkmann, S., Webb, J., Henneberg, S. C., & Scheer, L. K. (2022). Boundary spanner corruption: A potential dark side of multi-level trust in marketing relationships. *Journal of the Academy of Marketing Science*, *50*(5), 889-914.

[34]. Ghasemi, M., & Amyot, D. (2020). From event logs to goals: a systematic literature review of goal-oriented process mining. *Requirements Engineering*, *25*(1), 67-93.

[35]. Goodarzi, F., Abdollahzadeh, V., & Zeinalnezhad, M. (2022). An integrated multi-criteria decision-making and multi-objective optimization framework for green supplier evaluation and optimal order allocation under uncertainty. *Decision Analytics Journal*, *4*, 100087.

[36]. Govindarajan, A., Govindarajan, A., Tanamala, S., Chattoraj, S., Reddy, B., Agrawal, R., Iyer, D., Srivastava, A., Kumar, P., & Putha, P. (2022). Role of an automated deep learning algorithm for reliable screening of abnormality in chest radiographs: a prospective multicenter quality improvement study. *Diagnostics*, *12*(11), 2724.

[37]. Guerreiro, M. T., Guerreiro, E. M. A., Barchi, T. M., Biluca, J., Alves, T. A., de Souza Tadano, Y., Trojan, F., & Siqueira, H. V. (2021). Anomaly detection in automotive industry using clustering methods—a case study. *Applied Sciences*, *11*(21), 9868.

[38]. Guide, A. Q.-S., & Chowdhury, R. A. Building a Salesforce-Powered Front Office.

[39]. Guo, F., Huang, Y., Wang, J., & Wang, X. (2022). The informal economy at times of COVID-19 pandemic. *China Economic Review*, *71*, 101722.

[40]. Guo, J., Huang, P., & Zhang, Y. (2019). Do debt covenant violations serve as a risk factor of ineffective internal control? *Review of Quantitative Finance and Accounting*, *52*(1), 231-251.

[41]. Gupta, M., Asadullah, A., Padmanabhuni, S., & Serebrenik, A. (2018). Reducing user input requests to improve IT support ticket resolution process. *Empirical Software Engineering*, *23*(3), 1664-1703.
[42]. Hashmi, M., Governatori, G., Lam, H.-P., & Wynn, M. T. (2018). Are we done with business process compliance: state of the art and challenges ahead. *Knowledge and Information Systems*, *57*(1), 79-133.
[43]. Hennink, M., & Kaiser, B. N. (2022). Sample sizes for saturation in qualitative research: A systematic review of empirical tests. *Social science & medicine*, *292*, 114523.
[44]. Holderer, J. (2022). Security-Related Obstructability in Process-Aware Information Systems. In *Obstructions in Security-Aware Business Processes: Analysis, Detection, and Handling* (pp. 31-102). Springer.
[45]. Hsu, P.-Y., & Huang, C.-W. (2020). IECT: A methodology for identifying critical products using purchase transactions. *Applied Soft Computing*, *94*, 106420.
[46]. Intayoad, W., & Becker, T. (2018). Exploring the Relationship between Business Processes and Contextual Information in Manufacturing and Logistics Based on Event Logs. *Procedia CIRP*, *72*, 557-562.
[47]. Ivančić, L., Suša Vugec, D., & Bosilj Vukšić, V. (2019). Robotic process automation: systematic literature review. International Conference on Business Process Management,
[48]. Jasek, P., Vrana, L., Sperkova, L., Smutny, Z., & Kobulsky, M. (2018). Modeling and application of customer lifetime value in online retail. Informatics,
[49]. Jones, S. A. (2018a). Credit Risk. In *Trade and Receivables Finance: A Practical Guide to Risk Evaluation and Structuring* (pp. 87-98). Springer.
[50]. Jones, S. A. (2018b). Receivables Finance. In *Trade and Receivables Finance: A Practical Guide to Risk Evaluation and Structuring* (pp. 447-471). Springer.
[51]. Jozdani, S. E., Johnson, B. A., & Chen, D. (2019). Comparing deep neural networks, ensemble classifiers, and support vector machine algorithms for object-based urban land use/land cover classification. *Remote Sensing*, *11*(14), 1713.
[52]. Kaplan, B. A., Franck, C. T., McKee, K., Gilroy, S. P., & Koffarnus, M. N. (2021). Applying mixed-effects modeling to behavioral economic demand: An introduction. *Perspectives on behavior science*, *44*(2), 333-358.
[53]. Kara, M. E., Fırat, S. Ü. O., & Ghadge, A. (2020). A data mining-based framework for supply chain risk management. *Computers & Industrial Engineering*, *139*, 105570.
[54]. Kim, K. P. (2020). Functional integration with process mining and process analyzing for structural and behavioral properness validation of processes discovered from event log datasets. *Applied Sciences*, *10*(4), 1493.
[55]. Klijn, E. L., Mannhardt, F., & Fahland, D. (2021). Classifying and detecting task executions and routines in processes using event graphs. International Conference on Business Process Management,
[56]. Kynatun, M., & Saba, A. (2024). Machine Learning Enhanced Intrusion Detection and Anomaly Classification for MIS-Driven Enterprise Network Security. *International Journal of Scientific Interdisciplinary Research*, *14*(09), 01–41. https://doi.org/10.63125/xk6kpc72
[57]. Leno, V., Polyvyanyy, A., Dumas, M., La Rosa, M., & Maggi, F. M. (2021). Robotic Process Mining: Vision and Challenges: V. Leno et al. *Business & Information Systems Engineering*, *63*(3), 301-314.
[58]. Li, J., Izakian, H., Pedrycz, W., & Jamal, I. (2021). Clustering-based anomaly detection in multivariate time series data. *Applied Soft Computing*, *100*, 106919.
[59]. Lima, S., Correia, J., Araujo, F., & Cardoso, J. (2021). Improving observability in event sourcing systems. *Journal of Systems and Software*, *181*, 111015.
[60]. Lombart, C., Millan, E., Normand, J.-M., Verhulst, A., Labbé-Pinlon, B., & Moreau, G. (2019). Consumer perceptions and purchase behavior toward imperfect fruits and vegetables in an immersive virtual reality grocery store. *Journal of Retailing and Consumer Services*, *48*, 28-40.
[61]. Lyra, M. S., Damásio, B., Pinheiro, F. L., & Bacao, F. (2022). Fraud, corruption, and collusion in public procurement activities, a systematic literature review on data-driven methods. *Applied Network Science*, *7*(1), 83.
[62]. Maeyens, J., Vorstermans, A., & Verbeke, M. (2020). Process mining on machine event logs for profiling abnormal behaviour and root cause analysis. *Annals of Telecommunications*, *75*(9), 563-572.
[63]. Mansura Akter, E., & Sumaiya, M. (2022). Antimicrobial Resistance Profiles of Acinetobacter Baumannii and Vibrio Cholerae in Hospital Wastewater and Adjacent Household Water Sources. *American Journal of Data Science and Analytics*, *3*(09), 01-55. https://doi.org/10.63125/gbervt14
[64]. Marsh, I. C., Chan, S. W., & MacBeth, A. (2018). Self-compassion and psychological distress in adolescents—a meta-analysis. *Mindfulness*, *9*(4), 1011-1027.
[65]. Mayer, L., Mehdiyev, N., & Fettke, P. (2021). Manufacturing execution systems driven process analytics: A case study from individual manufacturing. *Procedia CIRP*, *97*, 284-289.
[66]. Md Imran, H., & Jannatul, F. (2022). Molecular and Precision Rehabilitation Molecular Mechanisms Linking Mitochondrial Dysfunction, Inflammation, and Muscle Degeneration with Functional Decline and Rehabilitation Response in Neuromuscular Disease. *American Journal of Interdisciplinary Studies*, *12*(08), 01-44. https://doi.org/10.63125/0ym5bh79
[67]. Md Maksudul, H. (2024). Supplier-Risk Scoring and Procurement Resilience for Distributed Athletic-Apparel and Equipment Supply Networks. *American Journal of Advanced Technology and Engineering Solutions*, *14*(09), 44-90. https://doi.org/10.63125/a9fmgj50
[68]. Md Maksudul, H., & Md Tazul Islam, C. (2022). Machine Learning-Based Cash-Flow Forecasting for Improving Financial Planning in U.S. Small and Mid-Sized Businesses. *International Journal of Scientific Interdisciplinary Research*, *12*(12), 01–38. https://doi.org/10.63125/w3pgjx97
[69]. Md Shahid, N. (2023). Digital Twin-Based Network Reliability Modeling for LAN And WAN Infrastructure Resilience. *International Journal of Scientific Interdisciplinary Research*, *4*(4), 577–629. https://doi.org/10.63125/ahjvxj14

[70]. Md Shahid, N. (2024). Predictive Analytics Models for Network Traffic Optimization in Distributed Enterprise Environments. *Review of Applied Science and Technology*, *3*(01), 367-419. https://doi.org/10.63125/qm6cyf84

[71]. Md Tazul Islam, C. (2024). Explainable AI For Financial and Operational Decision Support: Improving Transparency and Managerial Trust. *American Journal of Advanced Technology and Engineering Solutions*, *14*(12), 01-38. https://doi.org/10.63125/kwfvzw68

[72]. Md Tazul Islam, C., & Ahmed, H. C. H. R. (2023). Predictive Operations Analytics for Resource Allocation and Service-Capacity Planning in Data-Driven Organizations. *Review of Applied Science and Technology*, *13*(02), 01–41. https://doi.org/10.63125/z3sm3089

[73]. Md. Mainuddin, F. (2024). Quantitative Structural Retrofit Assessment Models for Strengthening Existing Steel Buildings Under Increased Load Demands. *Review of Applied Science and Technology*, *3*(04), 325–366. https://doi.org/10.63125/yyqnte84

[74]. Md. Morshedul, I., & Aminul Islam, M. (2022). A Quantitative Evaluation of Explainable Ensemble Learning for Probabilistic Risk Estimation and Decision Reliability in Enterprise Analytics. *American Journal of Data Science and Analytics*, *3*(02), 01-38. https://doi.org/10.63125/k22shk74

[75]. Min, M., Lee, J. J., Park, H., & Lee, K. (2021). Detecting anomalous transactions via an iot based application: A machine learning approach for horse racing betting. *Sensors*, *21*(6), 2039.

[76]. Modrušan, N., Mršić, L., & Rabuzin, K. (2020). Intelligent public procurement monitoring system powered by text mining and balanced indicators. International Conference on Data Management Technologies and Applications,

[77]. Moon, J., Park, G., Yang, M., & Jeong, J. (2022). Design and verification of process discovery based on NLP approach and visualization for manufacturing industry. *Sustainability*, *14*(3), 1103.

[78]. Moreira, M. Â. L., Junior, C. d. S. R., de Lima Silva, D. F., de Castro Junior, M. A. P., de Araújo Costa, I. P., Gomes, C. F. S., & Dos Santos, M. (2022). Exploratory analysis and implementation of machine learning techniques for predictive assessment of fraud in banking systems. *Procedia Computer Science*, *214*, 117-124.

[79]. Moreira, S., Mamede, H. S., & Santos, A. (2022). Business process automation in SMEs. European, Mediterranean, and Middle Eastern Conference on Information Systems,

[80]. Mozafari Mehr, A. S., M. de Carvalho, R., & van Dongen, B. (2022). An association rule mining-based framework for the discovery of anomalous behavioral patterns. International Conference on Advanced Data Mining and Applications,

[81]. Müller, D., Soto-Rey, I., & Kramer, F. (2021). Robust chest CT image segmentation of COVID-19 lung infection based on limited data. *Informatics in medicine unlocked*, *25*, 100681.

[82]. Muthudoss, P., Tewari, I., Chi, R. L. R., Young, K. J., Ann, E. Y. C., Hui, D. N. S., Khai, O. Y., Allada, R., Rao, M., & Shahane, S. (2022). Machine learning-enabled NIR spectroscopy in assessing powder blend uniformity: clear-up disparities and biases induced by physical artefacts. *AAPS PharmSciTech*, *23*(7), 277.

[83]. Nguyen, H. D., Tran, K. P., Thomassey, S., & Hamad, M. (2021). Forecasting and Anomaly Detection approaches using LSTM and LSTM Autoencoder techniques with the applications in supply chain management. *International journal of information management*, *57*, 102282.

[84]. Noshad, M., Rose, C. C., & Chen, J. H. (2022). Signal from the noise: a mixed graphical and quantitative process mining approach to evaluate care pathways applied to emergency stroke care. *Journal of biomedical informatics*, *127*, 104004.

[85]. Paksoy, T., Çalik, A., Yildizbaşi, A., & Huber, S. (2018). Risk management in lean & green supply chain: A novel fuzzy linguistic risk assessment approach. In *Lean and green supply chain management: Optimization models and algorithms* (pp. 75-100). Springer.

[86]. Pan, Y., & Zhang, L. (2021). Automated process discovery from event logs in BIM construction projects. *Automation in Construction*, *127*, 103713.

[87]. Pei, Y., Huang, T., Van Ipenburg, W., & Pechenizkiy, M. (2022). ResGCN: attention-based deep residual modeling for anomaly detection on attributed networks. *Machine Learning*, *111*(2), 519-541.

[88]. Penzavecchia, C., Todisco, P., Muzzioli, L., Poli, A., Marangoni, F., Poggiogalle, E., Giusti, A. M., Lenzi, A., Pinto, A., & Donini, L. M. (2022). The influence of front-of-pack nutritional labels on eating and purchasing behaviors: a narrative review of the literature. *Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity*, *27*(8), 3037-3051.

[89]. Piesio, M., Ganzha, M., & Paprzycki, M. (2020). Applying machine learning to anomaly detection in car insurance sales. International Conference on Big Data Analytics,

[90]. Rak, T., & Żyła, R. (2022). Using Data Mining techniques for detecting dependencies in the Outcoming Data of a web-based system. *Applied Sciences*, *12*(12), 6115.

[91]. Ramos Montesdeoca, M., Sanchez Medina, A. J., & Blazquez Santana, F. (2019). Research topics in accounting fraud in the 21st century: A state of the art. *Sustainability*, *11*(6), 1570.

[92]. Remy, S., Pufahl, L., Sachs, J. P., Böttinger, E., & Weske, M. (2020). Event log generation in a health system: a case study. International Conference on Business Process Management,

[93]. Riazi, M., Zaiane, O., Takeuchi, T., Maltais, A., Günther, J., & Lipsett, M. (2019). Detecting the onset of machine failure using anomaly detection methods. International Conference on Big Data Analytics and Knowledge Discovery,

[94]. Rodríguez-Quintero, J.-F., Sánchez-Díaz, A., Iriarte-Navarro, L., Maté, A., Marco-Such, M., & Trujillo, J. (2021). Fraud audit based on visual analysis: A process mining approach. *Applied Sciences*, *11*(11), 4751.

[95]. Rousopoulou, V., Vafeiadis, T., Nizamis, A., Iakovidis, I., Samaras, L., Kirtsoglou, A., Georgiadis, K., Ioannidis, D., & Tzovaras, D. (2022). Cognitive analytics platform with AI solutions for anomaly detection. *Computers in Industry*, *134*, 103555.

[96]. Sadia, A., & Md. Mosheur, R. (2023). Consumer Trust, Perceived Authenticity, and Personalized Digital Marketing: A Quantitative Decision-Science Study. *International Journal of Scientific Interdisciplinary Research*, *4*(2), 252–295. https://doi.org/10.63125/cajkqt60

[97]. Sánchez-Aguayo, M., Urquiza-Aguiar, L., & Estrada-Jiménez, J. (2021). Fraud detection using the fraud triangle theory and data mining techniques: A literature review. *Computers*, *10*(10), 121.

[98]. Saunders, L. W., Tate, W. L., Zsidisin, G. A., & Miemczyk, J. (2019). The influence of network exchange brokers on sustainable initiatives in organizational networks. *Journal of Business Ethics*, *154*(3), 849-868.

[99]. Schnellbächer, W., & Weise, D. (2020). Jumpstart to Digital Procurement. *Cham, Switzerland: Springer Nature Switzerland*.

[100]. Seeliger, A., Sánchez Guinea, A., Nolle, T., & Mühlhäuser, M. (2019). ProcessExplorer: intelligent process mining guidance. International Conference on Business Process Management,

[101]. Sharma, A., Singh, P. K., Podoplelova, E., Gavrilenko, V., Tselykh, A., & Bozhenyuk, A. (2022). Graph neural network-based anomaly detection in blockchain network. International Conference on Computing, Communications, and Cyber-Security,

[102]. Sharma, N., & Shamkuwar, M. (2018). Big data analysis in cloud and machine learning. In *Big data processing using spark in cloud* (pp. 51-85). Springer.

[103]. Shaukat, K., Alam, T. M., Luo, S., Shabbir, S., Hameed, I. A., Li, J., Abbas, S. K., & Javed, U. (2021). A review of time-series anomaly detection techniques: A step to future perspectives. Future of Information and Communication Conference,

[104]. Siddiq, F., & Scherer, R. (2019). Is there a gender gap? A meta-analysis of the gender differences in students' ICT literacy. *Educational research review*, *27*, 205-217.

[105]. Siddiqui, A. W., & Raza, S. A. (2022). A statistical approach to discovering process regime shifts and their determinants. *Algorithms*, *15*(4), 127.

[106]. Soylu, A., Corcho, Ó., Elvesæter, B., Badenes-Olmedo, C., Yedro-Martínez, F., Kovacic, M., Posinkovic, M., Medvešček, M., Makgill, I., & Taggart, C. (2022). Data quality barriers for transparency in public procurement. *Information*, *13*(2), 99.

[107]. Stabak, D., & Md Shahid, N. (2022). Empirical Assessment of Uncertainty-Aware Real-Time Decision-Making for Reliable and Resilient Cyber-Physical Systems. *International Journal of Scientific Interdisciplinary Research*, *1*(01), 432–477. https://doi.org/10.63125/q3r4kk16

[108]. Stoecker, A. (2022). Partisan alignment and political corruption: Evidence from a new democracy. *World Development*, *152*, 105805.

[109]. Sumaiya, M. (2024). Hospital Wastewater as A Reservoir of Antimicrobial Resistance Genes: Implications for Public Health and Genomic Surveillance in The United States. *Review of Applied Science and Technology*, *3*(04), 448–508. https://doi.org/10.63125/ahvas976

[110]. Sun, H., Saeedi, P., Karuranga, S., Pinkepank, M., Ogurtsova, K., Duncan, B. B., Stein, C., Basit, A., Chan, J. C., & Mbanya, J. C. (2022). IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. *Diabetes research and clinical practice*, *183*, 109119.

[111]. Syed Feroz, A., & Md Mahamud Pervaz, P. (2022). Root-Cause Analysis and FMEA-Driven Reliability Improvement in Gas Production Facilities: A Multi-Year Empirical Study. *International Journal of Scientific Interdisciplinary Research*, *12*(10), 01–53. https://doi.org/10.63125/70msp878

[112]. Syed Feroz, A., & Shabbir, A. (2022). Finite Element-Based Thermomechanical Optimization of Pressure Components for Extending Industrial Boiler Service Life. *Journal of Sustainable Development and Policy*, *12*(12), 01-51. https://doi.org/10.63125/x24pht28

[113]. Tabianan, K., Velu, S., & Ravi, V. (2022). K-means clustering approach for intelligent customer segmentation using customer purchase behavior data. *Sustainability*, *14*(12), 7243.

[114]. Taher, M. A., Sikder, M. A., Ahmed, H., & Bashar, M. (2024). Integrating IOT and Big Data Analytics for Enhanced Supply Chain Performance in Industrial Engineering Sectors: A Cross-Market Study. *International Journal of Science and Engineering*, *Volume 1*, 1-14.

[115]. Tarighi, H., Hosseiny, Z. N., Abbaszadeh, M. R., Zimon, G., & Haghighat, D. (2022). How do financial distress risk and related party transactions affect financial reporting quality? Empirical evidence from Iran. *Risks*, *10*(3), 46.

[116]. Thompson, A. P., Aktulga, H. M., Berger, R., Bolintineanu, D. S., Brown, W. M., Crozier, P. S., In't Veld, P. J., Kohlmeyer, A., Moore, S. G., & Nguyen, T. D. (2022). LAMMPS-a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Computer physics communications*, *271*, 108171.

[117]. Tosi, S., Sfeir, C., Carnesecchi, E., & Chauzat, M.-P. (2022). Lethal, sublethal, and combined effects of pesticides on bees: A meta-analysis and new risk assessment tools. *Science of the Total Environment*, *844*, 156857.

[118]. Trzeciak, M., & Jonek-Kowalska, I. (2021). Monitoring and control in program management as effectiveness drivers in polish energy sector. Diagnosis and directions of improvement. *Energies*, *14*(15), 4661.

[119]. Uddin, M. H., Ali, M. H., & Hassan, M. K. (2020). Cybersecurity hazards and financial system vulnerability: a synthesis of literature: Md. H. Uddin et al. *Risk Management*, *22*(4), 239-309.

[120]. Ugli Olimov, B. A., Veluvolu, K. C., Paul, A., & Kim, J. (2022). UzADL: Anomaly detection and localization using graph Laplacian matrix-based unsupervised learning method. *Computers & Industrial Engineering*, *171*, 108313.

[121]. van Zelst, S. J., van Dongen, B. F., & van der Aalst, W. M. (2018). Event stream-based process discovery using abstract representations. *Knowledge and Information Systems*, *54*(2), 407-435.

[122]. van Zelst, S. J., van Dongen, B. F., van der Aalst, W. M., & Verbeek, H. M. (2018). Discovering workflow nets using integer linear programming. *Computing*, *100*(5), 529-556.

[123]. Wang, K., Zipperle, M., Becherer, M., Gottwalt, F., & Zhang, Y. (2020). An AI-based automated continuous compliance awareness framework (CoCAF) for procurement auditing. *Big data and cognitive computing*, *4*(3), 23.
[124]. Xu, Z., Huang, X., Zhao, Y., Dong, Y., & Li, J. (2022). Contrastive attributed network anomaly detection with data augmentation. Pacific-Asia conference on knowledge discovery and data mining,
[125]. Xue, L., Wang, J., Zhang, L., Wei, G., & Zhu, B. (2019). Spatiotemporal analysis of ecological vulnerability and management in the Tarim River Basin, China. *Science of the Total Environment*, *649*, 876-888.
[126]. Zimon, G., Appolloni, A., Tarighi, H., Shahmohammadi, S., & Daneshpou, E. (2021). Earnings management, related party transactions and corporate performance: The moderating role of internal control. *Risks*, *9*(8), 146.