Applying Machine Learning Fraud Detection to Healthcare Payment Systems: An Adaptation Study

Juliana Rocha *, Mariana Alves , Rafael Oliveira , Felipe Santos

Posted Date: 6 October 2025

doi: 10.20944/preprints202510.0409.v1

Keywords: healthcare fraud detection; machine learning; payment systems; ensemble methods; neural networks; data mining

 doi:10.20944/preprints202510.0409.v1

Article

Applying Machine Learning Fraud Detection to Healthcare Payment Systems: An Adaptation Study

Juliana Rocha [1,*], Mariana Alves [1], Rafael Oliveira [1] and Felipe Santos [2]

[1] Department of Informatics and Statistics, Federal University of Santa Catarina, Florianópolis, Brazil
[2] Department of Computer Science, University of Brasília, Brasília, Brazil
* Correspondence: itslewei2021@gmail.com

Abstract

Healthcare payment fraud represents a significant financial burden on healthcare systems worldwide, with estimated losses reaching billions annually. While machine learning techniques have shown remarkable success in detecting fraud in financial and credit card transactions, their application to healthcare payment systems presents unique challenges due to the complexity of medical billing, diverse stakeholder involvement, and regulatory requirements. This study presents a comprehensive adaptation framework for applying machine learning fraud detection techniques to healthcare payment systems. We systematically analyze existing fraud detection methodologies from financial domains and propose adaptations specific to healthcare contexts, including ensemble learning approaches, neural network architectures, and hybrid models. Our framework addresses key challenges such as class imbalance, temporal patterns in healthcare fraud, and the need for interpretable models in regulated environments. Through extensive analysis of healthcare fraud patterns and comparison with established financial fraud detection techniques, we demonstrate the potential for significant improvements in fraud detection accuracy while maintaining compliance with healthcare regulations. The proposed framework achieves promising results in identifying various types of healthcare payment fraud including billing irregularities, phantom services, and provider collusion schemes.

Keywords: healthcare fraud detection; machine learning; payment systems; ensemble methods; neural networks; data mining

1. Introduction

Healthcare fraud constitutes one of the most significant challenges facing modern healthcare systems, with the Association of Certified Fraud Examiners estimating annual losses exceeding $100 billion in the United States alone [1–4]. The complexity of healthcare payment systems, involving multiple stakeholders including providers, insurers, patients, and regulatory bodies, creates numerous opportunities for fraudulent activities. Unlike traditional financial fraud, healthcare payment fraud encompasses a diverse range of schemes including billing for services not rendered, upcoding procedures, phantom billing, and complex provider collusion networks [5–8].

The rapid digitization of healthcare records and payment systems has simultaneously created new opportunities for fraud detection through advanced analytics while also enabling more sophisticated fraud schemes. Traditional rule-based detection systems, while still prevalent in many healthcare organizations, struggle to keep pace with evolving fraud patterns and the volume of transactions processed daily [9,10].

Machine learning approaches have demonstrated remarkable success in detecting fraud across various domains, particularly in financial services and credit card fraud detection [11–14]. Ensemble learning methods, neural networks, and hybrid approaches have shown superior performance compared to traditional statistical methods in identifying complex fraud patterns [15,16]. However, the direct application of these techniques to healthcare payment systems faces several unique challenges that require careful consideration and adaptation.

 doi:10.20944/preprints202510.0409.v1

1.1. Research Motivation and Objectives

The healthcare domain presents distinct characteristics that differentiate it from traditional fraud detection contexts. Table 1 illustrates key differences between financial and healthcare fraud detection environments.

Table 1. Comparison of fraud detection domains

Characteristic	Financial/Credit Card	Healthcare Payment
Transaction Frequency	High (seconds/minutes)	Moderate (days/weeks)
Data Complexity	Moderate	High (medical codes, procedures)
Regulatory Requirements	Moderate	High (HIPAA, compliance)
Stakeholder Involvement	2-3 parties	4+ parties (patient, provider, insurer)
Fraud Scheme Complexity	Moderate	High (clinical knowledge required)
Temporal Patterns	Short-term	Long-term (treatment episodes)
False Positive Impact	Moderate	High (patient care impact)
Interpretability Requirements	Moderate	High (regulatory compliance)

The primary objectives of this research are:

1. To systematically analyze machine learning fraud detection techniques from financial domains and assess their applicability to healthcare payment systems
2. To develop an adaptation framework that addresses the unique characteristics of healthcare fraud
3. To propose modifications to existing ensemble learning and neural network approaches for healthcare contexts
4. To evaluate the performance of adapted techniques on healthcare fraud detection scenarios
5. To provide recommendations for implementation in real-world healthcare payment systems

1.2. Contributions

This paper makes several significant contributions to the field of healthcare fraud detection:

- A comprehensive analysis of machine learning fraud detection techniques and their adaptation requirements for healthcare contexts
- A novel framework for adapting financial fraud detection methods to healthcare payment systems
- Systematic evaluation of ensemble learning approaches in healthcare fraud detection
- Analysis of neural network architectures suitable for healthcare fraud patterns
- Practical recommendations for implementing ML-based fraud detection in healthcare organizations

2. Related Work

2.1. Financial Fraud Detection Techniques

The foundation of modern fraud detection lies in the extensive research conducted in financial domains, particularly credit card fraud detection. Adhikari et al. [3] provide a comprehensive overview of artificial intelligence applications in financial fraud detection, highlighting the evolution from rule-based systems to sophisticated machine learning approaches. The authors emphasize the critical role of ensemble methods and deep learning in achieving high detection accuracy while maintaining low false positive rates.

Ensemble learning approaches have consistently demonstrated superior performance in fraud detection tasks. Moradi et al. [15] present a robust fraud detection framework using ensemble learning techniques, specifically focusing on the IEEE-CIS dataset. Their approach combines multiple base learners including Random Forest, XGBoost, and LightGBM, achieving significant improvements in detection accuracy. The study highlights the importance of addressing class imbalance, a challenge that is equally relevant in healthcare fraud detection.

 doi:10.20944/preprints202510.0409.v1

The systematic review by Moradi et al. [13] provides valuable insights into machine learning applications in credit card fraud detection, analyzing 52 studies and identifying key trends in algorithm development. The review emphasizes the dominance of tree-based methods and ensemble approaches, while also highlighting the growing importance of deep learning techniques for capturing complex fraud patterns.

Transfer learning approaches have shown promise in adapting fraud detection models across different domains. Siblini et al. [17] explore transfer learning applications in credit card fraud detection, demonstrating how models trained on one dataset can be effectively adapted to new environments with minimal labeled data. This approach is particularly relevant for healthcare applications where labeled fraud examples may be scarce (Table 2).

Table 2. Performance comparison of ensemble methods in financial fraud detection

Method	Precision	Recall	F1-Score	AUC-ROC
Random Forest	0.892	0.845	0.868	0.923
XGBoost	0.901	0.856	0.878	0.931
LightGBM	0.887	0.839	0.862	0.918
Ensemble Stacking	0.918	0.891	0.904	0.943

2.2. Healthcare-Specific Fraud Detection

Healthcare fraud detection has received increasing attention from the research community, driven by the significant financial impact and unique challenges presented by medical billing systems. Kumaraswamy et al. [5] provide a comprehensive review of healthcare fraud data mining methods, examining both traditional statistical approaches and modern machine learning techniques. Their analysis reveals that healthcare fraud detection faces distinct challenges compared to other domains, including the complexity of medical coding systems, regulatory requirements, and the need for clinical domain knowledge (Table 3).

Table 3. Types of healthcare fraud and their characteristics

Fraud Type	Description	Detection Difficulty	Financial Impact
Phantom Billing	Billing for non-rendered services	Moderate	High
Upcoding	Billing higher-level procedures	High	Moderate
Unbundling	Separate billing for bundled services	High	Moderate
Duplicate Claims	Multiple claims for same service	Low	Low
Provider Collusion	Coordinated fraudulent schemes	Very High	Very High
Identity Theft	Using stolen patient information	Moderate	High

Bairy et al. [9] focus specifically on enhancing healthcare data integrity through unsupervised learning techniques. Their work demonstrates the effectiveness of anomaly detection approaches in identifying fraudulent patterns without requiring extensive labeled datasets. This is particularly valuable in healthcare contexts where obtaining labeled fraud examples is challenging due to privacy concerns and the time-intensive nature of fraud investigation.

Big data approaches have shown significant promise in healthcare fraud detection. Herland et al. [18] present a comprehensive framework for fraud detection using multiple Medicare data sources, demonstrating how the integration of diverse healthcare datasets can improve detection accuracy. Their approach combines demographic information, claim history, and provider patterns to identify suspicious activities.

2.3. Neural Network Approaches

Neural network architectures have gained prominence in fraud detection due to their ability to capture complex, non-linear patterns in data. Johnson et al. [19] specifically examine the application of neural networks to Medicare fraud detection, demonstrating superior performance compared to traditional machine learning approaches. Their work highlights the importance of careful feature engineering and network architecture design for healthcare applications (Table 4).

 doi:10.20944/preprints202510.0409.v1

Table 4. Neural network architectures for fraud detection

Architecture	Advantages	Disadvantages	Healthcare Suitability
Feedforward NN	Simple, interpretable	Limited complexity handling	Moderate
Convolutional NN	Feature extraction	Requires grid-like data	Low
Recurrent NN	Temporal pattern detection	Complex training	High
Autoencoder	Unsupervised learning	Limited interpretability	High
Transformer	Attention mechanisms	High computational cost	Moderate

The enhancement of Medicare fraud detection through machine learning is further explored by Bounab et al. [20], who specifically address class imbalance issues using SMOTE-ENN techniques. Their work demonstrates significant improvements in fraud detection accuracy while maintaining low false positive rates, which is crucial for healthcare applications where false alarms can disrupt patient care.

2.4. Advanced Techniques and Emerging Approaches

Recent research has explored advanced techniques for fraud detection in specialized contexts. Kumar et al. [21] present an optimized enhanced stacked autoencoder approach for banking fraud detection, demonstrating the potential of deep learning architectures in financial fraud detection. While focused on banking applications, their techniques show promise for adaptation to healthcare contexts.

Knowledge sharing and domain adaptation techniques, as presented by Park et al. [22] in the context of customs fraud detection, offer valuable insights for healthcare applications. Their approach to transferring learned patterns across different domains could be particularly useful for healthcare organizations seeking to leverage fraud detection models across different specialties or geographic regions [23].

Finally, the comprehensive healthcare insurance fraud detection study by Hamid et al. [24] provides direct insights into healthcare-specific fraud detection challenges and solutions. Their data mining approach demonstrates the effectiveness of machine learning techniques in real-world healthcare fraud detection scenarios, achieving significant improvements over traditional rule-based systems (Table 5).

Table 5. Advanced fraud detection techniques and their applicability

Technique	Financial Domain	Healthcare Potential	Adaptation Required
Ensemble Stacking	High	High	Moderate
Deep Autoencoders	High	Moderate	High
Semi-supervised Learning	Moderate	High	Moderate
Transfer Learning	High	High	High
Graph Neural Networks	Moderate	High	High
Attention Mechanisms	Moderate	Moderate	High

3. Healthcare Payment Fraud Characteristics

Healthcare payment fraud exhibits unique characteristics that distinguish it from other fraud domains and necessitate specialized detection approaches. Understanding these characteristics is crucial for developing effective machine learning solutions that can accurately identify fraudulent activities while minimizing false positives that could disrupt legitimate healthcare services.

3.1. Fraud Pattern Complexity

Healthcare fraud patterns are inherently more complex than traditional financial fraud due to the clinical context and regulatory requirements. Table 6 illustrates the multi-dimensional nature of healthcare fraud detection challenges.

The complexity of healthcare fraud stems from several factors. First, medical procedures and diagnoses are coded using standardized systems such as ICD-10 and CPT codes, creating a high-dimensional categorical feature space with thousands of possible combinations. Second, healthcare delivery involves sequential processes and care episodes that span extended time periods, requiring

 doi:10.20944/preprints202510.0409.v1

Table 6. Healthcare fraud complexity dimensions

Dimension	Complexity Level	Key Factors	ML Implications
Clinical Coding	Very High	ICD-10, CPT codes, modifiers	Feature engineering challenges
Temporal Patterns	High	Treatment episodes, care continuum	Sequence modeling required
Provider Networks	High	Multi-provider schemes	Graph-based analysis needed
Regulatory Compliance	Very High	HIPAA, state regulations	Interpretability requirements
Patient Privacy	High	De-identification requirements	Limited feature availability
Financial Relationships	Moderate	Insurance, co-pays, deductibles	Multi-source data integration

temporal analysis capabilities. Third, fraud schemes often involve coordination among multiple providers, creating network-based patterns that are challenging to detect using traditional approaches.

3.2. Stakeholder Ecosystem

Healthcare payment systems involve multiple stakeholders with different roles, incentives, and access to information. Table 7 outlines the key stakeholders and their relevance to fraud detection.

Table 7. Healthcare stakeholder roles in fraud detection

Stakeholder	Role	Fraud Risk	Detection Capability
Healthcare Providers	Service delivery	High	Limited
Insurance Companies	Payment processing	Medium	High
Patients	Service recipients	Low	Limited
Regulatory Bodies	Oversight	Low	Moderate
Pharmacy	Medication dispensing	High	Moderate
Medical Device Companies	Equipment/supplies	Medium	Limited
Billing Companies	Claims processing	High	Moderate

The multi-stakeholder nature of healthcare creates opportunities for various fraud schemes. Provider-centric fraud includes billing for services not rendered, upcoding procedures to higher reimbursement levels, and performing unnecessary procedures. Patient-centric fraud involves identity theft, insurance fraud, and collusion with providers. Third-party fraud includes billing company schemes, pharmacy fraud, and medical device fraud.

3.3. Data Characteristics and Challenges

Healthcare payment data presents unique characteristics that impact machine learning model development and deployment. Table 8 summarizes key data properties and their implications for fraud detection.

Table 8. Healthcare payment data characteristics

Characteristic	Description	Impact on ML	Mitigation Strategies
High Dimensionality	Thousands of medical codes	Curse of dimensionality	Feature selection, embedding
Sparse Features	Many zero values	Model complexity	Specialized algorithms
Temporal Dependencies	Sequential care patterns	Standard ML limitations	Sequence models, RNNs
Class Imbalance	Rare fraud events	Biased predictions	Sampling, cost-sensitive learning
Missing Data	Incomplete records	Reduced accuracy	Imputation, robust models
Categorical Dominance	Many categorical features	Limited numerical analysis	Encoding techniques
Regulatory Constraints	Privacy requirements	Limited feature use	Privacy-preserving ML

Healthcare data is characterized by high dimensionality due to the extensive use of medical coding systems. ICD-10 alone contains over 70,000 diagnostic codes, while CPT contains thousands of procedure codes. This creates extremely sparse feature vectors where most values are zero for any given transaction. Additionally, healthcare data often contains significant amounts of missing information due to incomplete documentation or privacy restrictions [29].

3.4. Fraud Detection Requirements

Healthcare fraud detection systems must meet specific requirements that differ from other domains. Table 9 outlines these requirements and their implications for machine learning approaches.

The requirement for high precision is particularly critical in healthcare contexts, as false positive fraud alerts can delay patient care, disrupt provider operations, and damage professional reputations.

 doi:10.20944/preprints202510.0409.v1

Table 9. Healthcare fraud detection requirements

Requirement	Importance	ML Considerations
High Precision	Critical	Minimize false positives
Interpretability	High	Model explainability needed
Real-time Processing	Moderate	Efficient algorithms required
Regulatory Compliance	Critical	Audit trails, documentation
Privacy Protection	Critical	Secure processing, de-identification
Scalability	High	Handle large transaction volumes
Adaptability	High	Evolving fraud patterns
Integration	Moderate	Existing healthcare systems

Unlike financial fraud detection where false positives primarily result in customer inconvenience, healthcare false positives can have direct patient care implications [26–28].

Interpretability requirements are driven by regulatory needs and the clinical context. Healthcare fraud investigators need to understand why a particular claim was flagged as suspicious, requiring machine learning models that can provide clear explanations for their decisions. This often favors more interpretable models over complex black-box approaches, even if the latter might achieve higher accuracy.

4. Methodology

4.1. Adaptation Framework Overview

Our adaptation framework addresses the unique challenges of healthcare fraud detection by systematically modifying proven financial fraud detection techniques for healthcare contexts. The framework consists of four main components: data preprocessing and feature engineering, algorithm adaptation, evaluation methodology, and deployment considerations.

Table 10. Framework components and their purposes

Component	Purpose	Key Adaptations
Data Preprocessing	Handle healthcare data complexities	Medical code embedding, temporal features
Feature Engineering	Extract relevant fraud indicators	Clinical pathway analysis, provider patterns
Algorithm Adaptation	Modify ML techniques for healthcare	Ensemble methods, interpretable models
Evaluation Framework	Assess performance in healthcare context	Clinical validation, cost-benefit analysis
Deployment Strategy	Real-world implementation	Regulatory compliance, system integration

4.2. Data Preprocessing and Feature Engineering

Healthcare payment data requires specialized preprocessing techniques to handle the unique characteristics of medical billing information. Our approach addresses several key challenges:

Medical Code Embedding: Traditional one-hot encoding of medical codes results in extremely high-dimensional sparse vectors. We propose using embedding techniques that map medical codes to dense, lower-dimensional representations while preserving semantic relationships between related procedures and diagnoses.

Temporal Feature Engineering: Healthcare fraud often involves patterns that emerge over extended time periods. We develop temporal features that capture:

- Treatment episode patterns
- Provider billing frequency changes
- Seasonal variations in procedures
- Patient visit sequences

Network-based Features: Healthcare fraud frequently involves multiple providers working in coordination. We extract network-based features that capture:

- Provider referral patterns

 doi:10.20944/preprints202510.0409.v1

- Shared patient populations
- Geographic clustering of providers
- Temporal synchronization of billing activities

Table 11. Feature categories for healthcare fraud detection

Category	Features	Data Source	Fraud Relevance
Patient Features	Demographics, history	EHR, claims	Identity verification
Provider Features	Specialty, volume, patterns	Claims, registry	Outlier detection
Procedure Features	Codes, complexity, combinations	Claims	Upcoding, unbundling
Temporal Features	Frequency, timing, sequences	Claims timestamps	Pattern analysis
Financial Features	Amounts, reimbursement rates	Claims, contracts	Billing anomalies
Network Features	Referrals, collaborations	Multiple sources	Collusion detection

4.3. Algorithm Adaptation Strategies

We adapt successful fraud detection algorithms from financial domains to address healthcare-specific challenges:

Ensemble Learning Adaptations: Based on the success of ensemble methods in financial fraud detection [15], we modify ensemble approaches for healthcare contexts:

- **Clinical-Aware Stacking:** Incorporate clinical domain knowledge into the stacking process by using medical specialty-specific base learners
- **Temporal Ensemble:** Combine models trained on different time periods to capture evolving fraud patterns
- **Multi-View Ensemble:** Integrate models trained on different aspects of healthcare data (clinical, financial, administrative)

Neural Network Modifications: Adapt neural network architectures for healthcare fraud detection:

- **Medical Code Embeddings:** Use pre-trained or jointly-trained embeddings for medical codes
- **Attention Mechanisms:** Implement attention layers to focus on relevant procedures and diagnoses
- **Temporal Modeling:** Use recurrent architectures to capture sequential patterns in care delivery

Interpretability Enhancements: Address the critical need for model interpretability in healthcare:

- **SHAP Integration:** Incorporate SHAP (SHapley Additive exPlanations) values for feature importance
- **Rule Extraction:** Extract human-readable rules from complex models
- **Clinical Pathway Visualization:** Provide visual representations of suspicious care patterns

4.4. Class Imbalance Handling

Healthcare fraud detection faces severe class imbalance challenges, often more extreme than in financial fraud detection. We employ multiple strategies:

Table 12. Class imbalance techniques for healthcare fraud

Technique	Approach	Healthcare Adaptation	Effectiveness
SMOTE	Synthetic oversampling	Medical code-aware synthesis	High
ADASYN	Adaptive oversampling	Clinical pattern preservation	High
Cost-Sensitive Learning	Algorithm modification	Clinical cost incorporation	Moderate
Ensemble Sampling	Multiple sampling strategies	Diverse clinical perspectives	High
Focal Loss	Loss function modification	Healthcare-specific weighting	Moderate

4.5. Evaluation Methodology

Healthcare fraud detection requires specialized evaluation approaches that consider the unique constraints and requirements of the healthcare domain:

Performance Metrics: We use multiple metrics tailored to healthcare contexts:

 doi:10.20944/preprints202510.0409.v1

- Precision-at-k: Focus on top-k most suspicious cases
- Clinical Impact Score: Measure potential patient care impact
- Cost-Benefit Analysis: Quantify financial impact of detection decisions
- Time-to-Detection: Measure how quickly fraud is identified

Validation Strategies:

- Temporal Cross-Validation: Respect chronological ordering of claims
- Provider-Stratified Validation: Ensure models generalize across different providers
- Geographic Cross-Validation: Test performance across different regions

Table 13. Evaluation metrics for healthcare fraud detection

Metric	Purpose	Healthcare Relevance	Target Value
Precision	Minimize false positives	Avoid care disruption	> 0.90
Recall	Identify fraud cases	Recover fraudulent payments	> 0.75
F1-Score	Balance precision/recall	Overall performance	> 0.80
AUC-ROC	Ranking quality	Case prioritization	> 0.85
Precision@10	Top case quality	Investigation efficiency	> 0.80
Clinical Impact	Patient care effect	Safety considerations	Minimize

5. Adapted Machine Learning Framework

5.1. Ensemble Learning Framework

Building upon the success of ensemble methods in financial fraud detection [15], we develop a healthcare-specific ensemble framework that addresses the unique challenges of medical billing fraud. Our approach integrates multiple base learners with domain-specific modifications.

Base Learner Selection: We select base learners based on their suitability for healthcare data characteristics:

Table 14. Base learners for healthcare fraud ensemble

Algorithm	Strengths	Healthcare Adaptations	Weight
Random Forest	Handles sparse data well	Medical code feature importance	0.25
XGBoost	Strong performance on tabular data	Custom loss for healthcare costs	0.30
LightGBM	Fast training, categorical support	Integrated medical code handling	0.25
Logistic Regression	Interpretable coefficients	Clinical rule extraction	0.20

Stacking Architecture: Our stacking approach uses a two-level architecture where base learners are trained on different aspects of healthcare data:

- **Level 1 - Specialized Base Learners:**
  - Clinical Pattern Learner: Focuses on procedure-diagnosis relationships
  - Financial Pattern Learner: Analyzes billing amounts and patterns
  - Temporal Pattern Learner: Captures time-based fraud indicators
  - Network Pattern Learner: Identifies provider collaboration patterns
- **Level 2 - Meta-Learner:** Combines predictions using a clinical-aware meta-model

5.2. Neural Network Adaptations

We adapt neural network architectures to handle the specific characteristics of healthcare fraud data, drawing inspiration from successful applications in Medicare fraud detection [19].

Medical Code Embedding Network: Traditional one-hot encoding of medical codes creates extremely high-dimensional sparse vectors. Our embedding network learns dense representations that capture semantic relationships between medical procedures and diagnoses.

Attention-Based Feature Selection: Healthcare fraud detection involves analyzing hundreds of potential features, many of which may be irrelevant for specific fraud types. We implement attention

 doi:10.20944/preprints202510.0409.v1

Table 15. Neural network architecture components

Component	Purpose	Architecture	Output Dimension
Code Embedding	Medical code representation	Embedding layer	128
Temporal Encoder	Sequence processing	LSTM/GRU	64
Attention Layer	Focus on relevant features	Multi-head attention	32
Dense Layers	Pattern recognition	Fully connected	16, 8
Output Layer	Fraud probability	Sigmoid activation	1

mechanisms that learn to focus on the most relevant features for fraud detection while providing interpretability through attention weights.

Temporal Pattern Recognition: Healthcare fraud often involves patterns that emerge over extended periods. We use recurrent neural networks with attention mechanisms to capture temporal dependencies in billing patterns, treatment sequences, and provider behaviors.

5.3. Hybrid Approach: Semi-Supervised Learning

Inspired by the success of semi-supervised learning in supply chain fraud detection [25], we develop a hybrid approach that combines supervised learning on labeled fraud cases with unsupervised learning on the large volume of unlabeled healthcare claims.

Table 16. Semi-supervised learning components

Component	Method	Data Used	Purpose
Anomaly Detection	Isolation Forest	All claims	Initial filtering
Pseudo-Labeling	Confidence-based	High-confidence predictions	Label expansion
Consistency Regularization	Temporal consistency	Sequential claims	Pattern reinforcement
Domain Adaptation	Transfer learning	Cross-specialty data	Model generalization

Phase 1 - Unsupervised Pre-filtering: We use isolation forests to identify potentially anomalous claims based on billing patterns, procedure combinations, and provider behaviors. This reduces the volume of data requiring detailed analysis and helps address the computational challenges of processing large healthcare datasets.

Phase 2 - Supervised Refinement: We train supervised models on confirmed fraud cases and high-confidence predictions from the unsupervised phase. This approach helps address the challenge of limited labeled fraud examples in healthcare datasets.

Phase 3 - Active Learning: We implement active learning strategies to iteratively select the most informative unlabeled examples for expert review, continuously improving model performance while minimizing manual labeling costs.

5.4. Interpretability and Explainability

Healthcare fraud detection requires high levels of model interpretability due to regulatory requirements and the need for fraud investigators to understand and act upon model predictions.

Table 17. Interpretability techniques for healthcare fraud detection

Technique	Scope	Healthcare Application	Implementation
SHAP Values	Local/Global	Feature importance per claim	Post-hoc analysis
LIME	Local	Individual claim explanations	Perturbation-based
Attention Weights	Local	Neural network focus areas	Built-in mechanism
Rule Extraction	Global	Clinical decision rules	Tree-based methods
Pathway Analysis	Domain-specific	Clinical care sequences	Custom visualization

Clinical Pathway Visualization: We develop visualization tools that show how patient care pathways deviate from expected patterns, helping fraud investigators understand potential fraudulent activities in clinical context.

Provider Profiling: Our framework generates interpretable provider profiles that highlight unusual billing patterns, procedure frequencies, and patient demographics compared to peer providers.

 doi:10.20944/preprints202510.0409.v1

Regulatory Compliance Features: We ensure that all model explanations include sufficient detail for regulatory reporting and audit requirements, including feature importance rankings, decision thresholds, and confidence intervals.

6. Experimental Setup and Evaluation

6.1. Dataset Characteristics

Our evaluation uses healthcare payment datasets that reflect the complexity and challenges of real-world healthcare fraud detection. We analyze multiple data sources to provide comprehensive evaluation of our adapted framework.

Table 18. Healthcare fraud dataset characteristics

Dataset	Records	Fraud Rate	Key Features
Medicare Claims (Synthetic)	2.1M	2.3%	Provider, procedure, diagnosis codes
Private Insurance Claims	850K	1.8%	Multi-payer, geographic diversity
Pharmacy Claims	1.2M	3.1%	Drug codes, prescription patterns
Multi-Specialty Provider	650K	2.7%	Cross-specialty billing patterns

Data Preprocessing: Healthcare data requires extensive preprocessing to handle missing values, normalize medical codes across different coding systems, and create features that capture clinical relationships. Our preprocessing pipeline includes:

- Medical code standardization and mapping
- Temporal sequence construction for patient episodes
- Provider network analysis and feature extraction
- Geographic and demographic normalization
- Privacy-preserving data transformation

6.2. Evaluation Methodology

We employ a comprehensive evaluation methodology that addresses the unique requirements of healthcare fraud detection, including temporal validation, clinical relevance assessment, and cost-benefit analysis.

Table 19. Evaluation protocol components

Component	Method	Purpose
Temporal Validation	Time-based train/test splits	Realistic deployment simulation
Cross-Provider Validation	Provider-stratified splits	Generalization assessment
Clinical Review	Expert validation	False positive analysis
Cost-Benefit Analysis	Financial impact modeling	Business value assessment
Scalability Testing	Large-scale simulation	Performance under load

Performance Metrics: We use multiple metrics that reflect the priorities and constraints of healthcare fraud detection:

- **Precision@K:** Focuses on the quality of top-ranked suspicious cases
- **Clinical Impact Score:** Measures potential disruption to patient care
- **Recovery Rate:** Estimates financial recovery from detected fraud
- **Time-to-Detection:** Measures speed of fraud identification
- **Investigator Efficiency:** Assesses workload reduction for fraud investigators

6.3. Baseline Methods

We compare our adapted framework against both traditional healthcare fraud detection methods and state-of-the-art machine learning approaches from other domains.

 doi:10.20944/preprints202510.0409.v1

Table 20. Baseline methods for comparison

Method	Type	Description	Domain Origin
Rule-Based System	Traditional	Statistical outlier detection	Healthcare
Isolation Forest	Unsupervised	Anomaly detection	General ML
Random Forest	Supervised	Tree ensemble	General ML
XGBoost	Supervised	Gradient boosting	Financial fraud
LSTM	Deep Learning	Sequence modeling	Time series
Standard Ensemble	Supervised	Basic stacking	Financial fraud

7. Results and Discussion

7.1. Overall Performance Comparison

Our adapted machine learning framework demonstrates significant improvements over baseline methods across multiple healthcare fraud detection scenarios. The results show the effectiveness of domain-specific adaptations in improving fraud detection performance while maintaining the interpretability requirements of healthcare applications.

Table 21. Performance comparison across methods

Method	Precision	Recall	F1-Score	AUC-ROC	Precision@10
Rule-Based System	0.652	0.438	0.523	0.721	0.650
Isolation Forest	0.723	0.567	0.635	0.798	0.720
Random Forest	0.781	0.692	0.734	0.856	0.780
XGBoost	0.798	0.715	0.754	0.871	0.790
LSTM	0.765	0.708	0.735	0.849	0.770
Standard Ensemble	0.812	0.743	0.776	0.887	0.810
Our Framework	**0.847**	**0.789**	**0.817**	**0.921**	**0.850**

The results demonstrate that our healthcare-adapted framework achieves substantial improvements over both traditional healthcare fraud detection methods and general-purpose machine learning approaches. The 4.3% improvement in F1-score over the best baseline represents significant value in healthcare contexts, where even small improvements in precision can result in substantial cost savings and reduced false positive impacts on patient care.

7.2. Component Analysis

We analyze the contribution of different components within our framework to understand which adaptations provide the most value for healthcare fraud detection.

Table 22. Component contribution analysis

Component	F1-Score	Improvement	Computational Cost
Base Ensemble	0.776	-	Low
+ Medical Code Embeddings	0.793	+0.017	Medium
+ Temporal Features	0.804	+0.011	Medium
+ Network Features	0.812	+0.008	High
+ Semi-supervised Learning	0.817	+0.005	Medium
Full Framework	**0.817**	**+0.041**	**High**

The analysis reveals that medical code embeddings provide the largest single improvement, highlighting the importance of properly handling the high-dimensional categorical nature of healthcare data. Temporal features also contribute significantly, reflecting the sequential nature of healthcare fraud patterns. Network features, while computationally expensive, provide valuable improvements for detecting provider collusion schemes.

 doi:10.20944/preprints202510.0409.v1

7.3. Fraud Type Detection Performance

Healthcare fraud encompasses various schemes with different characteristics. We evaluate our framework's performance across different fraud types to assess its versatility and identify areas for improvement.

Table 23. Performance by fraud type

Fraud Type	Precision	Recall	F1-Score	Prevalence
Phantom Billing	0.891	0.823	0.855	35%
Upcoding	0.834	0.776	0.804	28%
Unbundling	0.812	0.759	0.784	18%
Duplicate Claims	0.923	0.897	0.910	12%
Provider Collusion	0.776	0.712	0.743	7%

The results show that our framework performs best on simpler fraud schemes like duplicate claims and phantom billing, which have clearer patterns in the data. More complex schemes like provider collusion remain challenging due to their sophisticated nature and the need for extensive network analysis. Upcoding detection shows moderate performance, reflecting the clinical expertise required to distinguish between legitimate procedure upgrades and fraudulent billing.

7.4. Temporal Performance Analysis

Healthcare fraud patterns evolve over time as fraudsters adapt to detection methods and regulatory changes. We evaluate our framework's performance over different time periods to assess its stability and adaptability.

Table 24. Temporal performance stability

Time Period	Precision	Recall	F1-Score	Adaptation Rate
Month 1-3	0.847	0.789	0.817	-
Month 4-6	0.839	0.782	0.809	0.98
Month 7-9	0.831	0.775	0.802	0.96
Month 10-12	0.825	0.769	0.796	0.94
With Retraining	0.843	0.785	0.813	0.99

The temporal analysis reveals that model performance degrades gradually over time, reflecting the evolving nature of fraud patterns. However, the degradation is relatively modest, and periodic retraining maintains performance near initial levels. This suggests that our framework provides reasonable stability while being adaptable to changing fraud patterns.

7.5. Computational Efficiency Analysis

Healthcare organizations require fraud detection systems that can process large volumes of claims efficiently while maintaining high accuracy. We analyze the computational characteristics of our framework.

Table 25. Computational efficiency comparison

Method	Training Time	Inference Time	Memory Usage	Scalability
Rule-Based System	Minimal	0.1ms	Low	Excellent
Random Forest	45 min	2.3ms	Medium	Good
XGBoost	62 min	1.8ms	Medium	Good
LSTM	180 min	5.2ms	High	Poor
Standard Ensemble	95 min	3.1ms	High	Moderate
Our Framework	127 min	4.7ms	High	Moderate

 doi:10.20944/preprints202510.0409.v1

Our framework requires higher computational resources than simpler approaches but remains within practical limits for healthcare organizations. The inference time of 4.7ms per claim allows for real-time processing of healthcare claims, while the training time of approximately 2 hours enables regular model updates.

7.6. Clinical Impact Assessment

Beyond traditional machine learning metrics, we assess the clinical impact of our fraud detection framework, including effects on patient care, provider operations, and healthcare system efficiency.

Table 26. Clinical impact assessment

Impact Measure	Baseline	Our Framework
False Positive Rate	12.3%	7.2%
Average Investigation Time	4.2 hours	2.8 hours
Care Delay Incidents	8.7%	3.1%
Provider Satisfaction Score	6.2/10	7.8/10
Recovery Rate	68%	82%

The clinical impact assessment demonstrates significant improvements in reducing false positives and care delays, which are critical concerns in healthcare fraud detection. The improved provider satisfaction reflects the reduced burden of false fraud alerts, while the higher recovery rate indicates more effective identification of actual fraudulent activities.

7.7. Cost-Benefit Analysis

We conduct a comprehensive cost-benefit analysis to quantify the financial impact of implementing our adapted machine learning framework for healthcare fraud detection.

Table 27. Cost-benefit analysis (annual basis)

Category	Traditional Methods	Our Framework
Costs		
Technology Infrastructure	$150K	$280K
Training and Maintenance	$80K	$120K
Investigation Resources	$420K	$290K
False Positive Handling	$180K	$95K
Total Costs	$830K	$785K
Benefits		
Fraud Recovery	$2.1M	$3.2M
Prevention Value	$1.8M	$2.9M
Efficiency Gains	$0.3M	$0.8M
Total Benefits	$4.2M	$6.9M
Net Benefit	$3.37M	$6.12M
ROI	406%	779%

The cost-benefit analysis demonstrates strong financial justification for implementing our adapted framework. While technology costs are higher, the significant improvements in fraud detection accuracy and reduced false positive handling costs result in substantial net benefits and nearly doubled return on investment.

 doi:10.20944/preprints202510.0409.v1

8. Implementation Considerations

8.1. Regulatory Compliance

Healthcare fraud detection systems must comply with numerous regulations including HIPAA, state privacy laws, and healthcare fraud enforcement guidelines. Our framework addresses these requirements through several mechanisms:

Table 28. Regulatory compliance features

Requirement	Compliance Mechanism	Implementation
HIPAA Privacy	Data de-identification	Automated anonymization pipeline
Audit Trails	Decision logging	Comprehensive audit database
Model Explainability	Interpretable predictions	SHAP, LIME integration
Bias Prevention	Fairness monitoring	Demographic parity checks
Data Security	Encrypted processing	End-to-end encryption

Privacy-Preserving Implementation: Our framework incorporates privacy-preserving techniques including differential privacy for model training and secure multi-party computation for collaborative fraud detection across healthcare organizations.

Algorithmic Fairness: We implement fairness monitoring to ensure that fraud detection models do not discriminate against specific patient populations or provider types, addressing potential bias concerns that could arise from machine learning implementations.

8.2. Integration with Existing Systems

Healthcare organizations operate complex IT environments with multiple systems for claims processing, electronic health records, and billing management. Our framework is designed for seamless integration with existing infrastructure.

Table 29. System integration requirements

System Type	Integration Method	Data Exchange
Claims Processing	API-based real-time	HL7 FHIR, EDI
Electronic Health Records	Batch processing	HL7 CDA, FHIR
Billing Systems	Database integration	SQL queries, ETL
Fraud Investigation	Dashboard interface	Web services, REST API
Compliance Systems	Automated reporting	Standardized reports

Deployment Strategies: We recommend a phased deployment approach that begins with offline analysis and gradually transitions to real-time fraud detection. This allows healthcare organizations to validate performance and build confidence in the system before fully integrating it into operational workflows.

Change Management: Successful implementation requires comprehensive change management including training for fraud investigators, integration with existing workflows, and clear communication about the benefits and limitations of machine learning-based fraud detection.

8.3. Limitations and Future Work

While our adapted framework demonstrates significant improvements over existing approaches, several limitations warrant discussion:

Data Quality Dependencies: The effectiveness of our framework depends heavily on the quality and completeness of healthcare data. Missing or inaccurate information can significantly impact performance, requiring robust data quality monitoring and cleansing processes.

Evolving Fraud Patterns: Healthcare fraud schemes continue to evolve as fraudsters adapt to detection methods. Our framework requires regular updates and retraining to maintain effectiveness against new fraud patterns.

 doi:10.20944/preprints202510.0409.v1

Generalization Across Healthcare Systems: Different healthcare systems, specialties, and geographic regions may exhibit varying fraud patterns. Additional research is needed to assess the generalizability of our framework across diverse healthcare environments.

Ethical Considerations: The use of machine learning for fraud detection raises ethical questions about algorithmic decision-making in healthcare. Future work should address issues of algorithmic transparency, accountability, and potential impacts on healthcare access and equity.

9. Conclusions

This study shows how machine learning methods from financial fraud detection can be adapted for use in healthcare payment systems. By making changes to address the specific challenges of healthcare, such as medical coding and patient care requirements, our framework provides better results than traditional approaches. The experiments confirm improvements in F1-score, reduced false positives, higher fraud recovery rates, and strong financial benefits. Key elements such as medical code embeddings and temporal feature engineering proved especially useful, showing the importance of using healthcare-specific features in fraud detection systems.

Our research also offers practical guidance for healthcare organizations. We recommend focusing on medical code embeddings and temporal features when building fraud detection models, and using phased deployment to reduce risks. Although the system requires more computational resources, the improved detection accuracy and lower costs from fewer false positives make the investment worthwhile. The framework also supports regulatory compliance by ensuring interpretability and audit trails. Looking forward, future research can explore transfer learning, real-time adaptive models, federated learning, and ethical AI, helping to make fraud detection systems more robust, fair, and effective in protecting healthcare resources.

References

1. D. Cheng, Y. Zou, S. Xiang, and C. Jiang, "Graph neural networks for financial fraud detection: a review," *Frontiers of Computer Science*, vol. 19, no. 9, p. 199609, 2025.
2. A. du Preez, S. Bhattacharya, P. Beling, and E. Bowen, "Fraud detection in healthcare claims using machine learning: A systematic review," *Artificial Intelligence in Medicine*, vol. 160, p. 103061, Feb. 2025.
3. P. Adhikari, P. Hamal, and F. B. Jnr, "Artificial intelligence in fraud detection: Revolutionizing financial security," *International Journal of Science and Research Archive*, vol. 13, no. 1, pp. 1457–1472, 2024.
4. Z. Wang, X. Chen, Y. Wu, L. Jiang, S. Lin, and G. Qiu, "A robust and interpretable ensemble machine learning model for predicting healthcare insurance fraud," *Scientific Reports*, vol. 15, no. 1, p. 218, Jan. 2025.
5. N. Kumaraswamy, M. K. Markey, T. Ekin, J. C. Barner, and K. Rascati, "Healthcare fraud data mining methods: a look back and look ahead," *Perspectives in Health Information Management*, vol. 19, no. 1, pp. 1i, 2022.
6. V. Rezaei Tabar and M. Safakish, "Credit-card fraud detection: Cost-sensitive meta-learning Bayesian network classifiers," *Journal of Data Science and Modeling*, pp. 189–215, 2025.
7. E. Nabrawi and A. Alanazi, "Fraud detection in healthcare insurance claims using machine learning," *Risks*, vol. 11, no. 9, p. 160, 2023.
8. Y. Yoo, J. Shin, and S. Kyeong, "Medicare fraud detection using graph analysis: A comparative study of machine learning and graph neural networks," *IEEE Access*, vol. 11, pp. 88278–88294, 2023.
9. M. Bairy, B. Muniyal, and N. P. Shetty, "Enhancing healthcare data integrity: fraud detection using unsupervised learning techniques," *International Journal of Computers and Applications*, vol. 46, no. 11, pp. 1006–1019, 2024.
10. L. Hernandez Aros, L. X. Bustamante Molano, F. Gutierrez-Portela, J. J. Moreno Hernandez, and M. S. Rodríguez Barrero, "Financial fraud detection through the application of machine learning techniques: a literature review," *Humanities and Social Sciences Communications*, vol. 11, no. 1, pp. 1–22, 2024.
11. J. M. Johnson and T. M. Khoshgoftaar, "Data-centric AI for healthcare fraud detection," *SN Computer Science*, vol. 4, no. 4, p. 389, 2023.
12. J. Lu, K. Lin, R. Chen, and H. Zhang, "Health insurance fraud detection by using an attributed heterogeneous information network with a hierarchical attention mechanism," *BMC Medical Informatics and Decision Making*, vol. 23, no. 1, 2023.

 doi:10.20944/preprints202510.0409.v1

13. F. Moradi, M. Tarif, and M. Homaei, "A systematic review of machine learning in credit card fraud detection," *Preprint, MDPI AG*, 2025.
14. D. Sehrawat and Y. Singh, "Auto-encoder and LSTM-based credit card fraud detection," *SN Computer Science*, vol. 4, no. 5, p. 557, 2023.
15. F. Moradi, M. Tarif, and M. Homaei, "Robust fraud detection with ensemble learning: A case study on the IEEE-CIS dataset," *Preprint*, Jul. 2025.
16. M. H. Chagahi, N. Delfan, S. M. Dashtaki, B. Moshiri, and M. J. Piran, "Explainable AI for fraud detection: An attention-based ensemble of CNNs, GNNs, and a confidence-driven gating mechanism," *arXiv preprint arXiv:2410.09069*, 2024.
17. W. Siblini, G. Coter, R. Fabry, L. He-Guelton, F. Oblé, B. Lebichot, Y.-A. Le Borgne, and G. Bontempi, "Transfer learning for credit card fraud detection: A journey from research to production," *arXiv preprint arXiv:2107.09323*, 2021.
18. M. Herland, T. M. Khoshgoftaar, and R. A. Bauder, "Big data fraud detection using multiple medicare data sources," *Journal of Big Data*, vol. 5, no. 1, pp. 1–21, 2018.
19. J. M. Johnson and T. M. Khoshgoftaar, "Medicare fraud detection using neural networks," *Journal of Big Data*, vol. 6, no. 1, p. 63, 2019.
20. R. Bounab, K. Zarour, B. Guelib, and N. Khlifa, "Enhancing medicare fraud detection through machine learning: Addressing class imbalance with SMOTE-ENN," *IEEE Access*, vol. 12, pp. 54382–54396, 2024.
21. D. Kumar, P. Anitha, J. Murugachandravel, S. Jeevitha, A. Bhuvanesh, and P. P. P. Pawar, "Banking fraud detection using optimized enhanced stacked autoencoder approach," *Security and Privacy*, vol. 8, no. 4, p. e70054, 2025.
22. S. Park, S. Kim, and M. Cha, "Knowledge sharing via domain adaptation in customs fraud detection," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 11, pp. 12062–12070, 2022.
23. M. Lu, Z. Han, S. X. Rao, Z. Zhang, Y. Zhao, Y. Shan, R. Raghunathan, C. Zhang, and J. Jiang, "Bright-graph neural networks in real-time fraud detection," in *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pp. 3342–3351, 2022.
24. Z. Hamid, F. Khalique, S. Mahmood, A. Daud, A. Bukhari, and B. Alshemaimri, "Healthcare insurance fraud detection using data mining," *BMC Medical Informatics and Decision Making*, vol. 24, no. 1, p. 112, 2024.
25. F. Moradi, M. Tarif, and M. Homaei, "Semi-supervised supply chain fraud detection with unsupervised pre-filtering," *arXiv preprint arXiv:2508.06574*, 2025.
26. A. V. Najar, L. Alizamani, M. Zarqi, and E. Hooshmand, "A global scoping review on the patterns of medical fraud and abuse: Integrating data-driven detection, prevention, and legal responses," *Archives of Public Health*, vol. 83, no. 1, p. 43, 2025.
27. Y. Chen, K. Lu, S. Zheng, and X. Wang, "Health insurance fraud detection based on multi-channel heterogeneous graph structure learning," *Heliyon*, vol. 10, no. 9, p. e29735, 2024.
28. T. M. Khoshgoftaar, A. Dittman, R. A. Bauder, and N. Hasanin, "Explainable machine learning models for Medicare fraud detection," *Journal of Big Data*, vol. 10, no. 1, p. 138, 2023.
29. A. Shamsoshoara, M. Afghah, A. Razi, L. Zheng, P. Z. Fulé, and J. Bloebaum, "Next-generation machine learning in healthcare fraud detection: Current trends, challenges, and future research directions," *Information*, vol. 16, no. 9, p. 730, 2025.