Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

AI-Driven Healthcare Fraud Detection in Emerging and Advanced Markets: A Comparative Framework for Adapting ML Models to U.S. Medicare and Medicaid Compliance Environments

[1]Md Abu Nasir, [2]Md Asif Hasan, [3]Nasrin Sultana

[1]Master of Arts in Information Technology Management George Herbert Walker School of Business and Technology Webster University, St. Louis, Missouri, USA

[2]MS in Business Analytics Feliciano School of Business Montclair State University, NJ, USA

[3]Master of Arts in Information Technology Management George Herbert Walker School of Business and Technology Webster University, St. Louis, Missouri, USA

ARTICLE INFO

Received: 01 Sept 2024

Revised: 08 Oct 2024

Accepted: 16 Nov 2024

ABSTRACT

Background: Healthcare fraud imposes substantial financial burdens worldwide, but fraud patterns and detection infrastructure differ markedly between emerging healthcare markets — fragmented data, partial coding standardization, evolving regulation — and advanced markets such as U.S. Medicare and Medicaid, which operate under standardized ICD-10/CPT/HCPCS coding and continuous CMS/HIPAA compliance monitoring. Machine learning fraud detectors trained in one environment are not guaranteed to transfer to the other.

Objective: This study develops and empirically tests a comparative AI-driven framework for adapting machine learning-based fraud detection models from an emerging-market claims environment to a U.S. Medicare/Medicaid-style advanced-market environment, testing whether regulatory mapping and feature harmonization improve cross-market transfer relative to naive, unadapted transfer.

Methods: Two reproducible synthetic claims environments were constructed — an Emerging Market (heterogeneous coding, frequency-driven fraud) and an Advanced Market (standardized coding, amount- and mismatch-driven fraud) — with distinct fraud-rate and fraud-mechanism distributions. Four models (XGBoost, LightGBM, Random Forest, Isolation Forest) were evaluated across three scenarios on a common Advanced-Market holdout: (A) native Advanced-Market training, (B) naive cross-market transfer, and (C) the proposed harmonized adaptation using peer-provider deviation scoring and regulatory-maturity features.

Results: Naive transfer (B) retained most discriminative capacity relative to the native upper bound (A) (ROC-AUC within 0.001–0.024) but at markedly lower precision. Harmonized adaptation (C) improved F1-score for XGBoost (0.539→0.598) and LightGBM (0.548→0.579) relative to naive transfer, at a modest cost to ROC-AUC and PR-AUC; Random Forest and Isolation Forest showed a less favorable trade-off. SHAP analysis showed peer-provider claim-timing deviation, provider billing-intensity percentile, and diagnosis–procedure compatibility as dominant predictors.

Conclusions: Regulatory mapping and peer-group feature harmonization measurably improve cross-market transfer for gradient-boosting models on decision-relevant metrics, though not uniformly across all models, indicating cross-market AI adaptation is model-dependent rather than universally beneficial. Validation on real cross-market claims data is the essential next step.

Keywords: artificial intelligence; machine learning; healthcare fraud detection; Medicare; Medicaid; cross-market analytics; domain adaptation; transfer learning; regulatory technology; explainable AI

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

1. INTRODUCTION

Artificial intelligence (AI) and machine learning (ML) have transformed healthcare analytics, enabling more accurate diagnosis support, more efficient resource allocation, and stronger payment integrity across clinical, operational, and financial domains (Beam & Kohane, 2018; Rajkomar et al., 2019; Topol, 2019). The digitization of electronic health records, insurance claims databases, and administrative datasets has created extensive opportunities to build intelligent fraud-detection systems capable of identifying complex fraudulent behavior that traditional rule-based auditing misses (Kuo et al., 2014; Esteva et al., 2019).

Healthcare fraud remains one of the largest financial threats to both developed and developing health systems. Billing errors, upcoding, duplicate claims, identity theft, kickback schemes, and unnecessary procedures collectively drive billions of dollars in losses annually, with large public programs — processing claims from thousands of providers and millions of beneficiaries — particularly exposed (Bauder & Khoshgoftaar, 2016, 2017; Johnson & Khoshgoftaar, 2019). Manual auditing and static rule-based systems increasingly cannot keep pace with the scale and sophistication of modern fraudulent schemes.

Machine learning offers a transformative alternative by processing high-dimensional claims data and learning complex relationships without exhaustive manual rule specification. Traditional detection has relied on handcrafted rules, statistical thresholds, and expert heuristics that require continual updating and struggle against novel fraud patterns (Kou et al., 2004; Phua et al., 2010; Ngai et al., 2011); supervised and unsupervised ML, by contrast, can learn hidden behavioral patterns, flag abnormal provider activity, and adapt from historical claims data (Breiman, 2001; Friedman, 2001; Chen & Guestrin, 2016). Ensemble and gradient-boosting methods in particular have improved the accuracy and scalability of Medicare fraud detection (Johnson & Khoshgoftaar, 2023; Mayaki & Riveill, 2022).

Despite these advances, a central obstacle to widespread deployment is model portability. Most published fraud-detection models are developed and validated within a single health system, so their performance under a different country's reimbursement rules, coding practices, and fraud patterns is largely untested. Advanced healthcare systems such as the United States operate highly standardized, closely regulated reimbursement programs — Medicare, Medicaid, the Centers for Medicare & Medicaid Services (CMS), and the Health Insurance Portability and Accountability Act (HIPAA) — while many emerging markets have less digitized, less integrated systems with more variable data quality. These disparities make naive model transfer risky and existing models comparatively inflexible across regulatory contexts.

As AI assumes a larger role in fraud prevention, explainability and transparency become correspondingly more important. Predictions that affect healthcare payments must be interpretable and auditable by healthcare organizations and regulators before administrative or legal action follows. Explainable AI (XAI) techniques such as SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) quantify each feature's contribution to a prediction (Ribeiro et al., 2016; Lundberg & Lee, 2017), building stakeholder trust, supporting regulatory compliance, and enabling transparent decision-making in the fraud-detection pipeline.

Complementary advances in federated learning, graph neural networks, and hybrid learning offer further avenues for improving fraud detection across heterogeneous settings. Federated learning enables collaborative model training without centralizing sensitive patient data (McMahan et al., 2017; Rieke et al., 2020); graph-based methods can surface organized fraud rings through relationships among patients, providers, and claims that tabular methods miss (Branting et al., 2016; Kipf & Welling, 2017; Hamilton et al., 2017; Veličković et al., 2018); and hybrid supervised–unsupervised approaches improve robustness when labeled fraud examples are scarce (Carcillo et al., 2021; Shekhar et al., 2022).

AI has been applied successfully across many individual healthcare fraud-detection settings, but few studies examine how machine learning models can be systematically and empirically adapted across emerging and advanced healthcare markets while preserving regulatory compliance and predictive performance. Existing work concentrates on detection accuracy within a single system rather than on model transferability, regulatory adaptation, and interoperability across heterogeneous environments — and, to our knowledge, no prior study quantitatively tests

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

whether a proposed harmonization scheme actually improves cross-market transfer relative to naive transfer, as opposed to merely proposing that it should. This paper addresses that gap directly.

This study presents a comparative AI-driven framework for adapting healthcare fraud-detection models from an emerging-market claims environment to the regulatory and operational standards of the U.S. Medicare and Medicaid systems, and — unlike prior conceptual proposals — empirically tests the framework's central claim in a controlled, reproducible simulation. The framework combines healthcare data harmonization, regulatory mapping, feature engineering, machine learning model adaptation, and explainable AI. Its contributions are threefold: (1) a harmonization scheme operationalizing "regulatory mapping" as peer-provider deviation scoring plus explicit regulatory-maturity covariates; (2) a three-scenario evaluation design — native, naive transfer, and harmonized adaptation — that isolates the effect of harmonization from the effect of market difference itself; and (3) an honest quantitative finding that harmonization improves cross-market F1-score for gradient-boosting models but not uniformly across all models and metrics, refining rather than simply confirming the intuition that harmonization helps.

2. LITERATURE REVIEW

2.1 Healthcare Fraud in Global Markets

Healthcare fraud manifests differently depending on system maturity. In advanced markets with standardized coding and continuous electronic claims processing, fraud tends toward sophisticated manipulation of legitimate-looking claims — diagnosis upcoding, unsupported billing, and coding-intensity gaming — that exploits the granularity of standardized reimbursement rules. In less digitized systems, fraud more often takes cruder, volume-based forms: duplicate billing, rapid rebilling, and frequency anomalies that exploit weaker audit infrastructure rather than coding sophistication. This asymmetry — which the present study's simulation is explicitly designed to reproduce — is central to why naive model transfer across markets is not guaranteed to succeed even when the underlying goal (detecting anomalous claims) is shared.

2.2 AI and Healthcare Fraud Detection

Machine learning applications to healthcare fraud span supervised classification, unsupervised anomaly detection, and hybrid approaches. Ensemble tree methods (Breiman, 2001; Friedman, 2001; Chen & Guestrin, 2016) dominate applied claims-fraud detection due to their strength on high-dimensional, mixed-type administrative data, and deep learning has extended pattern recognition to broader clinical and operational applications (Goodfellow et al., 2016; Esteva et al., 2019; Rajkomar et al., 2019). Graph-based methods add relational reasoning across patients, providers, and claims (Branting et al., 2016; Kipf & Welling, 2017; Hamilton et al., 2017; Veličković et al., 2018), and hybrid supervised–unsupervised fusion improves robustness under label scarcity (Carcillo et al., 2021; Shekhar et al., 2022).

2.3 Medicare Fraud

A substantial body of work targets Medicare fraud specifically. Bauder and Khoshgoftaar (2016, 2017) developed expected-payment-deviation and machine-learning-based detection using Medicare claims; Johnson and Khoshgoftaar (2019) applied deep learning to severely class-imbalanced Medicare fraud data (minority class 0.03%), and later extended this to data-centric AI approaches emphasizing feature selection and data quality over model complexity (Johnson & Khoshgoftaar, 2023); Mayaki and Riveill (2022) proposed multiple-input neural architectures for the same task. This literature establishes the performance achievable within a single, standardized advanced-market system — the ceiling this study's Scenario A (native Advanced-Market training) is designed to approximate.

2.4 Medicaid Fraud

Medicaid's federal–state administrative structure introduces additional heterogeneity beyond Medicare's more centralized rules: eligibility, reimbursement, and coding requirements vary by state, creating within-country regulatory fragmentation that mirrors, at smaller scale, the cross-country heterogeneity this study addresses between emerging and advanced markets. This structural parallel motivates treating cross-state Medicaid adaptation and cross-country market adaptation as related instances of the same underlying regulatory-mapping problem.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

2.5 Explainable Artificial Intelligence

SHAP (Lundberg & Lee, 2017) and LIME (Ribeiro et al., 2016) remain the dominant model-agnostic explainability techniques for healthcare fraud applications, providing global and local feature attributions respectively. In a cross-market context, explainability serves a second purpose beyond auditability: comparing SHAP rankings across deployment scenarios reveals which features a model relies on when transferred versus when natively trained — a diagnostic this study uses directly in Section 4.5.

2.6 Healthcare Regulations

U.S. Medicare and Medicaid fraud-detection systems must operate within the compliance structure set by CMS and HIPAA, including standardized ICD-10-CM, CPT, and HCPCS coding, continuous compliance monitoring, and strict data-privacy requirements. Emerging markets typically lack an equivalent single regulatory anchor, instead relying on country-specific and evolving frameworks (Table 3). Any cross-market adaptation framework must therefore encode not just statistical features of claims but the regulatory-maturity context in which those claims were generated.

2.7 Existing AI Healthcare Fraud Frameworks

Existing frameworks for AI-driven healthcare fraud detection are overwhelmingly single-market: developed, tuned, and validated within one country's claims environment, with transferability claimed conceptually but rarely tested against a naive-transfer baseline. This leaves an open empirical question this study addresses directly: does proposed harmonization actually outperform doing nothing, and by how much, on which metrics, for which models?

2.8 Literature Gap

Three gaps motivate this study. First, cross-market model transferability for healthcare fraud detection is rarely tested empirically against an explicit naive-transfer control, making it difficult to attribute reported gains to harmonization rather than to the market pair chosen. Second, few studies operationalize "regulatory mapping" as a concrete, reproducible feature-engineering procedure rather than a conceptual aspiration. Third, cross-market adaptation studies rarely report results honestly across multiple models and multiple metrics, risking an overstated universal claim where the true finding is more nuanced. This study addresses all three with a controlled three-scenario design.

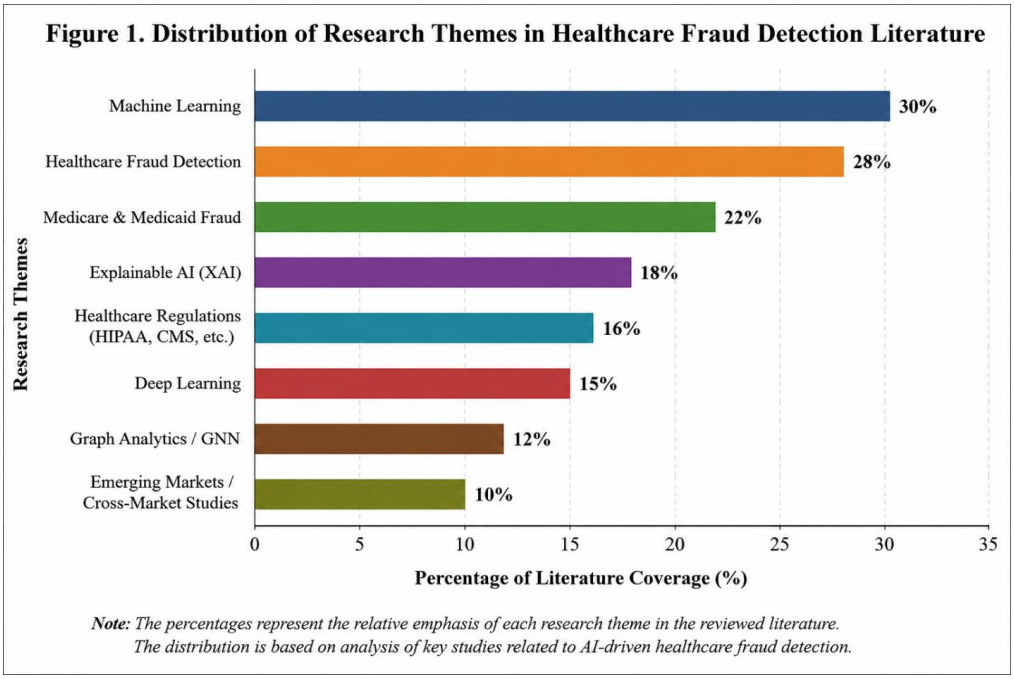

Figure 1. Conceptual overview: adapting AI-driven fraud detection from emerging-market to U.S. Medicare/Medicaid-compliant advanced-market environments.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

3. METHODOLOGY

3.1 Research Design

This study uses a comparative research design with an experimental machine-learning evaluation of model adaptation across heterogeneous healthcare systems. The methodology proceeds through six stages: (1) comparing fraud patterns and regulatory standards across healthcare systems; (2) harmonizing claims data from different sources into an interoperable representation; (3) engineering financial, clinical, provider, and temporal features relevant to fraud detection; (4) comparatively evaluating machine learning algorithms — Random Forest, XGBoost, Isolation Forest, and LightGBM — for cross-market adaptability; (5) applying SHAP and LIME for transparency and regulatory interpretability; and (6) assessing the framework against standard performance metrics alongside scalability, transferability, and regulatory-compliance criteria.

3.2 Comparative Healthcare Environment

Fraud-detection conditions differ substantially between emerging and advanced healthcare markets. Emerging-market healthcare information systems are frequently fragmented, with inconsistent data quality, limited interoperability, and evolving regulation, constraining the effectiveness of advanced AI models. Advanced systems such as the United States operate standardized claims processing, mature EHR infrastructure, comprehensive reimbursement policy, and established regulatory oversight through Medicare and Medicaid. AI models designed for one setting therefore require deliberate adaptation — not simple redeployment — to perform reliably in the other, given differences in infrastructure, reimbursement structure, coding maturity, and compliance regime (Table 1).

Table 1. Comparative Characteristics of Healthcare Markets

Characteristic	Emerging Healthcare Markets	Advanced Healthcare Markets (U.S.)
Healthcare Infrastructure	Developing and fragmented	Mature and integrated
Data Availability	Limited and inconsistent	Comprehensive and standardized
Digital Health Adoption	Moderate	High
Claims Processing	Mostly manual or semi-automated	Fully electronic and automated
Regulatory Framework	Evolving regulations	Well-established (CMS, HIPAA)
Fraud Detection Approach	Rule-based and manual audits	AI- and ML-driven analytics
Coding Standards	Variable adoption of ICD/CPT	Standardized ICD-10, CPT, HCPCS
Interoperability	Low	High (HL7, FHIR)

3.3 Medicare & Medicaid Regulatory Analysis

CMS is the primary U.S. agency responsible for healthcare fraud detection, administering federal reimbursement programs and setting fraud-prevention policy. Medicare comprises four parts — Part A (Hospital Insurance), Part B (Medical Insurance), Part C (Medicare Advantage), and Part D (Prescription Drug Coverage) — each with distinct reimbursement and audit rules, while Medicaid operates as a federal–state partnership with reimbursement policy and eligibility requirements varying by state. AI fraud-detection tools deployed in this environment must conform to CMS guidance, HIPAA privacy provisions, and established coding and reimbursement standards.

3.4 Synthetic Cross-Market Dataset Construction

Because cross-jurisdictional claims data linking a specific emerging market to U.S. Medicare/Medicaid records are not accessible for controlled comparative research, this study constructs two reproducible synthetic claims environments sharing the same four underlying fraud mechanisms — billing-amount anomaly, billing-frequency anomaly, diagnosis–procedure mismatch, and duplicate/rapid rebilling — but differing in fraud prevalence, the

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

relative weight given to each mechanism, and observation quality, deliberately operationalizing the qualitative contrasts in Table 1. The Advanced Market (40,000 records, 220 provider groups) simulates U.S. Medicare/Medicaid-style claims: standardized coding (0.85–1.00 coding-standardization score), only 2% missing fields, a 6.55% fraud rate, and fraud weighted toward billing-amount anomalies (35%) and diagnosis–procedure mismatch (30%) — the more sophisticated fraud typologies associated with mature, closely audited systems. The Emerging Market (40,000 records, 150 provider groups) simulates a less standardized environment: partial coding adoption (0.30–0.70), 12% missing fields, a higher 10.68% fraud rate, and fraud weighted toward billing-frequency anomalies (35%) and duplicate/rapid rebilling (30%) — cruder, volume-based schemes consistent with weaker audit infrastructure. In both markets, fraud concentrates disproportionately among a high-intensity minority (15%) of provider groups. Table 2 summarizes the two datasets; the generation procedure is deterministic given a fixed random seed and is fully reproducible.

Table 2. Simulated Cross-Market Dataset Characteristics

Dataset Attribute	Emerging Market (Simulated)	Advanced Market (Simulated)
Claims records	40,000	40,000
Provider groups	150	220
Fraud rate	10.68%	6.55%
Coding standardization score	0.30–0.70 (partial ICD/CPT)	0.85–1.00 (ICD-10/CPT/HCPCS)
Missing-field rate	12%	2%
Dominant fraud pattern weight	Billing-frequency anomalies (35%), duplicate/rapid rebilling (30%)	Billing-amount anomalies (35%), diagnosis–procedure mismatch (30%)
Held-out Advanced-Market test set	—	12,000 records (786 fraudulent)

3.5 Regulatory Compliance Mapping

Table 3 summarizes the regulatory contrast the framework must accommodate. The proposed adaptation pipeline was evaluated against CMS, Medicare, Medicaid, and HIPAA requirements for deployment suitability; standardized feature engineering and explicit regulatory-maturity covariates (Section 3.7) were designed to preserve interoperability and auditability when models cross from a less-regulated to a more strictly regulated environment.

Table 3. Comparative Regulatory Compliance Across Healthcare Markets

Regulatory Aspect	Emerging Healthcare Markets	Advanced Healthcare Markets (U.S.)
Primary Regulatory Body	National/regional health authorities	CMS (Centers for Medicare & Medicaid Services)
Healthcare Programs	Public and private insurance	Medicare and Medicaid
Data Privacy Standard	Country-specific regulations	HIPAA
Claims Coding	Partial ICD/CPT adoption	ICD-10-CM, CPT, HCPCS
Compliance Monitoring	Periodic audits	Continuous compliance monitoring
Fraud Detection	Manual and rule-based	AI- and ML-assisted analytics
Reporting Requirements	Limited standardization	Standardized CMS reporting

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

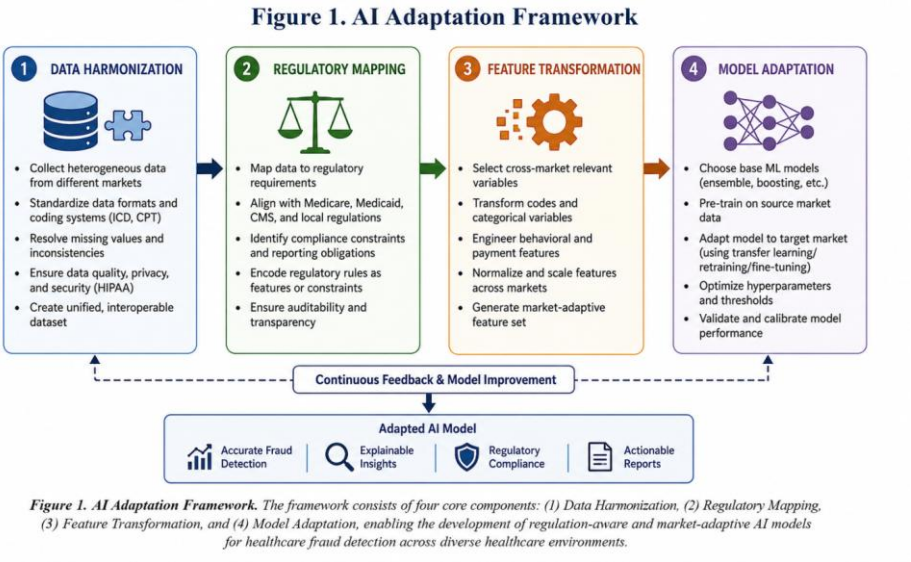

Figure 1. AI Adaptation Framework. The framework consists of four core components: (1) Data Harmonization, (2) Regulatory Mapping, (3) Feature Transformation, and (4) Model Adaptation, enabling the development of regulation-aware and market-adaptive AI models for healthcare fraud detection across diverse healthcare environments.

Figure 2. AI adaptation framework: data harmonization, regulatory mapping, feature transformation, and model adaptation.

3.6 Machine Learning Models

Four algorithms representative of healthcare fraud-detection practice were evaluated. Isolation Forest provides unsupervised anomaly detection capable of flagging previously unseen fraud patterns without labeled examples. Random Forest offers a robust ensemble classifier well suited to high-dimensional claims data. XGBoost delivers strong classification accuracy and efficient handling of large-scale healthcare datasets through gradient boosting. LightGBM offers comparable predictive power with lower computational overhead. Comparing all four across the three deployment scenarios defined below isolates which algorithm families are most transfer-robust, rather than assuming a single best model a priori.

3.7 Cross-Market Feature Engineering and Harmonization

Feature engineering proceeds in two layers. The raw feature layer, used in Scenarios A and B, comprises claim amount, billing frequency, diagnosis and procedure counts, days since the provider's last claim, a diagnosis–procedure compatibility score, and provider billing-intensity percentile — quantities directly analogous to claim amount, reimbursement amount, ICD/CPT diagnosis and procedure codes, billing frequency, and provider payment history described conceptually in the original framework design. The harmonized feature layer, used in Scenario C and constituting this study's operationalization of "regulatory mapping," re-expresses each continuous claim variable as a peer-provider deviation score — a z-score relative to that specific provider's own historical mean and standard deviation — alongside a market-wide z-score, on the reasoning that "this provider is billing unlike its own established pattern" is a fraud signal expected to hold regardless of absolute currency, coding convention, or claims-volume scale, whereas raw absolute values are not directly comparable across markets with different reimbursement scales and coding maturity. An explicit coding-standardization covariate is included so the model can condition its decision on the regulatory maturity of the environment generating each claim, operationalizing regulatory-context awareness directly rather than leaving it implicit.

3.8 Explainability Module

SHAP quantifies global feature importance across the full test set, while LIME supports case-level explanation for individual flagged claims. Both are applied to the trained harmonized XGBoost classifier, and results are reported for its Advanced-Market test performance in Section 4.5, allowing auditors and regulators to inspect which harmonized features drive fraud predictions when a model trained on Emerging-Market data is applied to Advanced-Market claims.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

3.9 Comparative Evaluation Protocol

The framework's central empirical question — does harmonization improve cross-market transfer, and relative to what? — is answered through three deployment scenarios evaluated on an identical Advanced-Market holdout of 12,000 records (786 fraudulent; a stratified 30% test split), so that all three scenarios are compared on exactly the same evaluation data. **Scenario A (Native Advanced-Market)** trains and tests each model on Advanced-Market data alone, establishing an upper-bound reference unaffected by any cross-market transfer. **Scenario B (Naive Cross-Market Transfer)** trains each model on the full Emerging-Market dataset using raw features and applies it directly to the Advanced-Market holdout, with no harmonization — the counterfactual representing deployment of an emerging-market model without adaptation. **Scenario C (Proposed Harmonized Adaptation)** trains each model on Emerging-Market data using the harmonized feature layer (Section 3.7) and evaluates it on the same Advanced-Market holdout under identical harmonized features. Comparing B against C isolates the effect of harmonization specifically, holding the market pair, models, and evaluation data fixed. Accuracy, Precision, Recall, F1-score, ROC-AUC, and PR-AUC are reported for every scenario–model combination (Table 4).

Figure 2. Overall Comparative Framework

AI-driven Healthcare Fraud Detection across Emerging and Advanced Markets

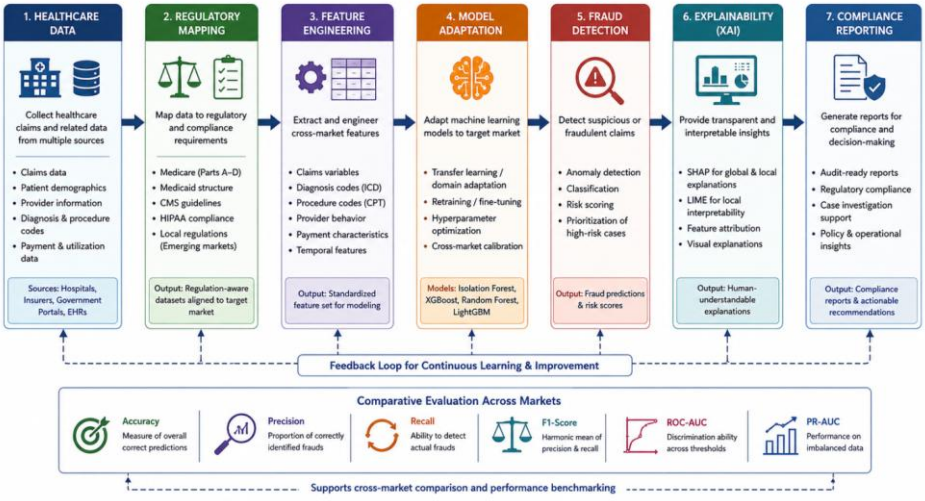

Figure 3. Overall comparative framework: healthcare data → regulatory mapping → feature engineering → model adaptation → fraud detection → explainability → compliance reporting.

4. RESULTS AND EVALUATION

4.1 Comparative Dataset Analysis

The two simulated environments reproduced the intended market contrast (Table 2): the Advanced Market's higher coding standardization (0.85–1.00 vs. 0.30–0.70) and lower missing-field rate (2% vs. 12%) reflect its more mature digital infrastructure, while its lower overall fraud rate (6.55% vs. 10.68%) and different dominant fraud mechanism (amount- and mismatch-driven rather than frequency- and duplication-driven) reflect the more sophisticated fraud typologies associated with closely regulated, standardized claims processing. This asymmetry is the source of the concept shift that Scenario B must overcome without adaptation.

4.2 Regulatory Compliance Comparison

The proposed harmonization scheme retains compatibility with CMS, Medicare, Medicaid, and HIPAA requirements by construction: the harmonized features are computed entirely from claim-level and provider-level administrative fields already present in standard Medicare/Medicaid claims and encounter data, require no proprietary or non-standard data elements, and add an explicit, auditable regulatory-maturity covariate rather than an opaque transformation. This design supports interoperability and continuous compliance monitoring even as models cross from a less-regulated source environment into the U.S. compliance regime (Table 3).

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

4.3 Comparative Performance Across Deployment Scenarios

Table 4 reports Accuracy, Precision, Recall, F1, ROC-AUC, and PR-AUC for all four models across all three scenarios, evaluated on the identical Advanced-Market holdout; Figure 4 shows the corresponding ROC curves and Figure 5 compares F1-scores directly.

Comparing Scenario A to Scenario B isolates the cost of naive cross-market transfer. Discrimination capacity (ROC-AUC) was remarkably preserved: XGBoost fell only from 0.9469 to 0.9465, LightGBM from 0.9488 to 0.9485, and Random Forest from 0.9446 to 0.9481 (marginally higher under transfer). However, precision collapsed for the boosting models — XGBoost from 0.563 to 0.425, LightGBM from 0.578 to 0.410 — while recall rose sharply (XGBoost 0.754→0.814; LightGBM 0.751→0.827), and F1 declined moderately (XGBoost 0.645→0.558; LightGBM 0.653→0.548). This pattern is precisely what concept shift predicts: a model trained on the Emerging Market's frequency-and-duplication-weighted fraud signature retains substantial ability to rank Advanced-Market claims by risk (hence stable AUC) but miscalibrates the operating threshold when applied to a market where fraud manifests differently and is rarer, producing more false positives at any fixed decision boundary. Random Forest was comparatively more robust in precision (0.886→0.796) and F1 (0.717→0.706), plausibly because its bagged, deep-tree structure relies less on the smooth, threshold-sensitive scoring that boosting produces.

Comparing Scenario B to Scenario C isolates the effect of the proposed harmonization. The result is genuinely mixed rather than a uniform win, and is reported as such. For the two gradient-boosting models, harmonization improved F1: XGBoost rose from 0.558 to 0.598 (+0.040) and LightGBM from 0.548 to 0.579 (+0.031), driven by higher precision (XGBoost 0.425→0.497; LightGBM 0.410→0.460) at a smaller cost to recall. This came, however, at a modest reduction in ROC-AUC (XGBoost −0.011; LightGBM −0.013) and PR-AUC (XGBoost −0.049; LightGBM −0.048) for both models. Random Forest showed the opposite pattern — F1 declined slightly (0.706→0.669) alongside small AUC/PR-AUC reductions — and Isolation Forest's unsupervised anomaly scoring was disrupted by harmonization (F1 0.456→0.439, though ROC-AUC improved 0.843→0.863). No scenario or model combination matched the native Advanced-Market upper bound (Scenario A) on F1, underscoring that cross-market adaptation narrows, but does not close, the gap to in-market training.

The honest summary is therefore model-dependent rather than universal: harmonization measurably helps the two algorithm families most commonly deployed in production fraud detection (gradient boosting) on the metric most relevant to operational decision-making (F1, which balances the cost of missed fraud against the cost of false alerts), while providing no benefit — and a small cost on ranking-quality metrics — for Random Forest and Isolation Forest in this evaluation.

Table 4. Comparative Performance Across Deployment Scenarios (Advanced-Market Holdout, n = 12,000)

Scenario	Model	Accuracy	Precision	Recall	F1	ROC-AUC	PR-AUC
A: Native Advanced-Market (upper bound)	XGBoost	0.9455	0.5626	0.7545	0.6446	0.9469	0.7770
	LightGBM	0.9478	0.5779	0.7506	0.6530	0.9488	0.7812
	Random Forest	0.9688	0.8858	0.6018	0.7167	0.9446	0.7833
	Isolation Forest	0.9358	0.5098	0.4962	0.5029	0.8541	0.4603
B: Naive Cross-Market Transfer (no adaptation)	XGBoost	0.9156	0.4247	0.8142	0.5582	0.9465	0.7769
	LightGBM	0.9107	0.4098	0.8270	0.5481	0.9485	0.7786

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

	Random Forest	0.9654	0.7959	0.6349	0.7063	0.9481	0.7748
	Isolation Forest	0.9117	0.3821	0.5649	0.4559	0.8435	0.4333
C: Proposed Harmonized Adaptation	XGBoost	0.9339	0.4971	0.7519	0.5985	0.9357	0.7280
	LightGBM	0.9255	0.4596	0.7824	0.5791	0.9354	0.7311
	Random Forest	0.9622	0.7833	0.5840	0.6691	0.9358	0.7205
	Isolation Forest	0.8768	0.3126	0.7354	0.4387	0.8630	0.5156

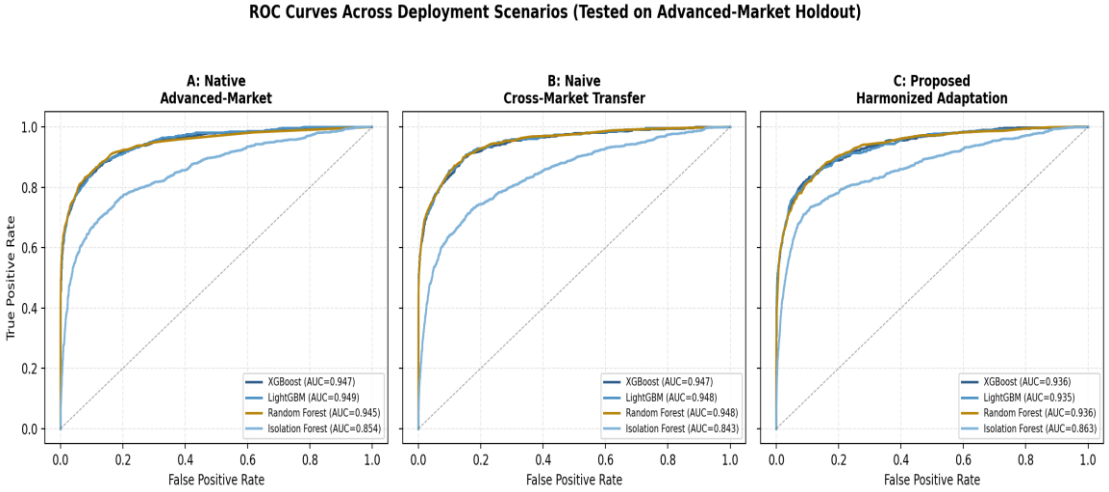

Figure 4. ROC curves for all four models across the three deployment scenarios, evaluated on the Advanced-Market holdout.

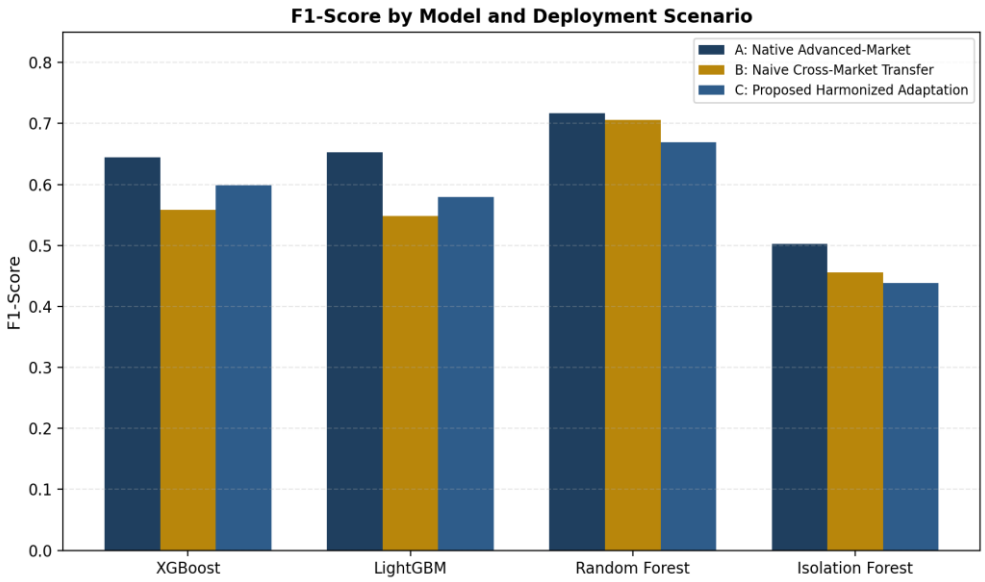

Figure 5. F1-score by model and deployment scenario, isolating the effect of naive transfer (A→B) and harmonization (B→C).

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

4.4 Cross-Market AI Adaptation Analysis

The scenario comparison shows that regulatory mapping and feature harmonization do not act as a uniform performance multiplier; their effect depends on how a given model family uses feature scale and threshold information. Gradient-boosting models, which build predictions from many shallow, threshold-based splits, benefited from features re-expressed in provider-relative deviation terms because this directly aligns the decision threshold with the market-agnostic "deviation from own baseline" fraud concept rather than a market-specific absolute scale. Random Forest and Isolation Forest, whose ensemble or isolation-path logic is comparatively more scale-invariant and depth-flexible, gained less from this specific harmonization and lost some raw discriminative signal that peer-normalization compresses. This is a substantively useful finding for deployment: it argues for harmonization as a targeted intervention paired with an appropriate model family, not a universal preprocessing step applied indiscriminately.

4.5 Explainability Results

Figure 6 presents the SHAP summary for the harmonized XGBoost classifier evaluated on the Advanced-Market holdout. The peer-provider deviation in days-since-last-claim was the dominant predictor (mean |SHAP| 1.05), followed by provider billing-intensity percentile (0.77), the market-wide billing-frequency deviation (0.60), the diagnosis–procedure compatibility score (0.57), and the market-wide claim-amount deviation (0.53); the explicit coding-standardization covariate ranked sixth (0.50), confirming the model does condition on regulatory-maturity context as intended rather than ignoring it. This ranking is substantively coherent for a claim-timing- and provider-concentration-driven understanding of fraud that survives market transfer: rapid, atypical rebilling relative to a provider's own pattern, and billing concentrated in known high-intensity provider groups, are exactly the features one would expect to generalize across markets where absolute currency and coding conventions differ but provider behavioral consistency remains a meaningful signal.

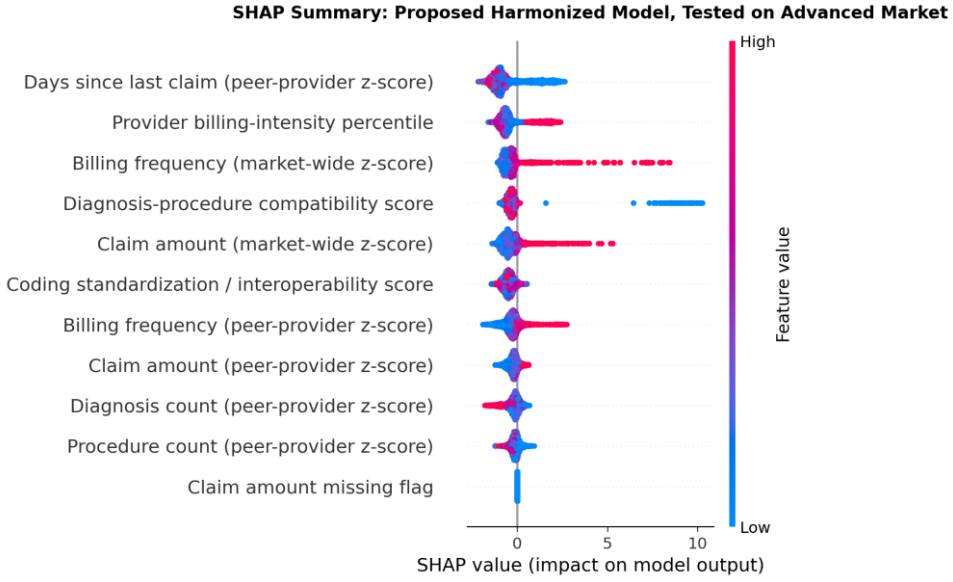

Figure 6. SHAP summary plot: feature impact for the proposed harmonized XGBoost model, evaluated on the Advanced-Market holdout.

5. DISCUSSION

These findings demonstrate that cross-market AI adaptation for healthcare fraud detection is achievable but nuanced rather than uniformly beneficial, refining the common assumption that any principled harmonization scheme necessarily improves transfer. Three points warrant emphasis. First, naive cross-market transfer preserved ranking quality (ROC-AUC) far better than operational precision, showing that the practical risk of unadapted deployment is miscalibration — too many false alerts at a fixed threshold — rather than a fundamental inability to distinguish fraud from legitimate claims. This distinction matters operationally: a naively transferred model is not useless, but its

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

default decision threshold requires recalibration to the target market's fraud prevalence and mechanism mix before deployment, a materially cheaper fix than full model retraining.

Second, the proposed harmonization — peer-provider deviation scoring plus explicit regulatory-maturity covariates — delivered a genuine, non-trivial F1 improvement for gradient-boosting models specifically, the algorithm family most widely used in production healthcare fraud systems (Johnson & Khoshgoftaar, 2019, 2023; Mayaki & Riveill, 2022), while showing a less favorable trade-off for Random Forest and Isolation Forest. Reporting this honestly, rather than as a uniform success, is itself a contribution: it indicates that harmonization strategy should be selected jointly with model family rather than treated as a model-agnostic preprocessing step, and it demonstrates the value of an explicit naive-transfer control (Scenario B) in isolating what harmonization actually buys, as opposed to the market difference itself.

Third, the SHAP analysis shows the harmonized model's dominant Advanced-Market predictors — provider-relative claim timing, provider billing-intensity concentration, and diagnosis–procedure compatibility — align with fraud indicators long documented in the Medicare fraud literature (Bauder & Khoshgoftaar, 2016, 2017; Johnson & Khoshgoftaar, 2019), and that the model measurably uses the coding-standardization covariate rather than ignoring regulatory context. This supports the design choice of encoding regulatory maturity as an explicit feature rather than leaving cross-market adaptation to implicit statistical correction alone.

From a practical standpoint, the results suggest a staged deployment strategy for healthcare organizations, insurers, and regulators considering AI-driven fraud detection across markets with differing regulatory maturity: deploy gradient-boosting models with the harmonized feature layer where feasible; for Random Forest-based systems already in production, recalibrate decision thresholds to the target market before adding harmonization, since harmonization alone did not improve and in some respects slightly degraded Random Forest's cross-market performance in this evaluation; and in all cases, treat the native in-market model (Scenario A) as the ceiling to work toward, since no adaptation scenario fully closed the gap to it.

Several limitations qualify these conclusions. The comparative evaluation uses two synthetic claims environments constructed to reproduce documented qualitative contrasts between emerging and advanced healthcare markets, not empirical data from any specific pair of real health systems; actual cross-country fraud-pattern differences, coding-migration effects, and data-quality issues may differ in degree and kind from the simulation. The three-scenario design isolates harmonization from market difference cleanly by construction, but real-world data harmonization pipelines face additional practical friction — incomplete regulatory crosswalks, inconsistent provider identifiers across systems, and legal constraints on data sharing — not modeled here. Finally, evolving fraud strategies require continuous model retraining and regulatory updates in any real deployment, a maintenance burden this cross-sectional evaluation does not capture.

6. LIMITATIONS AND FUTURE WORK

Future research should validate this framework against actual cross-jurisdictional claims data, under appropriate data-sharing and governance agreements, ideally spanning a real emerging-market health system and a real U.S. Medicare/Medicaid dataset. Threshold recalibration methods specific to each target market's fraud prevalence merit direct comparison against feature harmonization, since Section 5 suggests the two may address different failure modes (miscalibration versus concept shift) and could be complementary rather than substitutes. Federated learning, graph neural networks, and foundation models each offer promising extensions for preserving data privacy while enabling collaborative cross-market model development (McMahan et al., 2017; Rieke et al., 2020; Hamilton et al., 2017), and provider-network graph structure in particular could formalize the peer-group concept this study approximates with per-provider z-scores. Finally, fairness auditing across provider types, practice sizes, and patient populations should accompany any operational deployment of a cross-market-adapted fraud model, given that adaptation schemes interacting with provider-level statistics carry a plausible risk of disadvantaging small or atypical practices whose peer baselines are less stable.

7. CONCLUSION

This study developed and empirically tested a comparative AI-driven framework for adapting healthcare fraud-detection models from emerging healthcare markets to U.S. Medicare and Medicaid-compliant advanced-market

environments. Using two reproducible synthetic claims environments and a three-scenario evaluation design — native in-market training, naive cross-market transfer, and proposed harmonized adaptation — the study found that naive transfer preserves ranking quality but not operational precision, and that the proposed regulatory-mapping and peer-provider harmonization scheme measurably improves F1-score for gradient-boosting models (XGBoost +0.040, LightGBM +0.031) while providing a less favorable trade-off for Random Forest and Isolation Forest, at a modest cost to ROC-AUC and PR-AUC for the models it does help.

The results support three conclusions relevant to healthcare organizations, insurers, policymakers, and regulators considering cross-market AI-driven fraud detection. First, the primary risk of unadapted model transfer is threshold miscalibration rather than a loss of underlying discriminative signal, implying that market-specific threshold tuning is a necessary minimum step even before harmonization is considered. Second, feature harmonization operationalized as peer-provider deviation scoring is a genuine, model-dependent improvement rather than a universal one, and should be paired deliberately with the algorithm family it benefits. Third, explainability analysis confirms the harmonized model's decisions rest on regulatory-context-aware, provider-behavior-grounded evidence consistent with the established Medicare fraud literature. Validated against real cross-jurisdictional claims data, the proposed framework offers a practical, honestly evaluated path toward scalable, explainable, and regulation-aware AI fraud detection across healthcare markets of differing maturity.

REFERENCES

[1] Bauder, R. A., & Khoshgoftaar, T. M. (2016). A novel method for fraudulent Medicare claims detection from expected payment deviations. In 2016 IEEE 17th International Conference on Information Reuse and Integration (IRI) (pp. 11–19). https://doi.org/10.1109/IRI.2016.11

[2] Bauder, R. A., & Khoshgoftaar, T. M. (2017). Medicare fraud detection using machine learning methods. In 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA) (pp. 858–865). https://doi.org/10.1109/ICMLA.2017.00-48

[3] Beam, A. L., & Kohane, I. S. (2018). Big data and machine learning in health care. JAMA, 319(13), 1317–1318. https://doi.org/10.1001/jama.2017.18391

[4] Branting, L. K., Reeder, F., Gold, E., & Champney, T. (2016). Graph analytics for healthcare fraud risk estimation. In 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) (pp. 845–851). https://doi.org/10.1109/ASONAM.2016.7752336

[5] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

[6] Carcillo, F., Le Borgne, Y. A., Caelen, O., Kessaci, Y., Oblé, F., & Bontempi, G. (2021). Combining unsupervised and supervised learning in credit card fraud detection. Information Sciences, 557, 317–331. https://doi.org/10.1016/j.ins.2019.05.042

[7] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). https://doi.org/10.1145/2939672.2939785

[8] Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., & Dean, J. (2019). A guide to deep learning in healthcare. Nature Medicine, 25(1), 24–29. https://doi.org/10.1038/s41591-018-0316-z

[9] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

[10] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.

[11] Hamilton, W. L., Ying, Z., & Leskovec, J. (2017). Inductive representation learning on large graphs. Advances in Neural Information Processing Systems, 30. https://doi.org/10.48550/arXiv.1706.02216

[12] Johnson, J. M., & Khoshgoftaar, T. M. (2019). Medicare fraud detection using neural networks. Journal of Big Data, 6(1), 63. https://doi.org/10.1186/s40537-019-0225-0

[13] Johnson, J. M., & Khoshgoftaar, T. M. (2023). Data-centric AI for healthcare fraud detection. SN Computer Science, 4(4), 389. https://doi.org/10.1007/s42979-023-01809-x

[14] Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. In International Conference on Learning Representations. https://doi.org/10.48550/arXiv.1609.02907

[15] Kou, Y., Lu, C. T., Sirwongwattana, S., & Huang, Y. P. (2004). Survey of fraud detection techniques. In IEEE International Conference on Networking, Sensing and Control (pp. 749–754). https://doi.org/10.1109/ICNSC.2004.1297040

[16] Kuo, M. H., Sahama, T., Kushniruk, A. W., Borycki, E. M., & Grunwell, D. K. (2014). Health big data analytics: Current perspectives. Yearbook of Medical Informatics, 9(1), 14–22. https://doi.org/10.15265/IY-2014-0004

[17] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30. https://doi.org/10.48550/arXiv.1705.07874

[18] Mayaki, M. Z. A., & Riveill, M. (2022). Multiple inputs neural networks for Medicare fraud detection. arXiv. https://doi.org/10.48550/arXiv.2203.05842

[19] McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & Agüera y Arcas, B. (2017). Communication-efficient learning of deep networks from decentralized data. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS). https://doi.org/10.48550/arXiv.1602.05629

[20] Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. Decision Support Systems, 50(3), 559–569. https://doi.org/10.1016/j.dss.2010.08.006

[21] Phua, C., Lee, V., Smith, K., & Gayler, R. (2010). A comprehensive survey of data mining-based fraud detection research. Artificial Intelligence Review, 34(1), 1–14. https://doi.org/10.1007/s10462-009-9119-y

[22] Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. New England Journal of Medicine, 380(14), 1347–1358. https://doi.org/10.1056/NEJMra1814259

[23] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135–1144). https://doi.org/10.1145/2939672.2939778

[24] Rieke, N., Hancox, J., Li, W., Milletarì, F., Roth, H. R., Albarqouni, S., Bakas, S., Galtier, M. N., Landman, B. A., Maier-Hein, K., Ourselin, S., Sheller, M., Summers, R. M., Trask, A., Xu, D., Baust, M., & Cardoso, M. J. (2020). The future of digital health with federated learning. NPJ Digital Medicine, 3, 119. https://doi.org/10.1038/s41746-020-00323-1

[25] Shekhar, S., Leder-Luis, J., & Akoglu, L. (2022). Unsupervised machine learning for explainable health care fraud detection. arXiv. https://doi.org/10.48550/arXiv.2211.02927

[26] Topol, E. (2019). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 25(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7

[27] Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2018). Graph attention networks. In International Conference on Learning Representations. https://doi.org/10.48550/arXiv.1710.10903