Computer Fraud and Security
ISSN (online): 1873-7056

# AI-Driven Fraud Detection in Healthcare: Architecture, Implementation, and Impact

**Sri Venkata Aravindbabu Malempati**

California State University, Los Angeles, USA

**Abstract**

Healthcare fraud remains one of the most costly and structurally persistent threats confronting the United States healthcare system. It diverts immense resources away from legitimate care and erodes trust in the institutions involved. Furthermore, healthcare fraud has outstripped customary rule-based and manual audit processes both in velocity and complexity, and in the adaptive capacity of fraud networks. Artificial intelligence and machine learning-based systems have emerged as an alternative that can look at hundreds of variables across thousands of claims, identify anomalies in real time, and continuously learn as fraud schemes evolve. Technical architectures that include supervised ensemble models, unsupervised anomaly detectors, and graph network analyses have shown improved performance on insurer data. When data quality, algorithm fairness, model explainability, provider due process, and human intervention and monitoring are prioritized, AI-based fraud detection can add long-lasting value to patients, payers, and the healthcare system by recapturing payments to fraudsters and preventing future losses at scale.

**Keywords:** Healthcare Fraud Detection, Machine Learning, Anomaly Detection, Algorithmic Fairness, Claims Analysis

## 1. Introduction

Healthcare fraud remains one of the most meaningful active structural threats to the financial and systemic integrity of the healthcare system in the United States, with the Department of Justice opening over 802 criminal healthcare fraud investigations in fiscal year (FY) 2023. More than 346 criminal cases were filed charging at least 530 defendants, and over 476 healthcare fraud offenders were convicted in FY 2023 alone. [1] Civil health care fraud settlements and judgments under the FCA exceeded $1.8 billion in FY 2023. The FBI and its federal law enforcement Partners conducted over 620 operational disruptions of criminal fraud networks and dismantled over 127 healthcare fraud criminal networks. In addition to the financial impact, health care fraud raises insurance premiums, denies needed patient care, and sometimes harms patients sent for treatments that do not exist, services that are not necessary, or diagnoses that are fabricated. For instance, the HHS Office of Inspector General excluded 2,112 individuals and entities from participation in Medicare, Medicaid, and other Federal healthcare programs during FY 2023, showing the kinds of systemic fraud across the types of providers [1].

Conventional fraud detection infrastructure, dependent on large-scale rule-based systems and human auditors, struggles to adapt to the rapidly changing nature of healthcare transactions. Rule-based engines encode known fraud schemes in conditional logic but are unable to deal with new and evolving enterprise risks, such as those presented by sophisticated tactics that fraudsters employ to evade detection. Fraudsters create their plans to avoid being caught by using tricks like changing billing codes, spreading claims across various providers, and mimicking real billing practices. In FY 2023, the HHS-OIG Chief Data Office applied predictive analytics, geospatial analytics, machine learning, artificial intelligence (AI), including neural networks, and text mining to support 55 audits, 23 evaluations, and 161 criminal investigations in the Medicare and Medicaid portfolios. These levels for investment in data science can assess the level needed to create a real impact on healthcare fraud. In this paper, we look at the full data science-enabled life cycle of AI-enabled healthcare fraud detection (technical infrastructure, governance and regulatory landscape, organizational deployment, empirical evidence, challenges, and future research directions).

## 2. Technical Architecture And Machine Learning Approaches

The use of AI to detect healthcare fraud is supported by data fusion, where data from diverse sources is brought together for high-speed and continuous inference. To ease this, the HHS-OIG maintains the Integrated Data Repository and One Program Integrity platform to provide a single point of access to program integrity contractors, law enforcement, and investigators for consolidated Medicare and Medicaid data, including 9.7 billion encounter data records (as collected by

CMS's Encounter Data System) as of FY 2023 [1]. Production systems produce claims processing systems, electronic medical records, provider licensing and credentialing information systems, pharmacy dispensing systems, and lab results. Previous fraud investigation findings are another source of features. The most consequential step in model building is called feature engineering, the process of converting raw transactional data into features with predictive value. Predictive features considered in this analysis include extreme deviations in billing volume from specialty-adjusted baseline peer group levels, atypical temporal claim clustering, implausibility scores for pairs of diagnosis and procedure codes, networking metrics on provider groups, and outlier scores for beneficiary utilization. Ensemble models trained on such datasets have achieved an AUC-ROC performance of greater than 0.93 in test scenarios, compared to the baseline logistic regression score of 0.71 [3].

CMS's fraud prevention system (FPS) is an example of predictive analytics technology applied at scale, due to the Small Business Jobs Act of 2010. For FY 2023, the FPS system provided 1,137 new leads and 1,994 leads or investigations in progress to the program integrity contractors, who reported taking action against 1,095 providers based on FPS recommendations [1]. The Advanced Provider Screening system automatically screens all existing and potential Medicare providers and suppliers against several data sources. In FY 2023, the Centers for Medicare and Medicaid Services created over 700 criminal alerts based on information about potentially fraudulent providers for further investigation. Based on those alerts, CMS took action to revoke 237 provider enrollments based on felony convictions and 284 based on state medical license actions. Algorithms like gradient boosting machines such as XGBoost and LightGBM, random forests, and deep neural networks can achieve the best predictions when they are trained with labeled data. Algorithms such as isolation forests, autoencoders, and one-class support vector machines, which do not rely on supervised learning, can be useful for cases without historical labels of fraudulent transactions. In several benchmark evaluations, hybrid ensemble models of supervised/unsupervised combinations have been shown to outperform implementations of isolated algorithms by 8-12 percentage points [3].

Technical challenges facing a production system include class imbalance, concept drift, and explainability. In most payer environments, fraud events are a small fraction of claims, so SMOTE oversampling, cost-sensitive learning, and precision-recall optimization can be used. HHS-OIG CDO staff of analysts, data scientists, statisticians, and data engineers employed customized analytics using AI and machine learning in efforts to pursue criminal and civil enforcement for more than $2.1 billion in false claims. HHS-OIG analytic tools were used by more than 850 unique HHS-OIG staff in FY 2023.. Explainability limitations of deep learning architectures make it difficult to defend flagging decisions against provider pushback. In production, SHAP values, gradient-weighted class activation mapping, or simpler proxy model layers are used to produce human-understandable explanations of predictions.

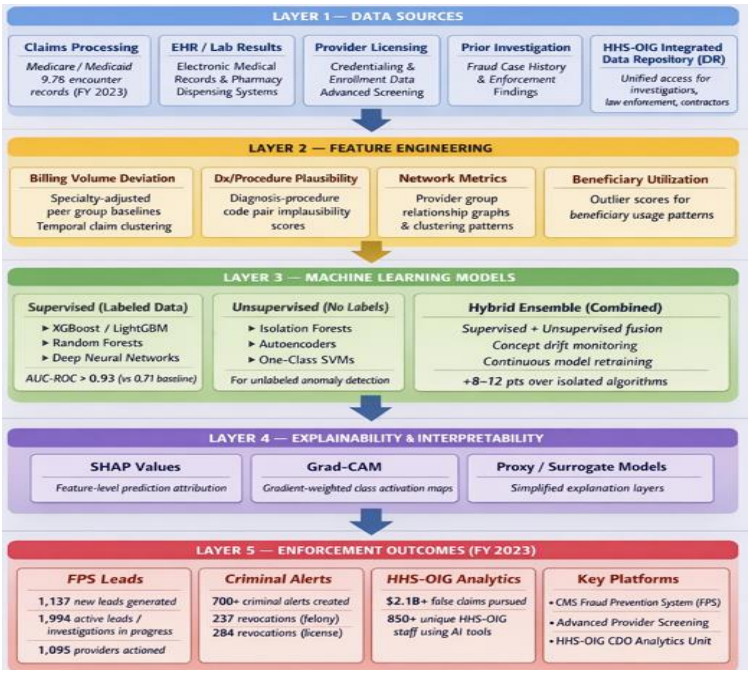


Fig 1: Healthcare Fraud Detection: Technical Architecture [1, 3]

### 3. Regulatory Framework and Compliance

A wide-ranging web of overlapping federal and state laws regulates the use of patient data in healthcare fraud detection and enforcement by establishing patient privacy standards, algorithmic explainability requirements, and provider due process rights. The primary civil enforcement mechanism in the United States for this type of fraud is the False Claims Act, which generated more than $1.8 billion in civil fraud settlements and judgments and hundreds of millions of dollars in qui tam relator payments to whistleblowers in fiscal year 2023. The Anti-Kickback Statute prohibits paying or receiving remuneration to induce someone to refer a patient or to furnish, arrange for, or recommend the purchase of federally reimbursable services or items. In the June 2023 National Health Care Fraud Enforcement Action, 78 defendants were charged with $2.5 billion in fraud, and the department seized or restrained millions of dollars in cash, vehicles, and real estate. While the Health Insurance Portability and Accountability Act Privacy Rule allows a covered entity to use and disclose protected health information for the purpose of detecting healthcare fraud and abuse without patient authorization, the covered entity must use the minimum necessary PHI that is required to accomplish the fraud detection function [5].

Technical and administrative safeguards required by the HIPAA Security Rule for protection of electronic PHI, applicable directly to fraud detection activities, include role-based access to claim-level PHI by analysts, AES-256 encryption of ePHI both in transit and at rest, audit logging of access, and documented incident response. Category and degree of willfulness determine civil monetary penalties for HIPAA violations, capped annually by statutory limits depending on the underlying conduct. HHS-OIG closed cases involving over $82.9 million in civil monetary penalties and assessments in FY 2023, thereby signaling the enforcement climate across the healthcare compliance ecosystem [1]. The Healthcare Fraud Prevention Partnership (HFPP) is a voluntary public-private partnership with 300 partner organizations at the end of FY 2023, including 136 private payers and 50 State Medicaid Agencies, resulting in over 334 billion professional claim lines that AI-based detection systems must process securely and compliantly [1].

The latest regulations from CMS and the Office of the National Coordinator for Health Information Technology are focusing on algorithmic accountability. For example, CMS guidance on fraud detection vendors in Medicare programs requires that these vendors make available in plain language the risk factors used to determine whether to flag an individual claim. Fairness specifications require that fraud detection systems do not produce unreasonably high false positive rates for any provider specialization, geographic area, or patient population. The national Medicaid improper payment rate in FY 2023 was $50.33 billion in gross Federal improper payments, or 8.58% of total program outlays. Insufficient documentation accounted for 81.84% of all improper payments, a property of data quality that is relevant to AI training data sets. In practice, important compliance steps could include written data use policies, BAAs with all PHI processing vendors, fairness impact assessments at deployment and every six months, and a formal provider appeals process with human review authority [6].

| Metric | Value | Context/Significance |
|---|---|---|
| False Claims Act Civil Settlements | $1.8B+ | Civil fraud settlements and judgments generated via 31 U.S.C. § 3729 in FY 2023 |
| Qui Tam Relator Payments | $100 Ms | Hundreds of millions paid to whistleblowers under the False Claims Act in FY 2023 |
| June 2023 Enforcement Action—Defendants | 78 | Defendants charged in the National Health Care Fraud Enforcement Action, June 2023 |
| June 2023 Enforcement Action—Fraud Amount | $2.5B | Total alleged fraud value in the June 2023 national enforcement sweep |
| HHS-OIG Civil Monetary Penalties (FY 2023) | $82.9M | Total civil monetary penalties and assessments closed by HHS-OIG in FY 2023 |
| HFPP Partner Organizations | 300 | Total number of organizations in the Healthcare Fraud Prevention Partnership at the end of FY 2023 |
| HFPP—Private Payers | 136 | Private payer members within the 300-organization HFPP |

| | | |
|---|---|---|
| | | coalition |
| HFPP—State Medicaid Agencies | 50 | State Medicaid Agency members participating in the HFPP |
| Professional Claim Lines Processed | 334B+ | Claim lines that AI-based detection systems must process securely across HFPP partners |
| Medicaid Gross Federal Improper Payments | $50.33B | National Medicaid improper payment total in FY 2023 |
| Medicaid Improper Payment Rate | 8.58% | Improper payments as a percentage of total Medicaid program outlays in FY 2023 |
| Improper Payments—Insufficient Documentation | 81.84% | Share of all improper payments attributable to documentation gaps—a key AI training data quality issue |
| Fairness Assessment Frequency | Every 6 months | Required interval for fairness impact assessments of fraud detection systems post-deployment |

Table 1: Regulatory Framework & Compliance: Key Numerical Data Points [1, 8]

**4. Implementation and Deployment**

Successful AI fraud detection deployment follows a structured, phase-gated development lifecycle beginning with rigorous problem definition and baseline measurement. Over the last three fiscal years (2021–2023), the overall HCFAC Program has generated a return on investment of $2.80 for every $1.00 expended, calculated as a three-year rolling average to smooth annual variation from case settlement and adjudication timing. HHS-OIG's program-specific ROI reached approximately $10.00 for every $1.00 obligated in FY 2023 when expected recoveries, fines, penalties, and stolen and misspent funds are considered against annual HHS-OIG obligations [1]. These benchmarks highlight the indispensability of establishing quantifiable baseline fraud metrics, such as confirmed cases, recovery dollars, false positive rate per investigation, and investigator hours expended per confirmed case, prior to model deployment, in order to demonstrate the value of the program after deployment. Problem definition must also resolve inherent objective tensions: maximizing recall systematically increases false positives, increasing investigator workload and provider friction; minimizing false positives may allow higher-value fraud to persist undetected. Organizations should explicitly prioritize these objectives based on investigative resource availability and regulatory exposure [7].

Exploratory data analysis routinely reveals data quality deficiencies requiring remediation before model training. The FY 2023 HCFAC report identifies insufficient documentation as accounting for 81.84%, or $41.19 billion, of total Medicaid errors—a finding that directly translates into degraded training data quality for supervised models that rely on labeled claims records [1]. Baseline models—typically logistic regression applied to historical labeled data—establish minimum performance thresholds that advanced architectures must exceed to justify additional complexity. The HCF Unit has a data team of nine in-house analysts who are expert users of Medicare/Medicaid data and financial analysis. In FY 2023 alone, the unit processed 2838 requests for data analysis support from US Attorneys' Offices, including providing the infrastructure to translate model outputs into actionable prosecutorial leads [1]. The production deployment involves a shadow mode, typically lasting between 60 and 90 days. The model creates predictions on live claims but does not trigger an investigation. However, the production deployment may be implemented gradually (at 10%, 25%, 50%, and 100% of claim volume) so that organizations can observe precision, recall, and false positive rates at various thresholds and fine-tune the system before full implementation of the classifier [8].

It is clear that the infrastructure needed to support the implementation of any operational automation-assisted detection of possible abuse of data should include the CMS Medicare Major Case Coordination process, which has reviewed over 5000 cases and received over 3100 law enforcement requests [1]. In FY 2023, CMS reviewed 1,106 cases at Major Case Coordination meetings, and law enforcement partners made 538 requests for case referrals—a volume that underscores the requirement for investigative capacity planning calibrated to the expected lead volume generated by automated detection systems. Post-deployment monitoring infrastructure should track AUC-ROC, precision-recall area, and fairness metrics on a daily or weekly basis, with automated alerting when metrics degrade beyond defined tolerances. Investigator

feedback loops—structured workflows through which investigation outcomes are systematically tagged and returned to the model retraining pipeline—are essential to prevent performance degradation as fraud schemes evolve [8].

| Metric | Value | Context/Significance |
|---|---|---|
| HCFAC Program ROI (3-Year Rolling Average) | $2.80: $1.00 | Return on investment for the HCFAC Program averaged across FY 2021–2023 to smooth case settlement timing variation |
| HHS-OIG Program-Specific ROI (FY 2023) | ~$10.00: $1.00 | Expected recoveries, fines, penalties, and misspent funds recovered per dollar obligated by HHS-OIG in FY 2023 |
| HCFAC ROI Reference Period | FY 2021–2023 | Three fiscal years are used as the rolling average window for the HCFAC program ROI calculation |
| Medicaid Errors—Insufficient Documentation Share | 81.84% | Proportion of total Medicaid errors attributable to documentation gaps, directly degrading supervised model training data quality |
| Medicaid Errors—Insufficient Documentation Value | $41.19B | The dollar value of Medicaid errors caused by insufficient documentation in FY 2023 |
| HCF Unit In-House Analysts | 9 | Expert Medicare/Medicaid data and financial analysis analysts on the HCF Unit data team |
| Data Analysis Requests Processed (FY 2023) | 2,838 | Requests for data analysis support received from U.S. Attorneys' Offices, translating model outputs into prosecutorial leads |
| Shadow Mode Deployment Duration | 60–90 days | Typical duration for shadow mode, during which the model scores live claims without triggering investigations |
| Gradual Rollout Thresholds | 10%, 25%, 50%, 100% | Progressive claim volume stages for phased production deployment, allowing precision, recall, and false positive rate monitoring before full rollout |
| CMS Major Case Coordination—Total Cases Reviewed | 5,000+ | Cumulative cases reviewed through the CMS Medicare Major Case Coordination process |
| CMS Major Case Coordination—Law Enforcement Requests (Cumulative) | 3,100+ | Total law enforcement requests received through the CMS Major Case Coordination process |
| CMS Major Case Coordination—Cases Reviewed (FY 2023) | 1,106 | Cases reviewed at Major Case Coordination meetings in FY 2023 alone |
| Law Enforcement Case Referral Requests (FY 2023) | 538 | Requests made by law enforcement partners for case referrals in FY 2023, informing investigative capacity planning |

Table 2: Implementation & Deployment: Key Numerical Data Points [1, 7, 8]

**5. Demonstrated Impact and Case Studies**

The outcome of meaningful enforcement activity in FY 2023 provides a wealth of empirical evidence to assess the impact of large-scale data-driven fraud detection efforts. In FY 2023, more than $3.4 billion was deposited with the

Department of the Treasury and CMS or transferred to other Federal agencies that manage healthcare programs. In particular, Medicare Trust Funds received nearly $974 million in transfers [1]. For FY 2023, final HHS and DOJ allocated mandatory and discretionary resources for the HCFAC program yielded a rolling three-year average annual ROI of $2.80 for every $1.00 spent; this figure continues to be adversely impacted by pandemic-related enforcement delays, as the report states [1]. HHS-OIG's expected recoveries from audits and investigations totaled more than $3.35 billion, and potential savings from legislative and administrative actions supported by HHS-OIG recommendations were estimated at $1.362 billion, of which $821.4 million was in Medicare savings and $541 million in savings to the Federal share of Medicaid [1].

Specific enforcement outcomes illustrate the impact of targeted data analytics at the case level. The HCF Unit's data analytics team identified a COVID-19 laboratory testing scheme that resulted in the conviction of two defendants for a $455 million fraud, with over $15.7 million seized—demonstrating an outsized return on investment attributable directly to data-driven targeting [1]. In the genetic testing domain, the owner of LabSolutions, LLC submitted more than $463 million in claims to Medicare between July 2016 and August 2019, of which Medicare paid over $187 million; the data analytics infrastructure supporting the Operation Double Helix takedown was essential to identifying the scale and structure of this scheme, which ultimately resulted in a 27-year prison sentence [1]. In the telemedicine exploitation domain, the HCF Unit has charged over $11 billion in fraud committed using or exploiting telemedicine over four and a half years, with the Operation Brace Yourself enforcement action alone producing savings of more than $1.9 billion in the amount paid by Medicare for orthotic braces in the 20 months following that enforcement action [1].

The Healthcare Fraud Prevention Partnership's analytical capabilities further illustrate the value of AI-supported cross-payer detection. Over 334 billion professional, institutional, and pharmacy claim lines were submitted by HFPP partners through FY 2023 for cross-payer analyses; studies initiated in FY 2023 examined problematic billing in COVID-19 add-on laboratory testing, excessive telehealth billing, applied behavioral analysis therapy, genetic testing, outlier billing for members with substance use disorders, and evaluation and management of improbable days [1]. HHS-OIG analytics supported 11 Medicare Advantage audits that collectively identified $78 million in overpayments to MA plans, and six Medicaid managed care audits identified more than $49 million in overpayments for deceased beneficiaries and concurrent multi-state enrollments [1]. Across all documented deployments, consistent findings emerge: AI-driven detection substantially amplifies investigative reach, and data analytics support produces outsized enforcement returns relative to the cost of the analytical infrastructure [3].

**6. Challenges and Limitations**

Despite the substantial performance gains documented in production deployments, AI-driven fraud detection systems face a persistent set of technical, organizational, and ethical challenges. Data quality deficiencies represent the most commonly cited barrier to model performance: the FY 2023 HCFAC report identifies insufficient documentation as accounting for 81.84%, or $41.19 billion, of total Medicaid FFS, managed care, and eligibility errors, and 68.05%, or $1.45 billion, of total CHIP errors [1]. This systemic documentation deficiency—missing diagnosis codes, inconsistent provider taxonomy classifications, and structural coding variations across regional payer systems—directly degrades the quality of training datasets on which supervised models depend for accurate fraud classification. Severe class imbalance—fraud constituting fewer than 1% of total claims in most environments—creates a pathological incentive for naive models to classify all claims as legitimate. This class imbalance problem is addressed via algorithmic approaches (e.g., SMOTE oversampling, class-weighted loss functions, and precision-recall optimization), as well as evaluation metrics that are suited for dealing with imbalanced datasets (e.g., F1 score, precision-recall AUC, and Matthews Correlation Coefficient) [4].

Structural concept drift, where the characteristics of a problem change over time, is another challenge not usually addressed in static evaluations. The report for FY 2023 found that fraud schemes transitioned over time from telehealth to genetic testing to COVID-19 relief funds. Such variation is simply a reflection of how fraud networks adapt schemes based on enforcement signals [1]. The Strike Force's stated use of advanced data analysis to identify "aberrant billing levels in health care fraud hot spots—cities with high levels of billing fraud—and target suspicious billing patterns, as well as emerging schemes and schemes that migrate from one community to another" [1] acknowledges that an adaptive system requires retraining. Organizational challenges also remain. The FY 2023 report discussed sequestered mandatory HCFAC funds totaling $264.7 million since FY 2013, including $180.5 million over the past 11 years, leaving limited funding for Strike Force investigators and analysts to process leads generated by AI [1]. Investigation resource

saturation—producing more leads than investigators can handle—is a known operational failure mode. For example, the I-MEDIC contractor started 692 investigations in FY 2023, sent 174 leads to law enforcement, and sent 181 leads to other groups. This shows how hard it is to manage a lot of leads with limited investigative resources [1].

Algorithmic fairness and bias concerns represent the most ethically consequential limitation category. Machine learning models trained on historical fraud investigation data systematically encode the biases present in those investigations. The FY 2023 enforcement report stresses geographic Strike Force cities (Miami, Houston, Detroit, Brooklyn, Los Angeles, and Chicago), which may reflect a geographic concentration of fraud but may also cause geographic overrepresentation in training data [1]. Providers who engage in atypical but clinically justified billing, such as caring for populations with complex medical needs and limited ability to pay or practicing in specialty areas with high baseline costs, could be systematically disadvantaged by peer-comparison features that treat statistical deviation from specialty norms as a fraud signal. Some of these strategies include fairness audits to evaluate false positive disparities by provider type, specialty, geography, and patient demographic grouping; incorporating algorithmic fairness constraints into model training objectives; and publicly available provider appeals policies with human review authority that cannot be overruled by automated algorithms [4].

| Metric / Indicator | Category | Value | Context/Significance |
|---|---|---|---|
| Medicaid Errors—Insufficient Documentation | Data Quality | 81.84% | Share of total Medicaid FFS, managed care, and eligibility errors caused by documentation deficiencies |
| Medicaid Errors—Insufficient Documentation Value | | $41.19B | Dollar value of Medicaid errors attributed to insufficient documentation in FY 2023 |
| CHIP Errors—Insufficient Documentation | | 68.05% | Share of CHIP program errors caused by documentation deficiencies |
| CHIP Errors—Insufficient Documentation Value | | $1.45B | The dollar value of CHIP errors linked to insufficient documentation |
| Fraud Share of Total Claims | Model Training / Class Imbalance | <1% | Fraud typically represents fewer than 1% of total claims, creating severe class imbalance for ML models |
| Sequestered HCFAC Funds (Since FY 2013) | Organizational/Funding | $264.7M | Total HCFAC funds sequestered since FY 2013, limiting fraud enforcement capacity |
| Sequestered Funds (Last 11 Years) | | $180.5M | Portion of sequestered funds accumulated over the past 11 years |
| I-MEDIC Investigations Initiated | Operational Capacity | 692 | Investigations started by the I-MEDIC contractor in FY 2023 |
| I-MEDIC Leads Sent to Law Enforcement | | 174 | Fraud leads escalated to law enforcement agencies |
| I-MEDIC Leads Sent to Other Groups | | 181 | Leads referred to other enforcement or oversight groups |

Table 3: Key Quantitative Indicators of Challenges and Limitations in AI-Driven Healthcare Fraud Detection Systems [1, 4]

**7. Future Directions and Emerging Technologies**

Several intersecting technology trends are positioned to considerably improve the effectiveness of AI-driven healthcare fraud detection over the next decade. Within this set of trends, GNNs are the most immediately promising in their ability to detect fraud rings and collusion networks through the learned representations of provider-patient-payer relationships. The FY 2023 enforcement record documents multi-hub, multi-layered fraud networks—including telemedicine

companies coordinating with patient brokers, laboratory owners, and call centers—that are individually unremarkable but collectively anomalous; GNNs are uniquely suited to detect such structures, with published evaluations reporting precision improvements of 15–22 percentage points over feature-based ensemble baselines on labeled fraud ring datasets [7]. The National Rapid Response Strike Force's multi-jurisdictional mandate—established specifically because "the nature and scope of health care fraud has evolved rapidly with the advent of new technologies"—mirrors the computational logic of graph-based detection: fraud increasingly operates across networks of entities and geographic boundaries that no single-provider analysis can capture [1].

Transformer architectures increasingly apply revolutionary performance in natural language processing, achieved through attention-based sequence modeling, to longitudinal health data representing patient care trajectories and provider billing histories. The transformer's capacity to model long-range temporal dependencies makes it particularly suited to detecting gradual upcoding schemes, slowly escalating unnecessary procedure patterns, and patient steering behaviors that unfold over extended time horizons—schemes that, as FY 2023 case records document, sometimes operate for a decade or more before enforcement action [1]. Federated learning architectures enable multiple payer organizations to collaboratively train shared fraud detection models without exchanging raw claims data or PHI. The HFPP's cross-payer analytical model—in which 85 actively submitting partner organizations contribute claim-level data for joint analysis while maintaining data security requirements—represents a governance precursor to federated learning at scale; the program analyzed over 334 billion professional claim lines in FY 2023, a dataset size that would be impossible for any single organization to assemble independently [1]. Causal inference methods represent a conceptually distinct advancement, moving beyond correlational pattern recognition toward causal understanding of billing behavior: isolating whether a provider's elevated billing rates are caused by fraudulent intent or legitimately explained by patient complexity, practice geography, or specialty case mix [8].

Explainability research advances are expected to increase provider acceptance and regulatory approval of algorithmic fraud decisions. SHAP values provide theoretically grounded, model-agnostic decompositions of individual prediction contributions; LIME generates locally faithful linear approximations of complex model behavior in the vicinity of specific flagged claims; and counterfactual explanation systems answer the provider-relevant question of what specific changes to billing patterns would have resulted in a claim not being flagged. Regulatory evolution is expected to formalize requirements for standardized fairness metrics, mandatory explainability disclosures to flagged providers, and validation standards analogous to FDA device approval requirements. The FY 2023 report's documentation of CMS's requirement that Part C organizations submit detailed encounter data for each item and service—now totaling over 9.7 billion encounter records—provides the data infrastructure upon which next-generation transformer and GNN models will be trained, positioning the Federal fraud detection ecosystem for a substantial capability advance as these architectural approaches mature [1].

**8. Recommendations and Best Practices**

For healthcare payers undertaking fraud detection system deployments, a structured set of evidence-based best practices has emerged from documented production deployments. The HCFAC Program's three-year rolling average ROI of $2.80 per dollar expended—and HHS-OIG's program-specific ROI of approximately $10.00 per dollar obligated—establishes a compelling economic benchmark for investment justification; payer organizations should adopt equivalent ROI tracking frameworks before and after AI system deployment to measure impact against institutional fraud baselines [1]. Pilot tests are strongly recommended, even for moderate claim flows representative of the claims universe. Pilots should last for at least six months, both to cover seasonal cycles and to produce stable estimates of the technology's performance. Reliable baseline fraud metrics (confirmed claim count and dollars, false positive rate per investigation) are needed to assist in impact measurement and adjustment of the system [7]. Detection capacity planning cannot exist without investigation planning. I-MEDIC's contractor commenced 692 investigations and made 174 law enforcement referrals in FY 2023 alone (reported in [7]). This highlights the triage capacity built into the design of I-MEDIC to prevent the investigators' bandwidth from being overwhelmed by AI leads [1].

A feedback loop infrastructure, in which the findings of investigations are fed back into model retraining pipelines, is required to maintain detection efficacy against evolving schemes. The HCF Unit's practice of distributing "targeting packages"—data summaries and descriptions of why a billing pattern is suspect, including examples such as claims submitted for dead beneficiaries or beneficiaries living great distances from frequently attended clinics—provides a structured template for the investigation-to-model feedback workflow that payer organizations should replicate in their

internal fraud detection operations [1]. Transparency and provider communication programs substantially improve system acceptance and reduce litigation risk. The HFPP's model of publishing cross-payer study results and sharing fraud scheme alerts across 300 partner organizations demonstrates the value of collaborative transparency in reducing system-wide fraud exposure [1]. Establishing formal appeals procedures with mandatory human review authority, defined response timelines, and documentation requirements for provider justifications provides due process protections consistent with emerging regulatory expectations. For regulators, establishing standardized fairness metric requirements, clarifying liability frameworks for AI-generated fraud determinations, mandating transparency disclosures to flagged providers, and funding independent third-party validation of deployed systems represent the highest-priority policy actions supported by current deployment evidence [6].

**Conclusion**

Perhaps the most transformative opportunity for health care payers, regulators, and law enforcement agencies confronting a heterogeneous and adaptive fraud landscape is AI-inspired fraud detection. Gradient increasing ensembles, graph neural networks, transformer sequence models, and federated learning architectures have reached a level of maturity that they can now outperform rule-based and manual health care fraud detection methods in production. The enormous array of data from federal mass enforcement efforts and commercial payers' mass deployments of data-driven detection systems has demonstrated the capacity of these systems, when paired with sufficient investigative resources and powerful feedback loops, to achieve large, lasting recoveries and deterrence of future fraud, abuse, and waste. Simply deploying a technical solution is not enough. AI-enabled fraud detection has documented failings of algorithmic bias due to historical investigative inequities, opaque documentation of data limitations, and investigative resources exceeding lead processing. These can be addressed by the implementation of bias audit, explainability, and timeliness; easing transparent rights of appeal by providers; regulation; and evolving codes of practice and industry standards with continuous supervision.

**References**

[1] U.S. Dept. of Health and Human Services and U.S. Dept. of Justice, "Annual Report of the Departments of Health and Human Services and Justice," 2023. [Online]. Available: https://oig.hhs.gov/documents/hcfac/10087/HHS%20OIG%20FY%202023%20HCFAC.pdf

[2] Richard A. Bauder and Taghi M. Khoshgoftaar, "The Detection of Medicare Fraud Using Machine Learning Methods with Excluded Provider Labels," Artificial Intelligence Research Society Conference, 2018. [Online]. Available: https://cdn.aaai.org/ocs/17617/17617-77660-1-PB.pdf

[3] Pooja Kushwaha and Prof. Hitesh Gupta, "Survey Paper on Fraud detection in Healthcare using Deep Learning," International Journal of Research and Technology, Volume 12, Issue 4, 2024. [Online]. Available: https://ijrt.org/j/article/view/160/140

[4] Lushun Jiang et al., "Opportunities and challenges of artificial intelligence in the medical field: current application, emerging problems, and problem-solving strategies," Journal of International Medical Research, 2021. [Online]. Available:https://pmc.ncbi.nlm.nih.gov/articles/PMC8165857/

[5] Anli du Preez et al., "Fraud detection in healthcare claims using machine learning: A systematic review," Artificial Intelligence in Medicine, Volume 160, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0933365724003038

[6] Hogan Lovells, "False Claims Act Guide 2026," 2026. [Online]. Available: https://www.hoganlovells.com/en/publications/false-claims-act-guide-2026

[7] Margaret Mitchell et al., "Model Cards for Model Reporting," arXiv:1810.03993v2, 2019. [Online]. Available: https://arxiv.org/pdf/1810.03993

[8] NHCAA, "The Challenge of Health Care Fraud." [Online]. Available: https://www.nhcaa.org/tools-insights/about-health-care-fraud/the-challenge-of-health-care-fraud/

[9] Zahra Sadeghi et al., "A review of Explainable Artificial Intelligence in healthcare," Computers and Electrical Engineering, Volume 118, Part A, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0045790624002982