



International Journal of Research Publication and Reviews

Journal Homepage: www.ijrpr.com ISSN 2582-7421

Machine Learning Detection of Revenue Leakage from Clinical Coding, Claims, and Reimbursement Discrepancies

Jine Laner

UConn Stamford — Stamford, Connecticut, USA

DOI : <https://doi.org/10.55248/gengpi.06.1226.2823>

ABSTRACT

Healthcare revenue cycles increasingly depend on accurate clinical documentation, coding, claims processing, and reimbursement, yet their complexity creates substantial opportunities for revenue leakage. Coding inaccuracies, incomplete documentation, claim denials, underpayments, contractual inconsistencies, and reimbursement discrepancies can reduce healthcare organizations' financial performance while increasing administrative burden. Conventional audit approaches are often retrospective, rule-based, and resource-intensive, limiting their ability to identify complex or evolving patterns across large volumes of clinical and financial data. This study examines a machine learning framework for detecting revenue leakage by integrating clinical coding, claims, payment, denial, and reimbursement information. The framework applies supervised and anomaly-detection techniques to identify discrepancies between expected and received reimbursement, coding irregularities, recurrent denial patterns, and financially significant claim anomalies. Predictive features incorporate diagnosis and procedure codes, billed charges, payer characteristics, reimbursement amounts, denial reasons, and historical payment behavior. Model performance can be evaluated using precision, recall, F1-score, ROC-AUC, and financial recovery indicators. The proposed approach supports risk-based claim prioritization, earlier identification of revenue loss, targeted auditing, and data-driven revenue-cycle management across healthcare reimbursement systems.

1. INTRODUCTION

1.1 Healthcare Revenue Cycle and Financial Sustainability

Healthcare organizations operate within increasingly complex financial environments in which the delivery of clinical care must be accurately translated into reimbursable financial transactions. The healthcare revenue cycle encompasses the interconnected processes of patient registration, clinical documentation, coding, charge capture, claims preparation, payer adjudication, reimbursement, and final revenue realization [1,2]. Each stage depends on accurate information generated during preceding activities, making revenue performance sensitive to errors throughout the clinical and administrative workflow. Clinical documentation establishes the basis for assigning diagnosis and procedure codes, while those codes influence claim construction, medical-necessity assessment, and reimbursement determination [3]. Effective revenue-cycle management is therefore important not merely for billing efficiency but also for maintaining liquidity, financing clinical operations, investing in healthcare infrastructure, and sustaining service delivery [4]. Consequently, weaknesses between documented clinical activity and realized reimbursement can progressively translate into significant financial losses.

1.2 Sources and Characteristics of Revenue Leakage

Revenue leakage describes legitimate income that a healthcare organization earns or reasonably expects but fails to realize because of deficiencies within coding, billing, claims, payment, or collection processes [5]. It is conceptually distinct from fraud, which involves intentional deception for financial benefit, and from waste, which generally reflects unnecessary expenditure or inefficient resource utilization. Leakage may arise through undercoding, omitted charges, incorrect modifiers, insufficient clinical documentation, coding inconsistencies, preventable claim denials, contractual underpayments, reimbursement shortfalls, and unresolved patient balances [6]. Some discrepancies are readily identifiable, whereas others emerge through complex relationships among clinical codes, payer rules, contractual terms, historical reimbursement patterns, and claim characteristics. Revenue leakage can therefore remain distributed across large transaction volumes, making individual losses appear insignificant while their cumulative financial effect becomes substantial [7]. Detecting these dispersed discrepancies requires analytical approaches capable of connecting clinical, claims, and reimbursement information.

1.3 Limitations of Conventional Detection and Audit Methods

Traditional revenue-integrity controls depend substantially on manual coding reviews, retrospective claim audits, predefined business rules, payer reconciliation, and exception-based reporting. Although these approaches remain valuable for compliance and financial control, they are often reactive and constrained by the volume and complexity of contemporary healthcare transactions [8]. Manual audits can examine only a fraction of claims, while fixed rules generally identify discrepancies that conform to previously recognized patterns. They may therefore overlook subtle interactions among diagnoses, procedures, payer behavior, reimbursement amounts, denial histories, and contractual adjustments. Retrospective detection also means that financial losses may already have occurred before intervention. Increasing data heterogeneity further complicates reconciliation across clinical, coding, claims, and payment systems, creating a need for more adaptive and scalable detection methods.

1.4 Study Aim, Scope, and Contribution

Against this background, this study develops a machine-learning framework for detecting potential revenue leakage from discrepancies across clinical coding, claims processing, and reimbursement data. Machine learning provides an opportunity to analyze multidimensional revenue-cycle information simultaneously and identify patterns that may not be captured through conventional rules or sampled audits [9]. The framework integrates coding characteristics, claim attributes, denial information, expected reimbursement, actual payments, payer behavior, and historical financial patterns to estimate leakage risk. Its contribution extends beyond binary anomaly identification by combining predictive detection with financial-exposure prioritization and interpretable decision support. This enables potentially high-value discrepancies to be identified earlier and directed toward appropriate coding review, claims investigation, or reimbursement recovery. Building on this foundation, the next section examines the specific coding, claims, and reimbursement mechanisms through which revenue leakage occurs and the analytical patterns machine-learning models must distinguish.

2. REVENUE LEAKAGE MECHANISMS AND ANALYTICAL FOUNDATIONS

2.1 Clinical Coding and Documentation Discrepancies

Clinical coding converts documented diagnoses, procedures, and services into standardized information used for claims processing and reimbursement. International Classification of Diseases (ICD) diagnosis codes, procedure codes, and modifiers therefore influence how payers interpret clinical encounters and determine payment obligations [7,8]. Revenue leakage can occur when documentation does not adequately support coded services, clinically relevant diagnoses are omitted, procedures are undercoded, or modifiers are incorrectly applied. Undercoding may reduce claim valuation, while inconsistent documentation-code alignment can trigger claim edits, medical-necessity reviews, or reimbursement adjustments [9,10]. Coding discrepancies may also arise from incomplete physician documentation, differences in coding interpretation, or failure to update records following clinical changes. Although some coding errors are individually small, repeated discrepancies across high claim volumes can produce substantial cumulative revenue losses. Clinical coding must therefore be evaluated alongside documentation quality and reimbursement outcomes rather than treated as an isolated administrative process [11].

2.2 Claims Processing and Denial-Related Leakage

Accurately coded services may still generate revenue leakage when claims encounter administrative or adjudication failures. Claims can be rejected before formal adjudication because of missing information, formatting errors, invalid identifiers, or incomplete demographic data, whereas denials generally occur after payer assessment against coverage, coding, authorization, or medical-necessity requirements [12]. Eligibility discrepancies, missing prior authorization, duplicate submissions, claim-edit failures, and missed timely-filing deadlines can prevent otherwise reimbursable services from generating expected revenue [13]. Unresolved or repeatedly resubmitted denials further increase administrative costs while extending accounts-receivable periods. Importantly, not every difference in reimbursement represents preventable leakage. Some payment reductions reflect legitimate payer policies, benefit limitations, contractual exclusions, or patient-specific coverage provisions [14]. Effective leakage detection must therefore distinguish avoidable administrative failures from valid payer-dependent reimbursement outcomes, reducing unnecessary investigations while directing attention toward claims where corrective intervention could realistically recover revenue.

2.3 Reimbursement and Contractual Discrepancies

Revenue realization ultimately depends on whether adjudicated payments correspond with the amounts legitimately expected under applicable reimbursement arrangements. Billed charges represent amounts submitted by providers, whereas allowed amounts reflect payer-recognized values after contractual rules and benefit provisions are applied [15]. Expected reimbursement may subsequently be distributed among payer payment, contractual adjustments, deductibles, coinsurance, copayments, and other patient responsibilities. Revenue leakage occurs when actual reimbursement falls below the correctly calculated expected amount without an appropriate contractual or adjudication explanation. Such discrepancies may result from incorrect fee schedules, inappropriate adjustments, payer processing errors, overlooked secondary coverage, or underpayment of eligible services [16]. Payment-variance detection consequently requires comparison of expected and actual reimbursement while accounting for payer contracts, service characteristics, adjustments, and patient liability. Repeated variance patterns can also reveal systematic underpayment behavior that may remain difficult to identify through individual claim reviews [17].

2.4 Machine Learning for Revenue-Cycle Anomaly Detection

The multidimensional nature of revenue leakage makes machine learning particularly applicable to healthcare revenue-cycle analytics. Conventional rules generally evaluate predetermined conditions independently, whereas machine-learning models can identify nonlinear relationships among clinical codes, claim characteristics, payer behavior, denial history, reimbursement amounts, and temporal payment patterns. Supervised classification can learn from previously confirmed leakage cases to estimate the probability that new claims contain recoverable discrepancies. Where reliable labels are limited, unsupervised anomaly-detection techniques can identify claims whose coding, payment, or reimbursement characteristics differ substantially from expected patterns. Ensemble methods can further combine multiple predictive models to improve robustness across heterogeneous claim populations. Rather than treating every anomaly equally, probability-based risk scoring can rank cases according to predicted leakage likelihood and potential financial exposure. This enables scarce audit resources to be concentrated on discrepancies with greater expected recovery value while preserving human review for final determination.

Figure 1. Conceptual Framework for Machine-Learning Detection of Healthcare Revenue Leakage

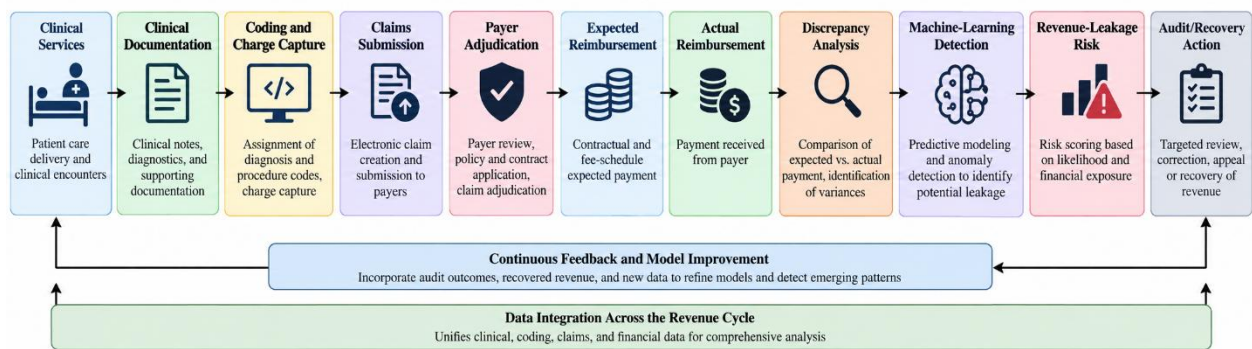


Figure 1. Conceptual Framework for Machine-Learning Detection of Healthcare Revenue Leakage

Together, these mechanisms demonstrate that revenue leakage is not generated at a single stage of the revenue cycle but can emerge through interconnected clinical, coding, claims, and reimbursement processes. Once these potential leakage pathways are established, reliable machine-learning detection depends on constructing an integrated analytical dataset that links the clinical origin of each service with its coding characteristics, claim history, adjudication outcome, and ultimate financial result.

3. DATA ACQUISITION, INTEGRATION, AND FEATURE ENGINEERING

3.1 Healthcare Revenue-Cycle Data Acquisition

Reliable machine-learning detection of healthcare revenue leakage requires an integrated data environment connecting clinical activity with the financial outcome of each encounter [15]. Electronic health records provide diagnoses, procedures, clinical documentation, service dates, provider information, and encounter characteristics that establish the clinical basis for billing [16]. Coding systems translate documented services into standardized diagnosis and procedure classifications used during claim generation and payer adjudication [17]. Charge-master systems provide billable service descriptions, units, departmental charges, and established prices, allowing documented clinical activity to be reconciled with captured charges [18]. Claims databases subsequently provide claim-line information, billed amounts, payer characteristics, submission histories, and adjudication outcomes [19]. Electronic remittance advice contributes allowed amounts, payments, adjustment codes, and denial information required for payment reconciliation [20]. Denial-management systems provide rejection, appeal, resubmission, and resolution histories [21]. Payer contracts and fee schedules establish benchmarks for expected reimbursement, whereas payment and accounts-receivable records determine whether expected revenue was received, delayed, adjusted, written off, or remained outstanding [22].

Table 1. Healthcare Data Sources and Revenue-Leakage Predictors

Data Source	Representative Variables	Leakage Signal	Analytical Function	Preprocessing Requirement
Electronic health records	Diagnoses, procedures, encounter, provider, service date	Documentation-service mismatch	Clinical validation	Missing-value handling and encounter matching
Coding systems	ICD codes, procedure codes, modifiers	Undercoding, omissions, coding inconsistency	Coding-discrepancy detection	Code harmonization

Data Source	Representative Variables	Leakage Signal	Analytical Function	Preprocessing Requirement
Charge master	Service, unit, charge, department	Missing or inaccurate charges	Charge-capture reconciliation	Service mapping and duplicate removal
Claims database	Claim type, payer, billed amount, claim status	Rejection, denial, abnormal claim pattern	Claims modeling	Claim-line aggregation
Remittance advice	Allowed amount, payment, adjustment code	Reimbursement shortfall	Payment reconciliation	Adjustment-code standardization
Denial-management system	Denial reason, appeal, resubmission	Recurrent preventable denial	Denial-risk modeling	Reason-code harmonization
Payer contracts	Contractual rate, fee schedule, payment terms	Contractual underpayment	Expected-payment estimation	Contract and rate standardization
Payment/accounts receivable	Payment, balance, write-off, ageing	Delayed or unrealized revenue	Financial-outcome analysis	Account and date reconciliation

3.2 Data Cleaning, Standardization, and Record Linkage

Healthcare revenue-cycle data originate from heterogeneous systems developed for different clinical, administrative, and financial functions, resulting in variations in structure, terminology, completeness, and granularity [23]. Data preparation should therefore begin by evaluating missing observations according to their clinical and financial significance rather than removing incomplete records indiscriminately [24]. Duplicate encounters, claims, and claim lines must be detected to prevent repeated transactions from distorting model training. Diagnosis codes, procedure codes, modifiers, denial codes, payer identities, and claim-status categories should also be transformed into consistent analytical formats [25].

Continuous variables with substantially different measurement scales can be standardized before model development. For variable x_i , standardization is expressed as:

$$z_i = \frac{x_i - \mu_x}{\sigma_x}$$

where z_i represents the standardized observation, μ_x is the mean, and σ_x is the standard deviation. Temporal alignment should preserve the sequence of service delivery, claim submission, adjudication, appeal, and payment. Patient identifiers, encounter numbers, claim identifiers, and service dates can then support record linkage across clinical and financial systems [15]. Remittance transactions should ultimately be matched to corresponding claims to establish an end-to-end representation of reimbursement [20].

3.3 Clinical Coding and Claims Feature Engineering

Feature engineering transforms standardized revenue-cycle records into measurable variables representing potential leakage mechanisms [16]. Clinical coding predictors can include diagnosis-code counts, procedure-code counts, modifiers, code specificity, diagnosis-procedure combinations, omitted-code indicators, and differences between initially submitted and subsequently corrected codes [17]. Such variables provide quantitative representations of coding complexity and potential documentation-code inconsistencies [18]. Claims-related predictors can additionally include billed amount, claim type, place of service, number of claim lines, payer category, service frequency, and submission behavior [19].

The delay between service delivery and claim submission can be represented as:

$$SD_i = D_{submit,i} - D_{service,i}$$

where SD_i denotes submission delay for claim i , $D_{submit,i}$ represents the submission date, and $D_{service,i}$ represents the corresponding service date.

Repeated resubmissions, claim edits, prior authorization deficiencies, and historical denials can provide indicators of administrative instability and potential preventable leakage [21]. Provider-level patterns can also capture systematic variation in coding practices, claim values, and denial experience [22]. For provider j , the historical denial rate can be expressed as:

$$DR_j = \frac{N_{denied,j}}{N_{submitted,j}}$$

where $N_{denied,j}$ is the number of denied claims and $N_{submitted,j}$ represents total submitted claims. Combining encounter-, provider-, coding-, and claim-level features allows machine-learning algorithms to evaluate multidimensional relationships that conventional isolated rules may not adequately capture [23].

3.4 Reimbursement and Payment-Variance Features

Financial feature engineering focuses on quantifying differences between anticipated reimbursement and realized revenue [24]. Expected reimbursement should reflect applicable contractual rates, fee schedules, service characteristics, payer arrangements, and relevant reimbursement rules [22]. Actual reimbursement represents the payment received following adjudication and applicable financial adjustments [20]. For claim i , the absolute reimbursement variance is defined as:

$$RV_i = ER_i - AR_i$$

where RV_i is reimbursement variance, ER_i represents expected reimbursement, and AR_i represents actual reimbursement. A positive variance identifies a potential financial shortfall but should not automatically be classified as leakage because legitimate contractual adjustments, benefit limitations, and patient responsibility may explain the difference [15].

Because absolute variance is influenced by claim value, a proportional reimbursement-variance rate can be calculated as:

$$RVR_i = \frac{ER_i - AR_i}{ER_i} \times 100$$

Payment realization can additionally be measured as:

$$PR_i = \frac{AR_i}{ER_i}$$

where a value approaching 1 indicates closer correspondence between expected and realized reimbursement. Other predictive variables can include allowed amount, contractual adjustment, patient responsibility, write-offs, outstanding balance, denial frequency, and historical underpayment patterns [19]. Payment timeliness provides another potentially informative signal [21]:

$$PD_i = D_{payment,i} - D_{submission,i}$$

where PD_i measures the number of days between claim submission and payment. Payer-specific payment behavior can then establish historical baselines against which unusual reimbursement outcomes are evaluated [25].

3.5 Leakage Labels, Class Imbalance, and Data Partitioning

Supervised learning requires a clearly defined outcome distinguishing verified revenue leakage from legitimate reimbursement variation [23]. Audit-confirmed underpayments, recovered omitted charges, corrected coding discrepancies, and successfully appealed payment errors can provide positive labels, whereas claims verified as appropriately reimbursed constitute negative observations [24]. The binary outcome can be represented as:

$$y_i = \begin{cases} 1, & \text{confirmed revenue leakage} \\ 0, & \text{no confirmed revenue leakage} \end{cases}$$

The prevalence of leakage within N claims is:

$$\pi = \frac{\sum_{i=1}^N y_i}{N}$$

A low value of π indicates class imbalance, which can bias conventional classifiers toward the more frequent non-leakage class [25]. Class weighting or appropriate resampling can therefore be incorporated during model training [17]. Data should also be divided chronologically into training, validation, and independent test periods to approximate prospective deployment [18]. Variables generated only after final adjudication or payment resolution must be excluded when predicting an earlier outcome because their inclusion would create data leakage and overestimate predictive performance [21].

Collectively, these procedures transform fragmented clinical, coding, claims, and reimbursement records into an analytically coherent representation of healthcare revenue realization [16]. Record linkage preserves the relationship between delivered services and subsequent financial outcomes [20], while feature engineering converts those relationships into measurable indicators of coding discrepancies, administrative failures, payment variance, and potential underpayment [22]. These integrated signals provide the analytical inputs required for Section 4, where supervised classification, anomaly detection, ensemble learning, and financial risk scoring are developed to distinguish routine reimbursement variation from claims exhibiting elevated revenue-leakage risk.

4. MACHINE-LEARNING REVENUE-LEAKAGE DETECTION FRAMEWORK

4.1 Supervised Revenue-Leakage Classification

When historical audits, corrected claims, successful appeals, or confirmed reimbursement discrepancies provide reliable labels, revenue-leakage detection can be formulated as a supervised binary-classification problem [23]. For claim i , the model estimates the probability of leakage conditional on its clinical, coding, claims, and reimbursement characteristics:

$$P(y_i = 1|X_i) = \hat{p}_i$$

where X_i represents the predictor vector and $\hat{p}_i$ is the estimated leakage probability. Logistic regression provides an interpretable baseline by modeling log-odds as a linear combination of predictors [24]:

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \sum_{j=1}^m \beta_j x_{ij}$$

Random forest extends this framework by aggregating multiple decision trees and capturing nonlinear relationships and feature interactions [25]. Gradient-boosting models sequentially construct weak learners that correct preceding prediction errors, making them suitable for heterogeneous claims data [26]. Comparing interpretable and nonlinear models helps determine whether additional predictive complexity produces meaningful improvements in identifying financially consequential discrepancies.

4.2 Unsupervised and Anomaly-Detection Approaches

Confirmed leakage labels may be incomplete because many underpayments or coding discrepancies remain undiscovered until targeted review [27]. Unsupervised methods can therefore identify unusual claim-payment combinations without requiring previously verified outcomes. Isolation Forest detects anomalies by recursively partitioning observations, with unusual transactions generally requiring fewer partitions for isolation [28]. Clustering can identify groups of claims exhibiting comparable coding, payer, service, and reimbursement characteristics, after which observations distant from their assigned cluster may warrant investigation [29]. A general distance-based anomaly measure can be represented as:

$$A_i = d(X_i, C_k)$$

where A_i is the anomaly score and C_k represents the relevant cluster center. Autoencoders provide another approach by learning compressed representations of normal transaction patterns [30]. Reconstruction error,

$$RE_i = \frac{1}{m} \sum_{j=1}^m (x_{ij} - \hat{x}_{ij})^2$$

can identify observations poorly represented by learned normal patterns. Statistical outlier measures can similarly flag unusually large payment variances [31]. Such anomalies should initiate review rather than automatically establish revenue leakage.

4.3 Model Training, Validation, and Hyperparameter Optimization

Model development should preserve temporal ordering so that training observations precede validation and test periods, thereby approximating prospective revenue-cycle deployment [32]. Cross-validation within the training period can support hyperparameter selection while preventing the independent test set from influencing model configuration. Class weighting or carefully controlled resampling can address the relative scarcity of confirmed leakage cases [23]. For binary classification, weighted cross-entropy may be expressed as:

$$L = - \sum_{i=1}^N [w_1 y_i \log(\hat{p}_i) + w_0 (1 - y_i) \log(1 - \hat{p}_i)]$$

where w_1 and w_0 control the contribution of leakage and non-leakage observations. Grid search, randomized search, or Bayesian optimization can be used to identify suitable hyperparameters [26]. Regularization, tree-depth constraints, early stopping, and validation monitoring reduce overfitting [24]. Probability calibration should additionally be evaluated because downstream financial prioritization depends on predicted probabilities representing risk reliably [33].

4.4 Integrated Leakage Risk Scoring

A claim's predicted leakage probability alone does not indicate whether investigating it is financially worthwhile. Risk prioritization should therefore integrate coding discrepancies, denial risk, reimbursement variance, payer history, anomaly evidence, and estimated monetary exposure [27]. For claim i , a normalized composite leakage score can be formulated as:

$$RLS_i = w_1C_i + w_2D_i + w_3V_i + w_4H_i + w_5A_i + w_6F_i$$

subject to:

$$\sum_{j=1}^6 w_j = 1, w_j \geq 0$$

where C_i denotes coding-discrepancy probability, D_i claim-denial risk, V_i normalized payment variance, H_i historical payer-risk behavior, A_i anomaly score, and F_i normalized financial exposure. Weighting can be determined using validation performance, expert input, or optimization against historical recovery outcomes [28].

The expected recoverable financial value can subsequently be estimated as:

$$EV_i = \hat{p}_i \times FE_i$$

where FE_i represents estimated financial exposure. Thus, a moderately probable discrepancy involving substantial financial exposure may receive greater audit priority than a highly probable but financially negligible discrepancy [29]. Ranking claims according to both likelihood and monetary consequence aligns machine-learning detection with revenue-integrity objectives and limited audit capacity.

4.5 Explainability and Root-Cause Attribution

Revenue-cycle predictions require interpretable outputs because coders, auditors, and financial specialists must understand why individual claims are prioritized [30]. Global feature importance can identify variables that influence predictions across the overall population, while local explanation methods characterize the factors contributing to an individual claim's risk score. SHAP-based analysis can represent a prediction as:

$$f(x_i) = \phi_0 + \sum_{j=1}^m \phi_{ij}$$

where ϕ_0 represents the baseline model output and ϕ_{ij} represents the contribution of feature j to claim i 's prediction [31]. Rather than presenting only a numerical risk probability, the system can generate reason codes indicating suspected undercoding, documentation mismatch, recurrent denial behavior, abnormal contractual adjustment, payer underpayment, or unusual payment delay [32]. Such root-cause attribution supports targeted intervention and allows human reviewers to determine whether a detected anomaly represents genuine recoverable leakage or legitimate reimbursement variation [33].

Figure 2. Predictive Machine-Learning Pipeline for Revenue-Leakage Detection

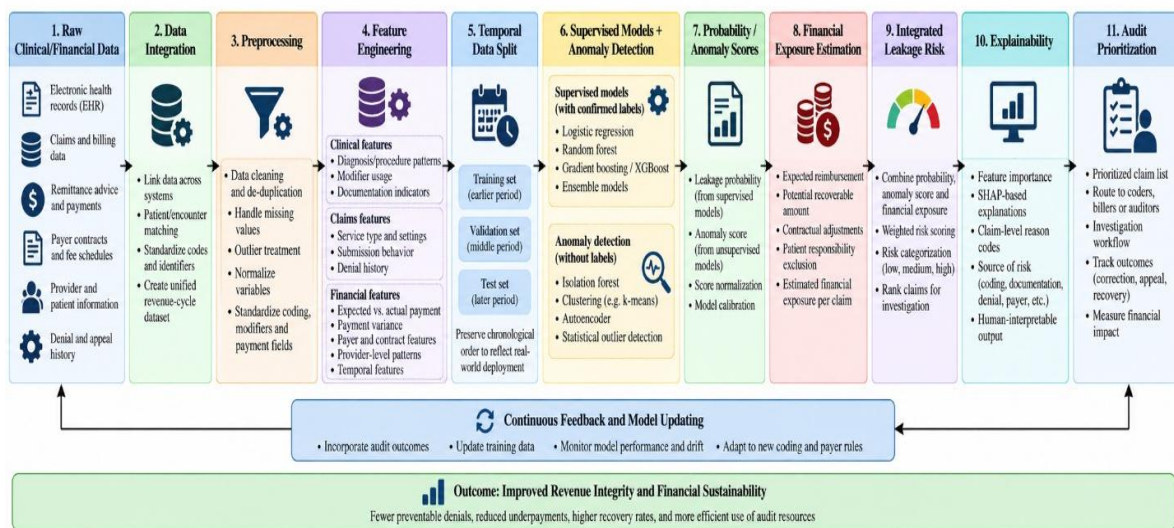


Figure 2. Predictive Machine-Learning Pipeline for Revenue-Leakage Detection

The framework therefore progresses from identifying statistical irregularities to estimating their practical financial significance. Supervised models exploit confirmed historical outcomes, anomaly-detection methods extend surveillance to previously unrecognized patterns, and integrated risk scoring

translates these outputs into prioritized financial investigations [25]. Explainability further connects prediction with the probable source of leakage, allowing intervention to be directed toward coding, documentation, claims processing, denial management, or payer reimbursement [30]. However, a high-risk score has limited operational value unless it reliably separates recoverable leakage from normal reimbursement variation. Section 5 therefore evaluates the framework using both conventional predictive-performance measures and financially oriented metrics that reflect the economic value of detection.

5. PERFORMANCE EVALUATION AND COMPARATIVE ANALYSIS

5.1 Leakage-Sensitive Evaluation Metrics

Evaluation of revenue-leakage detection should reflect the operational reality that confirmed coding errors, payer underpayments, and recoverable claim discrepancies may represent only a small fraction of processed claims [31]. Consequently, overall accuracy can be misleading when correctly classified routine claims dominate the test population. Precision measures how many claims referred for investigation actually contain confirmed leakage, directly reflecting unnecessary audit workload [32]:

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall measures the proportion of known leakage cases captured:

$$\text{Recall} = \frac{TP}{TP + FN}$$

Specificity and false-positive rate quantify the model's ability to avoid incorrectly escalating correctly reimbursed claims [33]:

$$\text{Specificity} = \frac{TN}{TN + FP}, \text{ FPR} = \frac{FP}{FP + TN}$$

The F1-score balances precision and recall, while PR-AUC evaluates performance across decision thresholds and is particularly informative for imbalanced leakage populations [34]. ROC-AUC remains useful for overall discrimination, but PR-AUC should receive greater emphasis where positive cases are uncommon [35]. Calibration should additionally verify whether claims assigned, for example, a 70% leakage probability exhibit approximately comparable observed leakage frequencies [36].

5.2 Model Performance Across Leakage Mechanisms

Model comparison should examine which revenue-loss mechanisms each algorithm captures rather than ranking models solely by a single discrimination statistic. Logistic regression provides a transparent baseline for relationships such as increasing denial history, payment variance, or coding inconsistency and can support straightforward interpretation by revenue-integrity teams [31]. Random forest is better positioned to capture interactions such as particular procedure-modifier combinations producing abnormal reimbursement only for specific payers or service locations [37]. Gradient boosting and XGBoost can progressively model more complex relationships involving contractual adjustments, resubmission history, authorization status, coding patterns, and payer-specific payment behavior [38]. Anomaly-detection approaches serve a different purpose by identifying previously uncommon claim-payment configurations that may lack confirmed historical labels [39]. However, an anomalous claim is not necessarily financially incorrect. Comparative assessment should therefore consider leakage detection, recoverable monetary value, false audit referrals, and interpretability alongside conventional predictive performance [40].

Table 2. Comparative Predictive and Financial Performance of Revenue-Leakage Detection Models

Model	Precision	Recall	PR-AUC	Leakage Value Captured	Recovery Yield per 100 Reviews	False Alerts per 100 Reviews	Audit Workload
Logistic Regression	0.71	0.68	0.69	68.4%	£31,800	29	Moderate
Random Forest	0.82	0.79	0.83	81.7%	£43,600	18	Moderate
Gradient Boosting/XGBoost	0.88	0.85	0.90	89.3%	£52,400	12	Low-Moderate
Anomaly Detection	0.64	0.74	0.67	72.6%	£27,900	36	High

5.3 Financial Impact and Recovery-Oriented Evaluation

For revenue integrity, detecting ten low-value discrepancies may be less valuable than identifying one substantial contractual underpayment. Evaluation should therefore weight detection by recoverable monetary exposure [32]. Dollar-weighted leakage recall measures the proportion of total confirmed recoverable value captured by the model:

$$DWLR = \frac{\sum_{i \in TP} L_i}{\sum_{i \in TP \cup FN} L_i}$$

where L_i denotes the confirmed recoverable value of claim i . This metric differentiates models with similar claim-level recall but substantially different financial consequences [34]. Recovery yield per reviewed claim can further be calculated as:

$$RY = \frac{\sum_{i \in R} A_i}{|R|}$$

where A_i represents revenue actually recovered and R represents investigated claims. Revenue captured within the highest-risk percentile should also be examined because audit teams normally have finite review capacity [36]. A model concentrating substantial recoverable value within its highest-ranked claims can reduce unnecessary coding and claims reviews [38]. Financial evaluation should additionally distinguish potential leakage identified from revenue actually recovered, since detection does not guarantee successful correction, appeal, or payer repayment [39].

5.4 Financially Constrained Threshold Optimization and Risk Prioritization

Operational thresholds should be determined by the financial consequences of investigation rather than applying an arbitrary probability cut-off. For threshold τ , claims exceeding the threshold form an investigation set S_τ . Expected net financial benefit can be represented as:

$$NB(\tau) = \sum_{i \in S_\tau} (\hat{p}_i L_i - C_i)$$

where $\hat{p}_i$ is predicted leakage probability, L_i is estimated recoverable exposure, and C_i represents the cost of reviewing claim i [33]. Threshold selection can then be formulated as:

$$\tau^* = \arg \max_{\tau} NB(\tau)$$

subject to:

$$|S_\tau| \leq B$$

where B represents available audit capacity. This constraint is important because lowering thresholds may identify additional leakage while simultaneously overwhelming coding or reimbursement teams with false alerts [35]. Claims can subsequently be classified as low risk for automated monitoring, medium risk for targeted validation, and high risk for immediate coding review, contractual reconciliation, or payer appeal [37]. High financial exposure may justify investigation even where predicted leakage probability is moderate, whereas small discrepancies may be deprioritized despite high probability [40].

Figure 3. Revenue-Leakage Risk Prioritization and Intervention Framework

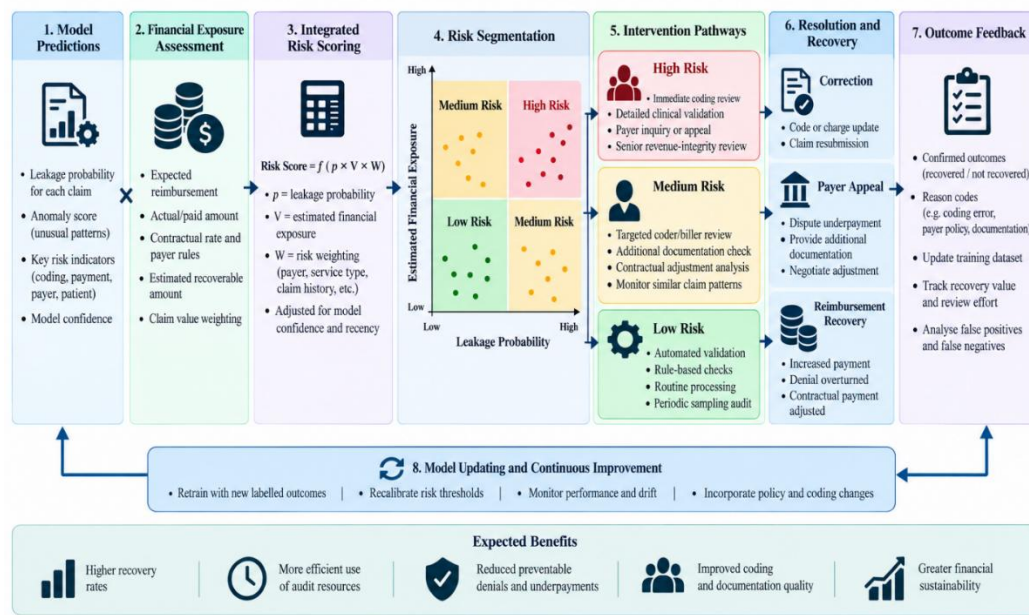


Figure 3. Revenue-Leakage Risk Prioritization and Intervention Framework

This evaluation framework therefore shifts model selection from the question of which algorithm predicts leakage most accurately to the more operationally relevant question of which model directs limited audit resources toward the greatest recoverable revenue. Statistical discrimination remains necessary, but financial concentration, recovery yield, false-review burden, and audit capacity determine whether predictive performance translates into practical revenue-cycle value [39]. These considerations lead directly to Section 6, where the framework is embedded within coding, claims, denial-management, and reimbursement workflows while addressing human oversight, privacy, compliance, and model governance.

6. OPERATIONAL DEPLOYMENT, GOVERNANCE, AND IMPLICATIONS

6.1 Integration Into Healthcare Revenue-Cycle Operations

The practical value of revenue-leakage detection depends on positioning model outputs at points where financial loss can still be prevented or recovered [39]. During coding review, risk scores can identify encounters with unusual diagnosis-procedure relationships, omitted charges, modifier inconsistencies, or documentation-code mismatches before billing [40]. Pre-bill validation can combine these signals with authorization status and expected reimbursement to hold financially material exceptions for targeted review rather than screening every claim manually [41]. Claims-scrubbing workflows can similarly prioritize predicted rejection or denial risks before payer submission. After adjudication, the analytical focus shifts from prospective prevention to retrospective recovery. Payment-variance models can compare contractual expectations with remittance transactions to identify probable underpayments [42], while denial-management systems can rank unresolved claims according to appeal probability and recoverable value. Accounts-receivable surveillance can identify abnormal payment delays, and post-payment audits can detect recurring payer, procedure, or adjustment patterns associated with unrealized revenue [43]. Thus, the framework supports both leakage prevention and targeted financial recovery.

6.2 Automated Detection and Human-in-the-Loop Review

Machine learning should function as an investigative prioritization mechanism rather than an autonomous determination that every reimbursement discrepancy represents an error [44]. A high-risk claim can be dynamically routed according to its predicted leakage source: suspected coding discrepancies to certified coders, authorization or submission failures to billing teams, contractual underpayments to revenue-integrity specialists, and complex reimbursement anomalies to auditors. Routing priority can incorporate both leakage probability and expected recoverable value:

$$P_i^{review} = \hat{p}_i \times L_i$$

where $\hat{p}_i$ represents predicted leakage probability and L_i represents estimated recoverable exposure. Human reviewers can then confirm whether the discrepancy reflects an actual error, legitimate contractual adjustment, benefit restriction, or unusual but valid payment outcome [40]. Confirmed review outcomes should subsequently return to the analytical dataset as feedback, allowing future model development to distinguish actionable leakage from nonrecoverable exceptions more effectively [42].

6.3 Explainability, Privacy, Compliance, and Governance

Revenue-leakage analytics may combine clinical information with identifiable claims and financial records, making privacy and governance integral to deployment rather than secondary technical considerations [41]. Access to protected health information should follow role-based authorization, least-privilege principles, data minimization, secure processing, and auditable access controls [43]. Model outputs should also retain claim-level reason codes so reviewers can determine whether a risk score originated from coding inconsistency, unusual payment variance, repeated denial behavior, contractual deviation, or another identifiable factor [39]. Governance should extend to performance monitoring across payers, providers, service categories, and patient populations because systematic differences in reimbursement patterns can generate biased risk estimates [44]. Coding revisions, contractual renegotiations, payer-policy changes, and reimbursement-rule updates can alter relationships learned from historical data. Model versions, training periods, thresholds, overrides, reviewer decisions, and recovery outcomes should therefore remain auditable [45]. Human oversight is particularly important where an automated recommendation could delay billing, initiate payer disputes, or incorrectly characterize legitimate reimbursement behavior as leakage.

6.4 Practical Implications and Model Limitations

Deployment cannot assume that a model developed in one healthcare organization will retain equivalent performance elsewhere. Provider organizations differ in specialty mix, documentation practices, coding systems, payer composition, contractual arrangements, denial workflows, and reimbursement structures [40]. Data quality also varies across electronic health records, billing platforms, remittance systems, and historical audit repositories [42]. Incomplete labels present a particular challenge because transactions classified as non-leakage may include financial discrepancies that were never investigated. Moreover, concept drift can occur when payer policies, fee schedules, coding standards, or reimbursement models change after model training [45]. Drift in prediction performance can be monitored as:

$$\Delta M_t = M_t - M_{baseline}$$

where M_t represents a selected performance measure during period t and $M_{baseline}$ represents its validated reference performance. Persistent deterioration should trigger investigation, recalibration, or retraining rather than continued automated use [43]. External validation across different provider and payer environments, followed by prospective evaluation within live revenue-cycle workflows, is therefore necessary before broad implementation [44].

These operational requirements demonstrate that successful revenue-leakage detection cannot be judged by predictive performance alone. Its practical value depends on whether predictions reach the correct revenue-cycle function at an actionable point, identify financially meaningful discrepancies, remain understandable to reviewers, and operate within appropriate privacy and governance controls [41]. This operational perspective provides the basis for the conclusion, which considers how machine learning can support a shift from retrospective revenue-loss discovery toward targeted, explainable, and financially prioritized revenue-integrity management.

7. CONCLUSION

7.1 Principal Findings and Analytical Contribution

This study demonstrates that revenue leakage is best understood as a connected analytical problem spanning clinical documentation, coding, claim submission, payer adjudication, and reimbursement rather than as a series of isolated billing errors. Integrating these data domains enables machine-learning models to identify patterns such as coding-payment mismatches, recurrent denial pathways, unexpected contractual adjustments, payer-specific underpayments, and financially significant anomalies that may remain hidden within separate revenue-cycle systems. The principal analytical contribution is the combination of leakage probability, anomaly evidence, and estimated financial exposure within a unified prioritization framework. This directly addresses the study objective by moving revenue-leakage detection beyond retrospective rule-based auditing toward predictive identification of claims where both the likelihood and financial consequence of revenue loss justify investigation.

7.2 Healthcare and Financial Implications

For healthcare organizations, the framework provides a mechanism for directing limited revenue-integrity resources toward discrepancies with the greatest probability of producing recoverable value. Prospective application can strengthen coding validation, claims scrubbing, and denial prevention before revenue is lost, while retrospective application can identify underpayments and prioritize appeals or post-payment audits. Financially weighted prioritization can reduce low-value manual reviews and concentrate specialist effort on claims with greater recovery potential. This approach can improve audit productivity, shorten the interval between reimbursement discrepancy and intervention, strengthen accounts-receivable management, and support financial sustainability without requiring indiscriminate review of every claim or payment variance.

7.3 Future Research Direction

Future research should validate the framework prospectively across different healthcare organizations, specialties, payer populations, and reimbursement environments. Real-time detection should be investigated to determine whether leakage can be intercepted before claim submission or

payment finalization. Adaptive learning and payer-specific models could improve responsiveness to changing contractual and adjudication behavior. Further work should strengthen explainable AI, automated model-drift monitoring, and feedback from confirmed recovery outcomes. Integration with automated revenue-cycle workflows should ultimately be evaluated according to both predictive performance and measurable improvements in recovered revenue, review efficiency, and sustainable financial control.

REFERENCE

1. Zhu C, Attaluri PK, Wirth PJ, Shaffrey EC, Friedrich JB, Rao VK. Current applications of artificial intelligence in billing practices and clinical plastic surgery. *Plastic and Reconstructive Surgery Global Open*. 2024 Jul 1;12(7):e5939.
2. Kilanko V. Leveraging artificial intelligence for enhanced revenue cycle management in the United States. *International Journal of Scientific Advances*. 2023 Jan;4(4):505-14.
3. Shekhar S, Leder-Luis J, Akoglu L. Unsupervised machine learning for explainable health care fraud detection. *National Bureau of Economic Research*; 2023 Feb 13.
4. Ahaiwe, Emmanuel, Samuel Olawole Akande, and Susana Owusu-Ansah. 2026. "Towards Self-Healing Machine Learning Systems: Continual Learning and Drift-Aware MLOps Pipelines for Supply Chain Operations". *Journal of Economics, Management and Trade* 32 (2):1-15. <https://doi.org/10.9734/jemt/2026/v32i21390>.
5. Kilanko V. The transformative potential of artificial intelligence in medical billing: a global perspective. *Int J Sci Adv*. 2023 Jan;4(3):346.
6. Obunadike TC. Leveraging AI forecasting to quantify tariff-induced food price volatility in net-importing nations. *Int J Res Publ Rev*. 2025;6(6):12441-12458. doi:10.55248/gengpi.6.0625.23102.
7. Gabel E, Gal J, Grogan T, Hofer I. A retrospective analysis using comorbidity detecting algorithmic software to determine the incidence of International Classification of Diseases (ICD) code omissions and appropriateness of Diagnosis-Related Group (DRG) code modifiers. *BMC Medical Informatics and Decision Making*. 2024 Oct 23;24(1):309.
8. Vajiram J, Senthil N. Correlating Medi-Claim Service by Deep Learning Neural Networks. arXiv preprint arXiv:2308.04469. 2023 Aug 8.
9. Samuel Addo. Harnessing AI-powered financial forecasting tools to improve risk management, investment strategies, and overall corporate profitability across industries. *International Journal of Science and Research Archive*, 2023, 09(02), 1175-1192. Article DOI: <https://doi.org/10.30574/ijrsra.2023.9.2.0653>
10. Nabrawi E, Alanazi A. Fraud detection in healthcare insurance claims using machine learning. *Risks*. 2023 Sep 5;11(9):160.
11. Gaur M, Pal S, Chaudhuri R, Anto OB, Kalaivanan R. Medical reimbursement prediction using artificial intelligence. In *International Conference on Data Intelligence and Cognitive Informatics 2023 Jun 27 (pp. 327-339)*. Singapore: Springer Nature Singapore.
12. Ogan RK. Predictive anomaly detection for real-time virtual account transactions and API-driven payment processing across enterprise banking platforms. *Int J Comput Appl Technol Res*. 2024;13(12):298-308. doi:10.7753/IJCATR1312.1023.
13. Pendyala SK. Enhancing Healthcare Pricing Transparency: A Machine Learning and AI-Driven Approach to Pricing Strategies and Analytics. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT)*. 2024 Nov;10(6):2334-44.
14. Valiki SV. Detecting Healthcare Billing Anomalies with Graph Neural Networks. *The American Online Journal of Science and Engineering (AOJSE)*. 2024 Dec 1;2(04).
15. Samuel Addo. Causal AI for strategic business planning: uncovering latent drivers of long-term organizational performance and resilience. *World Journal of Advanced Research and Reviews*, 2025, 26(2), 895-912. Article DOI: <https://doi.org/10.30574/wjarr.2025.26.2.1738>
16. Miller AM. AI-Powered Real-Time Claims & Provider Intelligence. *Global Research Development (GRD)*. 2023 Dec 16;1(01).
17. Suesserman M, Gorny S, Lasaga D, Helms J, Olson D, Bowen E, Bhattacharya S. Procedure code overutilization detection from healthcare claims using unsupervised deep learning methods. *BMC Medical Informatics and Decision Making*. 2023 Sep 28;23(1):196.
18. Tosin Emmanuel Folarin. Embedded finance expansion strategies: Scaling contextual financial products across fragmented regulatory and consumer ecosystems. *Int J Finance Manage Econ* 2024;7(1):491-504. DOI: [10.33545/26179210.2024.v7.i1.948](https://doi.org/10.33545/26179210.2024.v7.i1.948)
19. Naga JP, Rao KR, Lakshmi R, Suresh S. Investigating and Identifying Fraudulent Behaviors of Medical Claims Data Using Machine Learning Algorithms. In *Handbook of Artificial Intelligence 2023 Nov 13 (pp. 231-254)*. Bentham Science Publishers.
20. Richardson ER. Detecting Healthcare Billing Fraud with Graph Neural Networks. *Journal of Artificial Intelligence and Big Data Disciplines*. 2024 Mar 15;2(01).
21. Ogan RK. Machine-learning framework for predicting point-of-sale integration failures across digital ordering, payment, and third-party delivery technology ecosystems. *Int J Res Publ Rev*. 2025;6(12):8887-8898.

22. Adeleke O, Ajayi SA. Transforming the healthcare revenue cycle with artificial intelligence in the USA. DOI: <https://doi.org/10.54660/IJMRGE.2024.May.3-1069>.
23. Racheal Kikachukwu Ogan. (2023). AUTOMATED CONFIGURATION COMPLIANCE ANALYTICS FOR DETECTING OPERATIONAL DRIFT AND PREVENTING SERVICE DEGRADATION ACROSS MULTI-LOCATION RESTAURANT TECHNOLOGY SYSTEMS. INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY RESEARCH & MANAGEMENT (IJETRM), 07(12), 807–819. <https://doi.org/10.5281/zenodo.22267240>
24. Iweriebor LE. *Approach to Medicare Provider Fraud Detection and Prevention* (Doctoral dissertation, Capitol Technology University).
25. Buker KL. Financial Impact When a Health System Automates Manual Insurance Verification Processes. Northcentral University; 2023.
26. Twum-Ampofoah, Juliana, Samuel Olawole Akande, Daniel Kofi Yeboah, and Attah Nnaemeka Melford. 2025. "Analytics of Third-Party Risk in U.S. Supply Chains: AI for Proactive Risk Scoring and Mitigation". *Journal of Engineering Research and Reports* 27 (11):178-93. <https://doi.org/10.9734/jerr/2025/v27i111693>.
27. Desai H. Transformative Synergy: Revolutionizing Healthcare Accounting Information Systems with AI. Available at SSRN 5271060. 2024 May 26.
28. Özdemir G. *Machine Learning Based for Insurance Claims Fraud Detection Technique for Safeguarding the Health Industry* (Master's thesis, Istanbul Aydin University (Turkey)).
29. Nayebele FI. Technical framework for real-time U.S. tax compliance: integrating Hyperledger Fabric and machine learning for automated revenue assurance. *Int J Comput Appl Technol Res*. 2026 Jun;13(10):1015. doi:10.7753/IJCATR1310.1015.
30. Carter B. Understanding the Cost of Healthcare Inpatient Services: Prediction of Medicare Inpatient Hospital Costs of Diagnosis-Related Groups. Truman State University; 2023.
31. Hoseini C, Badar MA, Shahhosseini AM, Kluse CJ. Identifying Predictors and Comparing Machine Learning Algorithms to Detect Erroneous Medicaid Adjudication in the USA. In *Handbook of Digital Innovation, Transformation, and Sustainable Development in a Post-Pandemic Era* 2024 Dec 30 (pp. 407-455). CRC Press.
32. Tosin Emmanuel Folarin. Adaptive revenue optimization strategies: Aligning dynamic pricing, interchange economics, and consumer incentives in emerging payment markets. *Int J Res Marketing Manage Sales* 2024;6(2):357-367. DOI: [10.33545/26633329.2024.v6.i2d.443](https://doi.org/10.33545/26633329.2024.v6.i2d.443)
33. Simbert A. *Machine Learning Methods to Detect Medicare Fraud and Abuse in US Healthcare* (Doctoral dissertation, University of Maryland University College).
34. Yoo Y, Shin J, Kyeong S. Medicare fraud detection using graph analysis: A comparative study of machine learning and graph neural networks. *Ieee Access*. 2023 Aug 17;11:88278-94.
35. Racheal Kikachukwu Ogan. Machine-learning-based API failure prediction and root-cause analytics for improving reliability of enterprise banking integration architectures at scale. *Int J Comput Artif Intell* 2021;2(2):142-153. DOI: [10.33545/27076571.2021.v2.i2a.378](https://doi.org/10.33545/27076571.2021.v2.i2a.378)
36. Ünal C, Erbuğa GS. Detection and prevention of medical fraud using machine learning. *Acta Infologica*. 2024 Dec 12;8(2):100-17.
37. Nweke OP. Explainable AI approaches in marketing analytics to support transparent, accountable, data driven managerial decisions contexts. *Int J Comput Artif Intell*. 2023;4(1):89-102. doi:10.33545/27076571.2023.v4.i1a.269
38. Hancock JT, Bauder RA, Wang H, Khoshgoftaar TM. Explainable machine learning models for Medicare fraud detection. *Journal of Big Data*. 2023 Oct 9;10(1):154.
39. Ogan RK. Risk-based API governance for predicting integration failures, security anomalies, and service degradation across hybrid enterprise application environments. *Int J Finance Manage Econ*. 2024;7(2):871-882. doi:10.33545/26179210.2024.v7.i2.946.
40. Adeleke O, Ajayi SA. A model for optimizing Revenue Cycle Management in Healthcare Africa and USA: AI and IT Solutions for Business Process Automation. DOI: <https://doi.org/10.54660/JFMR.2023.Jan.1-505>.
41. Nalluri V, Chang JR, Chen LS, Chen JC. Building prediction models and discovering important factors of health insurance fraud using machine learning methods. *Journal of Ambient Intelligence and Humanized Computing*. 2023 Jul;14(7):9607-19.
42. Balogun M, Fujita T, Yasmin F, Popoola S. Date: 06 February 2023.
43. Folarin TE. Data-centric product growth models: leveraging behavioral analytics to improve retention, monetization, and digital payment adoption. *International Journal of Engineering Technology Research & Management (IJETRM)*. 2023 Dec;7(12):820.
44. Albagmi S, Shawaheen M, Asiri EM, Alqadeer TA. Assessment of clinical miscoding errors and potential financial their implications on healthcare management—A case of local hospital in Najran, Saudi Arabia. *Saudi Pharmaceutical Journal*. 2024 Jan;32(1):1-7.

-
45. Hamid Z, Khaliq F, Mahmood S, Daud A, Bukhari A, Alshemaimri B. Healthcare insurance fraud detection using data mining. *Bmc medical informatics and decision making*. 2024 Apr 26;24(1):112.