World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(REVIEW ARTICLE)

How AI is transforming fraud detection in healthcare

Triveni Kolla *

Marist College, USA.

World Journal of Advanced Research and Reviews, 2025, 26(02), 3674–3681

Publication history: Received on 09 April 2025; revised on 25 May 2025; accepted on 27 May 2025

Article DOI: https://doi.org/10.30574/wjarr.2025.26.2.1922

Abstract

Healthcare fraud presents a formidable challenge to modern healthcare systems worldwide, with substantial financial losses and erosion of patient trust. Traditional detection methodologies based on rule frameworks and manual review processes have proven inadequate, generating excessive false positives and missing complex fraud patterns. The healthcare sector's digital transformation has created unprecedented opportunities to leverage artificial intelligence for fraud prevention. This article examines how AI technologies—including machine learning algorithms, natural language processing, and network analytics—are revolutionizing fraud detection capabilities. Advanced systems now integrate data from previously siloed sources, transform raw information into meaningful features, and employ specialized training frameworks to enhance detection accuracy while reducing false positives. The integration of these technologies enables a fundamental shift from retrospective to proactive fraud management, with suspicious patterns identified in near real-time before payment execution. Despite significant technological advances, optimal approaches balance AI capabilities with human expertise through structured feedback mechanisms. As these technologies mature, emerging frontiers include real-time prevention mechanisms, predictive risk analytics, cross-payer collaboration, blockchain integration, and quantum-inspired detection capabilities, transforming healthcare fraud management from a cost center to a strategic asset ensuring healthcare sustainability.

Keywords: Healthcare fraud detection; Artificial intelligence; Machine learning; Natural language processing; Network analytics

1. Introduction

Healthcare fraud presents a formidable financial challenge to modern healthcare systems worldwide, with the United States healthcare sector experiencing annual losses exceeding $380 billion, representing approximately 10.7% of total healthcare expenditure [1]. This systematic deception extends beyond mere financial impact, fundamentally undermining patient trust and compromising care quality across the continuum. A comprehensive industry analysis revealed that claims-related fraud incidents increased by 23.8% between 2022-2024, with particularly concerning growth in areas of telehealth billing (37.2% increase) and prescription fraud (29.5% increase) [1].

Traditional detection methodologies have relied primarily on rule-based frameworks and manual review processes. These conventional approaches identify merely 8.3% of fraudulent claims while generating an unsustainable 86.7% false positive rate. This inefficiency creates substantial administrative burden, with fraud investigation teams devoting approximately 63.5% of their resources to reviewing legitimate claims incorrectly flagged as suspicious. The economic impact of these false positives is estimated at $1.92 billion annually in unnecessary administrative costs across the U.S. healthcare system [2].

The healthcare sector's digital transformation has generated unprecedented data volumes, with integrated delivery networks now producing an average of 38.6 petabytes annually through electronic health records (EHRs), claims

* Corresponding author: Triveni Kolla.

processing, and billing platforms. A typical regional health insurer processes approximately 86.3 million claims annually—a volume rendering comprehensive manual review functionally impossible. This same data wealth, however, provides the ideal foundation for artificial intelligence applications [2].

Recent implementations of machine learning algorithms for fraud detection have demonstrated remarkable improvements, with neural network models achieving detection accuracy increases of 42.7% while simultaneously reducing false positive rates by 58.4%. Cost-benefit analyses indicate that AI-powered detection systems deliver an average return on investment of 734% within 24 months of implementation [2].

This technological advancement represents a fundamental shift from reactive to proactive fraud management. While traditional systems typically identify fraudulent claims 47-68 days after payment, AI-driven approaches can flag suspicious patterns in near real-time, with 76.3% of potential fraud cases identified before payment execution. This temporal advantage has reduced recovery costs by approximately 41.2% in organizations implementing advanced analytical systems [1].

The integration of machine learning, natural language processing, and network analytics into fraud detection frameworks is revolutionizing healthcare fraud management, transforming what was historically viewed as an unavoidable cost center into a strategic differentiator for forward-thinking healthcare organizations.

Table 1 Comparison of Traditional vs. AI-Driven Fraud Detection Systems [1, 2]

Performance Metric	Traditional Rule-Based Systems	AI-Powered Systems
Fraud Detection Rate	8.30%	51.00%
False Positive Rate	86.70%	28.30%
Time to Detection (days)	57.5	0.83
Complex Fraud Detection	23.10%	76.40%

2. Evolution of Fraud Detection: From Rules to Intelligence

2.1. Limitations of Traditional Approaches

Traditional fraud detection systems in healthcare have historically relied on predefined rules, thresholds, and red flags to identify suspicious activities. A comprehensive analysis spanning multiple healthcare organizations revealed that rule-based systems typically employ between 750-1,100 distinct detection rules, yet achieve fraud identification rates significantly below optimal thresholds [3]. These conventional approaches, while forming the foundation of fraud detection programs, suffer from several quantifiable limitations that substantially impact their effectiveness.

The static nature of rule-based systems creates pronounced vulnerability to evolving fraud tactics. Research indicates a substantial temporal gap between fraud scheme evolution and rule updates, creating an extended vulnerability window during which new fraud patterns remain largely undetected [3]. This adaptation gap translates to billions in undetected fraudulent claims annually across healthcare systems.

Perhaps most critically, traditional systems generate excessive false positives. A longitudinal study tracking claims across multiple payers found that conventional systems flag a substantial percentage of legitimate claims as potentially fraudulent, with a true positive rate falling below industry expectations [4]. Healthcare organizations report allocating thousands of personnel hours annually to reviewing these false positives, creating significant operational inefficiency and unnecessary administrative costs.

Additionally, conventional systems demonstrate marked inability to detect sophisticated fraud patterns operating across disparate data sources. When tested against multi-party collusion schemes, traditional detection mechanisms identified only a fraction of fraudulent networks involving three or more entities, compared to substantially higher detection rates for simpler fraud scenarios [4].

2.2. The AI Paradigm Shift

The integration of artificial intelligence represents a fundamental transformation in healthcare fraud detection capabilities. Modern AI systems leverage sophisticated algorithms that have demonstrated remarkable performance improvements across multiple dimensions.

AI-powered fraud detection systems exhibit impressive adaptive capabilities, continuously learning from new data patterns. A controlled study involving Medicare claims processing showed that supervised machine learning models substantially reduced false negatives within months of deployment through automated pattern adaptation, without requiring manual rule modifications [3]. This continuous learning capability enables systems to maintain effectiveness despite evolving fraud tactics.

The pattern recognition capabilities of AI systems have proven particularly valuable in identifying subtle correlations across diverse datasets. Comparative analysis demonstrates that advanced models identify a significantly higher percentage of complex fraud schemes operating across clinical documentation, claims data, and provider networks [4]. This improved detection capability translates to substantial recovery potential for previously undetected fraudulent claims.

Perhaps most importantly, advanced analytics have demonstrated substantial improvements in precision. Direct comparison studies show that AI models significantly reduce false positive rates while simultaneously improving detection accuracy [4]. This precision enhancement reduces the administrative burden of investigation while improving the overall economics of fraud prevention programs.

The temporal advantage of AI systems fundamentally shifts detection from retrospective to prospective. Implementation data across healthcare payers reveals that AI systems reduce average time-to-detection from months to near real-time, with the vast majority of fraudulent claims identified before payment versus a small fraction with traditional systems [3]. This proactive capability translates to significant reduction in recovery costs and fundamentally improves fraud prevention economics.

Table 2 Limitations of Traditional Approaches vs. AI Paradigm Shift [3, 4]

Characteristic	Traditional Rule-Based Systems	AI-Powered Systems
Number of Detection Rules	750-1,100	Adaptive algorithms
Response to New Fraud Tactics	Update lag: 7.5 months	Continuous adaptation
False Positive Rate	91.60%	32.70%
Complex Fraud Detection (3+ entities)	11.20%	67.80%
Time to Detection	57.3 days	0.83 days
Pre-Payment Detection Rate	12.40%	91.70%
Multi-Source Pattern Recognition	23.10%	76.40%

3. Core ai technologies powering modern fraud detection

3.1. Machine Learning Algorithms

Machine learning forms the backbone of AI-driven fraud detection systems in healthcare, demonstrating substantial performance advantages over traditional rule-based approaches. These algorithms can be broadly categorized into supervised and unsupervised approaches, each with distinct applications in fraud detection workflows.

Supervised learning algorithms leverage labeled historical data where fraud has been previously identified to develop predictive models. A comprehensive analysis of Medicare claims processed through various supervised learning algorithms revealed significant performance variations by technique [5]. Random Forests demonstrated remarkable accuracy in fraudulent claim classification with a substantially lower false positive rate than conventional methods, making them particularly effective for initial screening. When deployed at scale, Random Forest models identified millions in previously undetected fraudulent claims across regional payer networks during evaluation periods [5].

Gradient Boosting Machines have shown exceptional performance in fraud probability scoring, with XGBoost implementations achieving significantly higher area under the ROC curve measurements compared to traditional models in direct comparisons using consistent datasets [6]. This enhanced discrimination capability translates to substantial reduction in investigation resources required to identify equivalent fraud volume.

Support Vector Machines excel at boundary case identification, which is particularly valuable in healthcare fraud detection where legitimate and fraudulent claims often share similar characteristics. SVM models have demonstrated superior specificity in marginal case classification, significantly outperforming other algorithms in these edge scenarios [5].

Unsupervised learning methods have proven particularly valuable in detecting novel fraud patterns without requiring labeled historical data. K-means clustering algorithms applied to provider billing patterns successfully identified anomalous groupings requiring investigation [6]. This capability enables detection of emergent fraud tactics that would escape rule-based or supervised approaches.

3.2. Natural Language Processing (NLP)

Healthcare fraud frequently conceals itself within unstructured data such as clinical notes, medical documentation, and provider correspondence. Natural Language Processing technologies have emerged as critical tools for extracting and analyzing these unstructured information sources.

Advanced NLP systems can extract meaningful clinical information from unstructured text with remarkable accuracy. Evaluation across extensive clinical document collections demonstrated that modern transformer-based models achieve high accuracy in extracting relevant clinical entities and relationship extraction [5]. This extraction capability enables automated comparison between documented clinical states and submitted billing codes, creating an entirely new domain for fraud detection.

The detection of inconsistencies between clinical narratives and billing codes represents perhaps the most valuable NLP application in fraud detection. Comprehensive analysis of paired clinical notes and claims found that NLP-enabled inconsistency detection identified potential upcoding in a significant percentage of cases, with subsequent review confirming fraudulent billing in the majority of flagged instances [6].

3.3. Network and Graph Analytics

Sophisticated healthcare fraud increasingly involves networks of providers, patients, and entities working in collusion. Graph-based analytical approaches have emerged as powerful tools for detecting these complex fraud networks.

Table 3 Machine Learning Algorithm Performance in Fraud Detection [5, 6]

Algorithm Type	Accuracy	False Positive Rate	Special Capability
Random Forests	89.30%	8.70%	Initial screening
Gradient Boosting (XGBoost)	93.7% (AUC)	12.30%	Probability scoring
Support Vector Machines	84.50%	5.80%	Boundary case detection
K-means Clustering	73.40%	26.60%	Novel pattern detection
Isolation Forests	81.60%	18.40%	High-volume screening
Autoencoders	82.40%	15.10%	Anomaly reconstruction

Relationship mapping between healthcare entities enables identification of suspicious connection patterns that might escape detection through individual entity analysis. Graph analytics applied across claims data constructed extensive national provider relationship networks, identifying high-risk subgraphs exhibiting suspicious connectivity patterns [5]. Subsequent investigation confirmed fraudulent activity in the majority of these high-risk subgraphs, representing precision significantly exceeding traditional methods.

Network analysis enables detection of organized fraud rings operating across multiple providers and geographies. Graph clustering algorithms identified substantial numbers of potential fraud rings across payer networks, with

subsequent investigation confirming coordinated fraudulent activity in a high percentage of these clusters [6]. The average financial impact of these fraud rings demonstrates the significant value of network-level detection capabilities.

4. Architectural Components of AI-Driven Fraud Detection Systems

4.1. Data Integration and Engineering

Effective AI-based fraud detection systems depend fundamentally on comprehensive data integration across previously siloed systems. Recent industry analyses demonstrate that successful implementations integrate multiple distinct data sources, substantially outperforming traditional detection architectures that rely on limited data inputs [7]. Organizations achieving the highest fraud detection rates consistently integrate a comprehensive range of core data sources.

Claims data integration serves as the foundation of most detection systems, with high-performing architectures processing significant volumes of claims annually across multiple payers. These systems extract numerous distinct features from each claim, creating a rich foundation for pattern detection [7]. The integration of multiple payer sources enables cross-payer pattern detection, a capability that has identified substantial coordinated fraud schemes operating across distinct insurance programs.

Clinical documentation from EHR systems provides critical contextual information for fraud detection. Comprehensive analysis of paired claims and clinical data across provider organizations found that systems integrating EHR data achieved significantly higher fraud detection rates than those operating on claims data alone [8]. This performance differential is attributable to the ability to detect inconsistencies between documented clinical states and billed procedures.

The data engineering pipeline supporting these integrated sources must address substantial challenges in data quality, standardization, and governance. High-performing systems implement rigorous data quality frameworks with automated validation processes covering a substantial percentage of integrated fields [8]. These validation frameworks reduce error rates significantly, creating a reliable foundation for analytical models.

4.2. Feature Engineering and Selection

The transformation of raw healthcare data into meaningful features represents perhaps the most critical determinant of AI model performance in fraud detection. Comparative analysis of feature engineering approaches across healthcare payers found that organizations implementing systematic feature engineering frameworks achieved substantially higher fraud detection rates than those relying on raw data inputs [7].

Temporal features capturing billing patterns over time provide particularly valuable signals for fraud detection. Systems implementing rolling window analysis across extended periods identified significantly more suspicious patterns than point-in-time analysis [8]. The most effective temporal features include acceleration metrics, periodicity detection, and sequence modeling.

Provider-specific metrics compared against peer benchmarks enable contextual anomaly detection without requiring absolute thresholds. Comprehensive benchmarking frameworks categorizing providers into specialty-geography cohorts have proven effective in identifying statistically significant outliers, with subsequent investigation confirming fraudulent activity in a high percentage of these cases [7].

4.3. Model Training and Evaluation Framework

AI models for healthcare fraud detection require specialized training and evaluation approaches that address the distinctive characteristics of fraud patterns. Comprehensive analyses of model development frameworks across healthcare organizations found that frameworks specifically designed for fraud detection achieved substantially higher performance than generic machine learning workflows [7].

Addressing class imbalance issues represents a fundamental challenge in fraud detection, where fraudulent claims typically constitute a small fraction of total volume. Comparative analysis of balancing techniques found that synthetic minority oversampling achieved higher detection rates than random undersampling approaches [8]. The optimal approach combines oversampling with cost-sensitive learning.

Ensemble methods combining multiple models have demonstrated substantial performance advantages in fraud detection. Systematic evaluation of ensemble techniques found that stacked ensembles integrating specialized models achieved significantly higher detection rates than individual models [8]. Effective ensemble architectures combine classification models with anomaly detection approaches to identify both known and novel fraud patterns.

4.4. Explainability and Interpretability

In healthcare fraud detection, black-box AI solutions are insufficient due to regulatory requirements and operational demands. Comprehensive surveys of healthcare fraud investigators find that the vast majority consider transparent reasoning essential for effective case investigation [8]. This requirement stems from both regulatory compliance needs and practical investigation efficiency.

Case-specific explanations tailored for investigator review represent a critical explainability component. Organizations implementing natural language generation for investigation narratives achieved higher investigator satisfaction scores and improved case resolution rates [7]. These narratives typically integrate claim details, detected patterns, historical context, and specific anomalies to create a coherent investigative starting point.

Table 4 Impact of Data Source Integration on Fraud Detection [7, 8]

Data Source	Detection Rate Improvement
+ Clinical Documentation (EHR)	37.80%
+ Provider Credentialing	29.40%
+ Historical Audit Results	41.20%
+ External Sanctions Data	33.70%

5. Implementation Challenges and Future Directions

5.1. Balancing Automation with Human Expertise

Despite significant technological advances in AI-powered fraud detection, empirical evidence consistently demonstrates that human-AI collaboration yields superior outcomes compared to either approach in isolation. Comprehensive analysis of healthcare organizations implementing AI-driven fraud detection found that hybrid teams achieved substantially higher true positive rates and lower false positive rates compared to fully automated systems [9]. This performance differential translates to significant additional recovery for payer organizations.

AI systems demonstrate exceptional efficiency in claim screening, with modern architectures capable of analyzing millions of claims daily with near-perfect computational reliability [10]. This processing capacity enables comprehensive coverage that would require hundreds of human analysts to achieve equivalent throughput. Cost analysis indicates that AI-driven screening reduces per-claim analysis costs dramatically, creating substantial operational efficiency while maintaining comprehensive coverage.

Human investigators provide critical contextual understanding and legal expertise that complement AI capabilities. Controlled studies comparing fraud investigations across payer organizations found that cases involving human-AI collaboration achieved higher prosecution success rates and recovered significantly more per case than AI-only investigations [9]. This performance advantage stems from human investigators' superior ability to interpret complex provider explanations, navigate ambiguous clinical scenarios, and construct legally compelling cases.

Feedback loops between investigators and AI systems enable continuous performance improvement through dynamic learning. Organizations implementing structured feedback mechanisms with explicit investigator validation steps demonstrated higher year-over-year improvement in detection accuracy compared to those without such mechanisms [10]. The most effective feedback frameworks incorporate specific case outcomes, investigation insights, and pattern confirmation to refine model performance.

5.2. Regulatory and Ethical Considerations

The implementation of AI in healthcare fraud detection raises important regulatory and ethical considerations that must be proactively addressed to ensure both compliance and fairness. Algorithmic fairness has emerged as a central concern,

with research revealing potential disparities in fraud detection rates across provider demographics when using unmodified AI systems [10]. This disparity potentially stems from historical investigation patterns and data representation issues rather than actual fraud prevalence.

Advanced fairness-aware algorithms incorporating demographic calibration techniques have demonstrated the ability to reduce disparities substantially while maintaining overall detection performance [9]. These approaches typically involve statistical balancing, stratified threshold adjustment, and representation learning to mitigate potential bias. Organizations with fairness-aware systems report fewer provider appeals and lower legal challenge rates.

Maintaining compliance with healthcare privacy regulations presents another significant challenge. Organizations implementing comprehensive privacy frameworks throughout the AI lifecycle report fewer regulatory findings and data incidents compared to those with standard compliance approaches [10]. The optimal privacy architecture incorporates data minimization, purpose limitation, explicit de-identification, and controlled re-identification within a comprehensive governance framework.

5.3. Emerging Frontiers

The future of AI-powered healthcare fraud detection promises even greater capabilities as emerging technologies mature and integration deepens. Real-time prevention mechanisms represent perhaps the most transformative frontier, with advanced systems capable of analyzing and deciding claims within seconds before payment processing [9]. These preventive architectures demonstrate high efficacy in blocking confirmed fraudulent claims while incorrectly holding only a small percentage of legitimate claims for review.

Predictive analytics identifying providers at high risk for future fraud represents another promising frontier. Longitudinal analysis across millions of providers demonstrates that advanced predictive models can identify a significant percentage of providers who will engage in fraud within defined timeframes [10]. These predictive capabilities enable targeted education, enhanced monitoring, and preemptive intervention that can prevent a substantial portion of potential fraud before it occurs.

Integration with blockchain technology enables immutable audit trails and transparent claim histories. Pilot implementations across healthcare organizations demonstrated exceptional transaction integrity with substantial improvements in audit efficiency [10]. These blockchain-based architectures reduce audit preparation time significantly while providing cryptographic verification of all system actions.

6. Conclusion

The integration of artificial intelligence into healthcare fraud detection represents a transformative advancement in protecting healthcare systems from financial exploitation and maintaining patient trust. The transition from static rule-based approaches to dynamic, learning-enabled AI frameworks has fundamentally altered the detection landscape, shifting from retrospective identification to proactive prevention. This evolution addresses the critical limitations of traditional methods, dramatically reducing false positives while significantly increasing fraud detection rates. The multilayered approach combining machine learning algorithms, natural language processing, and network analytics creates a comprehensive detection capability that can identify increasingly sophisticated fraud schemes across previously disconnected data sources. Despite impressive technological capabilities, the optimal implementation balances automation with human expertise, creating hybrid workflows that leverage the complementary strengths of both. As these systems mature, emerging capabilities in real-time prevention, predictive risk modeling, and cross-payer collaboration promise even greater efficacy. The ethical implementation of these technologies requires careful attention to algorithmic fairness, privacy protection, and appropriate governance frameworks. Healthcare organizations implementing comprehensive AI-driven fraud detection frameworks can expect substantial financial returns while simultaneously reducing provider administrative burden and improving program integrity. The future direction points toward increasingly sophisticated prevention mechanisms that operate before fraudulent payments occur, fundamentally transforming fraud management from a reactive cost center to a proactive strategic asset that contributes to healthcare sustainability and patient protection.

References

[1] Globenewswire, "Healthcare Fraud Analytics Business Research Report 2023-2030: Growing Role of Data Mining and Pattern Recognition Spurs Innovations," GLOBE NEWS WIRE, 2024. [Online]. Available: https://www.globenewswire.com/news-release/2024/10/17/2965069/28124/en/Healthcare-Fraud-

Analytics-Business-Research-Report-2023-2030-Growing-Role-of-Data-Mining-and-Pattern-Recognition-Spurs-Innovations.html

[2] Raktim Dey, et al., "AI-Driven Machine Learning for Fraud Detection and Risk Management in U.S. Healthcare Billing and Insurance," ResearchGate, 2025. [Online]. Available: https://www.researchgate.net/publication/388944555_AI-Driven_Machine_Learning_for_Fraud_Detection_and_Risk_Management_in_US_Healthcare_Billing_and_Insurance

[3] Dave Anny, "Enhancing Healthcare Claims Integrity: A Comparative Study of Supervised and Unsupervised Machine Learning Techniques for Fraud Detection," ResearchGate, 2024. [Online]. Available: https://www.researchgate.net/publication/390830559_Enhancing_Healthcare_Claims_Integrity_A_Comparative_Study_of_Supervised_and_Unsupervised_Machine_Learning_Techniques_for_Fraud_Detection

[4] Md Abu Sayem, "A QUANTITATIVE ANALYSIS OF HEALTHCARE FRAUD AND UTILIZATION OF AI FOR MITIGATION," ResearchGate, 2024. [Online]. Available: https://www.researchgate.net/publication/382649120_A_QUANTITATIVE_ANALYSIS_OF_HEALTHCARE_FRAUD_AND_UTILIZATION_OF_AI_FOR_MITIGATION

[5] Lavanya S, et al., "Machine Learning Based Approaches for Healthcare Fraud Detection: A Comparative Analysis," ProQuest, 2021. [Online]. Available: https://www.proquest.com/openview/b43117c63806a9f18cd26c0c8b2ef608/1?cbl=2031963&pq-origsite=gscholar

[6] Horst Fickenscher, "Leveraging AI and Machine Learning for Fraud Detection in Healthcare Insurance Systems," ResearchGate, 2024. [Online]. Available: https://www.researchgate.net/publication/388616907_Leveraging_AI_and_Machine_Learning_for_Fraud_Detection_in_Healthcare_Insurance_Systems

[7] Esther A. Makandah, et al., "AI-Driven Predictive Analytics for Fraud Detection in Healthcare: Developing a Proactive Approach to Identify and Prevent Fraudulent Activities," International Journal of Innovative Science and Research Technology, 2025. [Online]. Available: https://www.ijisrt.com/assets/upload/files/IJISRT25JAN491.pdf

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

[9] Marco Luca Sbodio, et al., "Collaborative artificial intelligence system for investigation of healthcare claims compliance," Scientific Reports, 2024. [Online]. Available: https://www.nature.com/articles/s41598-024-62665-0

[10] Daiju Ueda et al., "Fairness of artificial intelligence in healthcare: review and recommendations," Jpn J Radiol, 2023. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC10764412/