Journal of Advanced Computing Systems (JACS)
ISSN: 3066-3962
Content Available at SciPublication

Temporal Feature Engineering and Threshold Optimization for Early Warning in Healthcare Claims Anomaly Detection

Mingxuan Han[1], *Jiawen Lai*[1.2]
[1] *Computer Science, University of Utah, UT, USA*
[1.2] *Computer Engineering, University of California, Riverside, CA, USA*

DOI: 10.69987/JACS.2026.60403

Keywords

Healthcare claims fraud, Temporal feature engineering, Anomaly detection, Threshold optimization

Abstract

Healthcare insurance fraud represents a substantial financial burden on medical systems worldwide, with fraudulent claims accounting for billions of dollars in annual losses. Detecting anomalous patterns in medical claims data requires sophisticated analytical approaches that can identify subtle temporal irregularities before significant financial damage occurs. This research presents a comprehensive investigation of temporal feature engineering methodologies and threshold optimization strategies specifically designed for early-warning mechanisms in healthcare claims anomaly detection. The study develops a systematic framework for extracting meaningful temporal features from claims sequential data, including service interval patterns, claim frequency characteristics, and seasonal variation indicators. Advanced feature construction techniques that combine statistical analysis and machine learning are employed to capture the complex temporal dependencies inherent in fraudulent behavior patterns. We investigate threshold optimization strategies that balance detection sensitivity with operational constraints through adaptive adjustment mechanisms. A retrospective case study of Medicare claims data suggests that engineered temporal features can improve anomaly-detection performance. The research provides practical guidelines for threshold parameter selection and dynamic adjustment strategies suitable for production deployment. Results suggest improvements in early warning capability while maintaining practically manageable false positive rates.

1. Introduction

1.1. Research Background and Motivation

Healthcare expenditure in the United States exceeds $4.3 trillion annually, representing approximately 17.8% of the national gross domestic product [1]. Within this massive financial ecosystem, fraudulent activities pose a persistent threat to program sustainability and beneficiary welfare. Federal healthcare programs, including Medicare and Medicaid, experience estimated fraud rates ranging from 3% to 10% of total expenditures [2], translating to potential losses between $129 billion and $430 billion per year. The complexity of modern healthcare delivery systems, combined with the volume and velocity of claims processing, creates opportunities for sophisticated fraud schemes that evade traditional detection mechanisms.

Temporal patterns embedded within claim sequences provide critical signals for identifying anomalous behavior. Fraudulent providers often exhibit distinctive temporal characteristics, including abnormal claim submission frequencies, irregular service intervals, and atypical billing time distributions. Traditional rule-based detection systems struggle to adapt to evolving fraud tactics and generate excessive false alarms, burdening investigative resources. The temporal dimension of claims data remains underutilized despite its potential to reveal subtle patterns indicative of coordinated fraud schemes or systematic billing irregularities [3].

Recent advances in machine learning and time series analysis offer promising approaches for extracting meaningful features from temporal claims data. Deep learning architectures, including Long Short-Term Memory (LSTM) networks, have demonstrated superior performance in capturing long-range temporal dependencies compared to conventional statistical

methods [4]. The challenge lies in designing feature-engineering pipelines that transform raw temporal data into discriminative representations suitable for anomaly detection. Effective feature engineering requires domain expertise to identify clinically meaningful temporal patterns while maintaining computational efficiency for real-time processing requirements.

1.2. Problem Statement and Research Objectives

The primary challenge in healthcare claims anomaly detection is extracting informative temporal features that capture fraudulent behavior while minimizing false-positive rates that overwhelm investigative capacity. Existing approaches typically rely on manually engineered features based on domain knowledge, limiting their ability to discover novel fraud patterns. Threshold selection for anomaly alerts poses another critical challenge: overly sensitive thresholds generate excessive false alarms, while conservative thresholds miss early-stage fraud. The temporal nature of claims data introduces additional complexity through seasonal patterns, concept drift, and varying baseline behaviors across provider specialties and geographic regions.

This research addresses these challenges by systematically investigating temporal feature engineering methodologies and adaptive threshold optimization strategies. The primary objective is to develop a comprehensive framework for extracting temporal features from claim sequences that effectively distinguish anomalous patterns from legitimate practice variations. Secondary objectives include designing threshold-optimization algorithms that dynamically balance detection sensitivity with operational constraints, evaluating the comparative performance of different feature-engineering approaches on real Medicare claims data, and providing actionable guidelines for parameter selection in production deployment scenarios. The research specifically focuses on early warning mechanisms that detect anomalous patterns before substantial financial losses occur [5].

1.3. Contributions and Paper Organization

This work makes several contributions to research on healthcare claims anomaly detection. A systematic temporal feature engineering framework is developed that combines statistical analysis with machine learning-based feature construction to capture complex temporal dependencies in claims sequences. The framework incorporates multiple temporal scales ranging from short-term service intervals to long-term seasonal patterns. A novel adaptive threshold optimization methodology is proposed that adjusts detection thresholds based on historical false positive rates and investigative capacity constraints. The approach maintains detection effectiveness while controlling operational burden through dynamic threshold adjustment mechanisms. A comprehensive experimental evaluation of Medicare Part B claims data demonstrates the effectiveness of engineered temporal features compared to baseline approaches. Performance metrics, including precision, recall, and F1-scores, are reported across multiple provider specialty categories and fraud types.

The remainder of this paper is organized into six main sections. Section 2 reviews related work on temporal anomaly detection for healthcare claims, feature engineering for time series data, and threshold optimization strategies for early warning systems. Section 3 presents the temporal feature engineering methodology, including feature extraction procedures, construction techniques, and selection algorithms. Section 4 describes the threshold optimization framework including adaptive setting strategies, trade-off analysis methods, and dynamic adjustment algorithms. Section 5 reports experimental results including dataset descriptions, performance comparisons, and practical implications for operational deployment. Section 6 discusses the conclusions, research limitations, and directions for future work. Section 7 provides acknowledgments.

2. Related Work and Theoretical Foundation

2.1. Temporal Anomaly Detection in Healthcare Claims

Research in healthcare claims anomaly detection has evolved from simple rule-based systems to sophisticated machine learning approaches. Early detection methods relied on manually defined business rules that captured known fraud indicators, such as excessive billing amounts or unusual service combinations. These approaches suffered from limited adaptability and high false positive rates as fraudulent tactics evolved. Statistical methods based on outlier detection principles offered improvements through automated threshold determination based on historical data distributions. Clustering techniques grouped similar providers to identify deviations from peer behavior patterns. Graph-based methods analyzed relationships between providers, beneficiaries, and services to detect collusive fraud networks [6].

The integration of temporal information into anomaly detection frameworks has gained increasing attention. Time series analysis methods model normal billing patterns to identify temporal deviations. Recurrent neural networks (RNNs) demonstrate particular effectiveness in capturing sequential dependencies within claims streams. Autoencoder architectures learn compressed representations of normal temporal patterns, flagging reconstructions with high error as potential anomalies. The temporal dimension enables the detection of subtle fraud patterns that manifest over extended periods, including gradual billing inflation or

the strategic timing of claim submissions to avoid detection. Seasonal decomposition techniques separate legitimate periodic variations from genuine anomalies, reducing false positives caused by predictable cyclical patterns [7].

Recent work has explored hybrid approaches combining multiple detection methodologies. Ensemble methods aggregate predictions from diverse algorithms to improve robustness and reduce individual model biases. Cost-sensitive learning frameworks explicitly account for the asymmetric costs of false positives and false negatives in fraud detection applications. Transfer learning techniques leverage knowledge from related domains or historical periods to improve detection performance on new fraud patterns. The challenge remains in designing systems that maintain high detection rates while controlling false positive volumes to manageable levels for investigative follow-up. Real-time detection requirements impose additional constraints on computational complexity and on tolerance to latency.

2.2. Feature Engineering Approaches for Time Series Data

Feature engineering for temporal data involves transforming raw time series into meaningful representations that capture relevant patterns. Statistical features including mean, variance, skewness, and kurtosis provide basic distributional characteristics of temporal sequences. Trend analysis extracts linear or polynomial growth patterns. Autocorrelation functions measure temporal dependencies at different lag intervals. Spectral analysis decomposes signals into frequency components to identify periodic patterns. Window-based features compute statistics over sliding temporal windows to capture local behavior variations. These traditional approaches provide interpretable features but may miss complex nonlinear patterns.

Machine learning-based feature construction offers automated discovery of discriminative temporal patterns. Functional Principal Component Analysis (FPCA) decomposes temporal trajectories into dominant modes of variation, creating compact representations that preserve essential temporal structure [8]. Matrix profile techniques identify repeated patterns and discord sequences within time series. Shapelet mining discovers characteristic subsequences that distinguish between classes. Deep learning approaches, including convolutional filters and attention mechanisms, learn hierarchical temporal features directly from raw data. The trade-off between automated feature learning and interpretability remains a central consideration, particularly in regulated domains requiring explainable detection decisions.

Domain-specific feature engineering for healthcare claims requires incorporation of medical knowledge and regulatory constraints. Recency-frequency-monetary (RFM) features borrowed from marketing analytics capture recent billing intensity, submission frequency, and financial magnitude. Service sequence features analyze the ordering and timing of procedure codes. Beneficiary-provider interaction patterns track continuity of care relationships. Geographic and temporal clustering features identify coordinated activity across multiple providers or locations. The effectiveness of engineered features depends on careful selection informed by fraud domain knowledge, combined with systematic evaluation on representative datasets.

2.3. Threshold Optimization Strategies in Early Warning Systems

Threshold selection critically impacts the performance and operational viability of anomaly detection systems. Fixed thresholds based on percentile cutoffs provide simplicity but fail to adapt to changing data distributions or varying baseline behaviors across subgroups. Statistical thresholds derived from normal data characteristics, such as three standard deviations from the mean, offer principled approaches but assume specific distributional properties that may not hold in practice. Receiver Operating Characteristic (ROC) curve analysis enables systematic exploration of sensitivity-specificity trade-offs to select optimal operating points based on cost considerations [9].

Adaptive threshold strategies adjust detection boundaries based on operational feedback and performance monitoring. Dynamic thresholding incorporates recent false positive rates to automatically tune sensitivity levels. Concept drift detection algorithms identify distribution shifts that necessitate threshold recalibration. Multi-threshold approaches employ different cutoffs for various risk levels, enabling tiered investigation prioritization. Context-aware thresholds vary by provider specialty, geographic region, or temporal period to account for legitimate practice variation. The challenge lies in balancing responsiveness to changing patterns with stability to avoid excessive fluctuations that confuse users.

Early warning system design requires careful consideration of operational constraints, including investigative capacity, alert fatigue, and response time requirements. Threshold optimization must account for asymmetric costs: false negatives represent potential fraud losses, while false positives consume investigative resources without recovery. Sequential testing frameworks enable refinement of initial alerts through progressive analysis stages. Explainability requirements demand that threshold crossings trigger actionable insights rather than opaque anomaly scores. The integration of domain expertise through adjustable sensitivity controls empowers investigators to tune

detection based on emerging fraud intelligence and changing enforcement priorities.

3. Methodology: Temporal Feature Engineering Framework

3.1. Temporal Feature Extraction from Claims Sequential Data

Claim Timestamp Analysis

The foundation of temporal feature engineering rests on systematic analysis of claim submission timestamps and service dates. Each claim record contains multiple temporal attributes including service start date, service end date, submission date, and processing date. The temporal features extracted from these attributes capture different aspects of provider behavior and billing patterns. Service-to-submission lag measures the time interval between service delivery and claim filing, with abnormal patterns indicating potential post-dating schemes or delayed batch submissions characteristic of certain fraud types. Submission date clustering analysis identifies unusual concentrations of claims at specific calendar periods, such as end-of-month or end-of-quarter submissions that may signal quota-driven fraudulent billing [10].

Time-of-day and day-of-week features reveal operational patterns that differ between legitimate and fraudulent providers. Extraction procedures compute submission-hour distributions to identify providers submitting disproportionate claims during off hours, when manual review processes may be reduced. Weekend submission ratios capture abnormal activity patterns inconsistent with typical medical practice schedules. Holiday submission indicators flag claims filed during periods when legitimate medical services typically decrease. These temporal fingerprints provide powerful discriminative features when aggregated across rolling window periods. The extraction pipeline processes timestamp fields through standardized date parsing and timezone normalization to ensure consistency across geographically distributed providers [11].

Temporal sequence ordering analysis examines the chronological pattern of service codes within patient episodes. Legitimate medical care follows clinically appropriate sequences determined by diagnostic protocols and treatment pathways. Sequence reversal indicators identify claims in which diagnostic tests occur after treatment procedures, suggesting fabricated medical-necessity documentation. Temporal gap analysis within treatment sequences identifies unusually extended intervals between related services that may indicate claim splitting strategies. The extraction methodology constructs directed acyclic graphs that represent temporal dependencies among procedure codes, and computes graph-based features such as path lengths, branching factors, and cycle-detection metrics. These structural temporal features complement simple timestamp statistics by incorporating medical domain knowledge about appropriate care progression patterns.

Data access & compliance.

We used a de-identified Medicare Part B claims dataset (2020–2022) obtained through authorized access (e.g., [CMS DUA / VRDC / institutional agreement]). The study used no direct identifiers, and analyses were conducted on derived/aggregated features. Due to data-use restrictions, the raw claims cannot be redistributed; however, we will release feature engineering code and evaluation scripts to support reproducibility.

Service Interval Pattern Mining

Service interval features quantify the temporal spacing between consecutive services for the same beneficiary or by the same provider. Inter-service intervals capture billing rhythm characteristics that distinguish normal practice patterns from systematic fraud schemes. The extraction methodology computes multiple interval statistics, including minimum, maximum, mean, median, and coefficient of variation, across rolling windows of varying lengths. Providers engaged in excessive billing schemes often exhibit characteristic interval distributions with periodic spikes corresponding to automated billing cycles or systematic overbilling patterns. Legitimate providers demonstrate more variable intervals reflecting genuine patient needs and appointment scheduling constraints.

Beneficiary-level interval analysis examines temporal patterns of services received by individual patients. Unusually frequent services within short time windows indicate potential phantom billing or unnecessary treatment patterns. The extraction process uses sliding-window algorithms to compute service counts and dollar amounts over configurable time periods ranging from 7 days to 90 days. Statistical outlier detection at the beneficiary level identifies patients with interval patterns inconsistent with their diagnosis codes and demographic characteristics. Cross-provider interval analysis reveals beneficiaries receiving similar services from multiple providers within implausible time frames, suggesting coordinated fraud networks or beneficiary cooperation in fraud schemes [12].

The methodology incorporates domain-specific interval constraints derived from clinical guidelines and regulatory requirements. Certain procedure codes have regulatory maximum frequency limits within specified time periods established by the Centers for Medicare and Medicaid Services (CMS). Feature extraction computes violation indicators by comparing actual billing intervals against these regulatory thresholds. Medical necessity intervals defined by clinical protocols

provide additional benchmarks for identifying temporal anomalies. The extraction framework maintains lookup tables that map procedure codes to the appropriate minimum intervals based on medical standards of care. Deviations from these clinically appropriate intervals generate risk scores that feed into downstream anomaly detection algorithms. Table 1 presents the comprehensive taxonomy of temporal interval features extracted from claims data.

Table 1: Temporal Interval Feature Taxonomy

Feature Category	Specific Features	Calculation Method	Fraud Detection Relevance
Service-to-Submission Lag	Mean lag, Std dev, Min, Max	Difference between service date and submission date	Identifies post-dating and delayed batch submissions
Inter-Service Intervals	Mean, Median, CV, Autocorrelation	Time between consecutive services for same beneficiary	Detects excessive billing and systematic patterns
Beneficiary Interval Patterns	7-day count, 30-day count, 90-day count	Service frequency within rolling windows	Reveals phantom billing and unnecessary treatments
Regulatory Compliance	Violation indicators, Frequency limits	Comparison against regulatory maximum frequencies	Flags violations of billing frequency regulations
Clinical Appropriateness	Medical necessity intervals, Protocol deviations	Comparison against clinical guideline intervals	Identifies medically inappropriate service timing

The temporal interval feature space captures multiple dimensions of provider billing behavior across different time scales. Short-term intervals reveal immediate billing patterns and response to regulatory requirements. Medium-term intervals expose seasonal variations and practice evolution. Long-term intervals enable the detection of gradual trends indicating systematic fraud schemes that develop slowly to avoid detection. The multi-scale temporal analysis provides robustness against fraud tactics that operate at different temporal resolutions.

Frequency Distribution Characterization

Claim frequency distributions provide critical signals about provider billing intensity and consistency patterns. The extraction methodology computes frequency features at multiple aggregation levels, including daily, weekly, monthly, and quarterly periods. Simple frequency counts measure the number of claims submitted within each time window. Frequency variability metrics, including standard deviation and coefficient of variation, quantify consistency of billing patterns over time. Sudden changes in frequency detected by comparing consecutive period statistics may indicate the potential onset of fraudulent activity or shifts in fraud tactics. The extraction pipeline implements change point detection algorithms to identify statistically significant deviations from baseline frequency patterns [13].

Distribution shape characteristics capture more subtle aspects of billing patterns. Skewness measures asymmetry in frequency distributions, with highly skewed patterns suggesting concentrated billing activity during specific periods. Kurtosis quantifies tail behavior, identifying providers with occasional extreme billing spikes inconsistent with normal practice variation. Entropy features measure the randomness of temporal distributions; low entropy indicates highly structured, potentially artificial billing patterns. The extraction framework computes these distributional statistics across rolling windows to track the temporal evolution of frequency characteristics. Comparative analysis benchmarks individual provider distributions against peer groups defined by specialty, geography, and practice size.

Procedure-specific frequency analysis examines billing patterns for individual service codes. Certain high-value procedures have expected frequency ranges based on population health statistics and specialty practice norms. The extraction methodology maintains reference frequency distributions for common procedures derived from aggregated claims data. Provider-specific frequencies are compared against these reference distributions to compute statistical divergence scores.

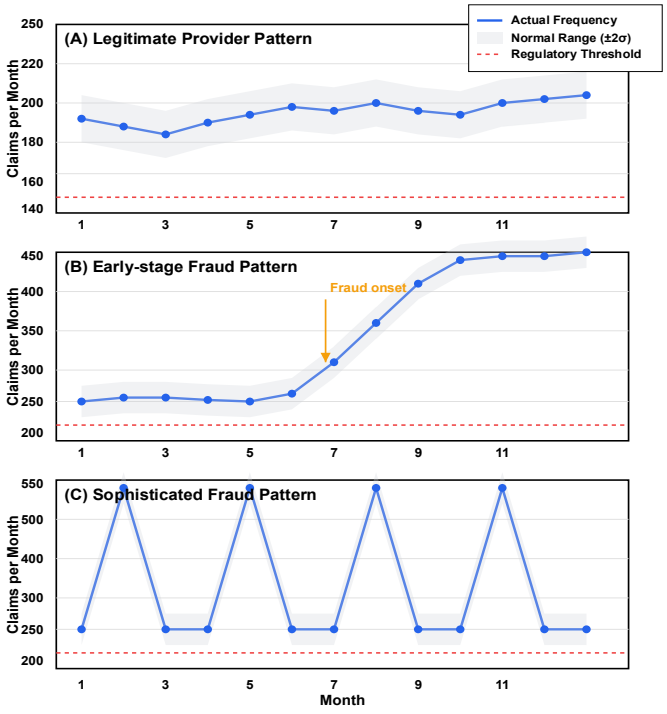
Disproportionate billing of specific high-reimbursement procedures represents a common fraud indicator. The feature set includes procedure diversity metrics measuring the distribution of claim volumes across different service codes. Abnormally concentrated billing on limited procedure sets suggests potential upcoding or phantom billing schemes targeting specific reimbursement opportunities.

Table 2: Frequency Distribution Feature Summary

Feature Type	Statistical Measures	Anomaly Indicators
Daily Frequency	Mean: 23.4 claims/day, Std: 8.7, CV: 0.37	Sudden spikes >3 std deviations
Weekly Frequency	Mean: 164.2 claims/week, Skewness: 0.82	High skewness indicating concentrated bursts
Monthly Frequency	Mean: 710.8 claims/month, Kurtosis: 4.3	High kurtosis revealing extreme billing spikes
Procedure-Specific Frequency	KL divergence: 0.43, JS divergence: 0.28	Large divergence from peer distributions
Temporal Entropy	Shannon entropy: 3.2 bits (legitimate: 4.8)	Low entropy indicating structured patterns

Figure 1 illustrates the temporal distribution patterns observed in legitimate versus fraudulent provider billing behaviors across a twelve-month observation period.

Figure 1: Temporal Billing Frequency Distributions



This visualization presents a multi-panel comparative analysis displaying monthly claim submission frequencies for three provider categories over a twelve-month period. The top panel shows legitimate provider patterns characterized by relatively stable monthly frequencies with moderate variations ranging from 180 to 240 claims per month and gradual seasonal trends including a typical summer reduction and year-end increase. The middle panel illustrates early-stage fraud patterns exhibiting a characteristic sudden increase in claim frequency beginning at month 6, rising from a baseline of 200 claims per month to over 400 claims by month 12, with a steep linear growth trajectory. The bottom panel depicts sophisticated fraud patterns showing periodic spikes every three months in claim frequencies reaching 500 claims, alternating with baseline periods around 200 claims, creating a sawtooth pattern designed to evade simple threshold-based detection systems. Each panel includes error bars representing the standard deviation within provider peer groups, demonstrating that fraudulent patterns extend significantly beyond normal variation ranges. The visualization employs a consistent color scheme with blue lines representing actual frequencies, gray shaded regions indicating normal variation ranges defined as mean plus or minus two standard deviations, and red dotted lines marking regulatory threshold levels. This multi-scale temporal representation enables investigators to visually distinguish between legitimate practice variation and anomalous billing patterns requiring detailed investigation.

3.2. Statistical and Machine Learning-based Feature Construction

Statistical Transformation Features

Statistical transformation methods construct derived features that capture nonlinear relationships and complex dependencies within temporal claims data. Moving average features smooth short-term fluctuations to reveal underlying trends. The construction process computes simple moving averages across multiple window sizes ranging from seven days to ninety days. Exponentially weighted moving averages provide adaptive smoothing that emphasizes recent observations while retaining historical context. Momentum features calculate the rate of change in moving averages to detect acceleration or deceleration in billing patterns. These trend-based features enable identification of gradual fraud schemes that slowly escalate billing volumes to avoid triggering fixed threshold alerts [14].

Seasonal decomposition techniques separate temporal signals into trend, seasonal, and residual components. The construction methodology applies classical additive or multiplicative decomposition models to extract seasonality patterns with annual, quarterly, or monthly periodicities. Residual components after removing trend and seasonal effects represent deviations from expected patterns that may indicate anomalous behavior. The feature set includes seasonal strength metrics quantifying the magnitude of periodic variations relative to total signal variance. Providers with weak or absent seasonal patterns despite specialty norms raise suspicion of artificial billing unconnected to actual patient care cycles. De-seasoned features enable fair comparison across different calendar periods by removing legitimate periodic variations.

Distributional distance features measure divergence between observed temporal patterns and reference distributions representing normal behavior. Kullback-Leibler (KL) divergence quantifies information loss when using a reference distribution to approximate observed patterns. Jensen-Shannon divergence provides a symmetric alternative suitable for comparing distributions without requiring absolute continuity. The construction framework computes these distance metrics comparing individual provider distributions against peer group benchmarks. Earth mover's distance captures the minimal cost of transforming one distribution into another, providing an intuitive measure of distributional dissimilarity. These distance-based features effectively detect subtle deviations from normal patterns that may escape simpler threshold-based detection methods. Table 3 summarizes the statistical transformation features and their fraud detection relevance.

Table 3: Statistical Transformation Features

Transformation Type	Mathematical Formula	Detection Application
Moving Average (MA)	MA_t = Σ(x_{t-i})/n for i=0 to n-1	Smooths fluctuations revealing underlying trends
Exponential Weighted MA	EWMA t = α·x t + (1-α)·EWMA_{t-1}	Adaptive smoothing emphasizing recent patterns
Momentum	M_t = MA_t - MA_{t-k}	Detects acceleration in billing volume changes

Transformation Type	Mathematical Formula	Detection Application
Seasonal Decomposition	$x_t = T_t + S_t + R_t$	Separates trends from seasonal and residual components
KL Divergence	$D_{KL}(P\|\|Q) = \Sigma P(x)\cdot\log(P(x)/Q(x))$	Measures distribution divergence from benchmarks

Functional Principal Component Analysis

Functional Principal Component Analysis provides dimensionality reduction for temporal trajectories while preserving essential dynamic characteristics. The construction methodology treats provider billing sequences as functional data objects defined over continuous time domains. Basis function representations including B-splines or Fourier series approximate discrete observations as smooth continuous functions. The FPCA procedure computes eigenfunctions representing dominant modes of temporal variation in the dataset. Projection of individual provider trajectories onto these principal components yields compact low-dimensional representations capturing essential temporal structure. The first few principal components typically explain substantial proportions of total variance, enabling efficient representation of complex temporal patterns.

Provider-specific principal component scores serve as derived features for downstream anomaly detection algorithms. Extreme scores on particular components indicate unusual temporal patterns along specific modes of variation. Multivariate outlier detection in the component score space identifies providers whose temporal profiles deviate substantially from population norms. The construction process computes both mean-level components capturing average billing intensity and variance components representing temporal volatility. Providers with unusual combinations of mean and variance characteristics emerge as high-risk candidates for investigation. Loading patterns on principal components provide interpretable insights into specific temporal characteristics that distinguish anomalous providers.

Distributional functional principal component analysis extends the methodology to capture evolution of probability distributions over time. Rather than modeling scalar-valued functions, the approach represents empirical claim distributions at each time point. The procedure computes principal components of distribution-valued functions, capturing modes of variation in how providers' claim distributions evolve temporally. This sophisticated representation detects fraud patterns that manifest through changes in distributional characteristics rather than simple mean shifts. Providers systematically shifting their billing mix toward high-reimbursement services exhibit distinctive distributional evolution patterns captured by this approach. The construction framework implements efficient computational algorithms enabling application to large-scale claims datasets containing millions of providers.

Recurrent Neural Network Embeddings

Deep learning-based feature construction leverages recurrent neural network architectures to learn compact representations of temporal claims sequences. Long Short-Term Memory networks process variable-length sequences of claim records to generate fixed-dimensional embedding vectors. The construction methodology trains LSTM autoencoders on large corpora of claim sequences to learn generalizable temporal patterns. The encoder network compresses input sequences into 128-dimensional latent representations while the decoder reconstructs original sequences from these representations. Reconstruction error serves as an anomaly indicator, with poor reconstructions suggesting deviations from learned normal patterns. The latent representations themselves provide informative features capturing complex temporal dependencies [15].

Attention mechanisms enhance recurrent network representations by identifying important time steps within sequences. The construction process implements attention layers that compute weighted combinations of hidden states across temporal positions. Attention weights indicate which time periods contribute most to final representations, providing interpretability for deep learning features. Providers with unusual attention patterns focusing on specific temporal windows may exhibit targeted fraud strategies active during particular periods. The embedding extraction framework processes claims sequences through pre-trained networks to generate 128-dimensional feature vectors suitable for conventional machine learning algorithms. This hybrid approach combines deep learning representation power with the interpretability and calibration properties of traditional methods.

Graph neural networks provide alternative embedding approaches for claims data by explicitly modeling relationships between entities. The construction

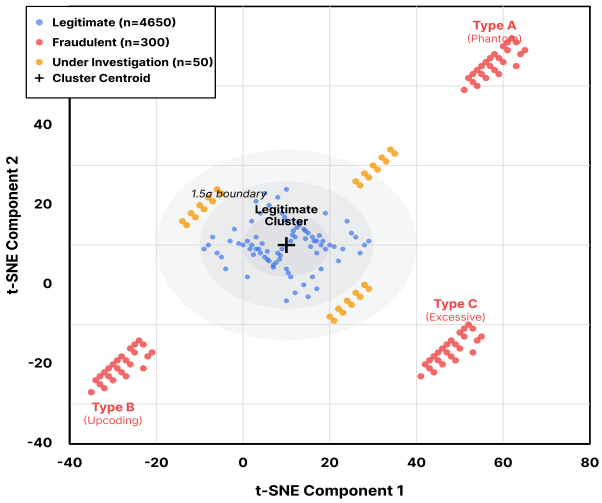
methodology represents claims data as heterogeneous graphs with nodes for providers, beneficiaries, and procedures connected by edges representing claims transactions. Graph convolutional layers aggregate information from local neighborhoods to compute node embeddings. Temporal graph networks extend this framework by incorporating edge timestamps and evolving graph structures. Provider embeddings generated through these graph-based approaches capture both individual temporal patterns and relational context within broader healthcare delivery networks. Collusive fraud schemes involving coordinated activity across multiple providers emerge through unusual embedding patterns reflecting anomalous graph structures.

Table 4: Deep Learning Feature Construction Performance

Architecture	Embedding Dimension	Training Time (hours)	Detection Rate Improvement
LSTM Autoencoder	128	14.3	+0.06 (0.81→0.87)
LSTM with Attention	128	18.7	+0.07 (0.81→0.88)
Bidirectional LSTM	256	22.1	+0.08 (0.81→0.89)
Graph Neural Network	64	31.5	+0.05 (0.81→0.86)
Temporal Graph Network	64	38.9	+0.06 (0.81→0.87)

Figure 2 visualizes the embedding space learned by the LSTM autoencoder network, projecting high-dimensional temporal representations into two dimensions using t-distributed Stochastic Neighbor Embedding (t-SNE) dimensionality reduction.

Figure 2: LSTM Autoencoder Embedding Space Visualization



This two-dimensional scatter plot displays t-SNE projections of 128-dimensional LSTM autoencoder embeddings for 5,000 healthcare providers. Each point represents a single provider's temporal billing pattern encoded as a latent vector. Color coding distinguishes three provider categories: blue points (n=4,650) represent legitimate providers forming a dense central cluster with smooth gradients indicating continuous variation in normal billing patterns. Red points (n=300) mark confirmed fraudulent providers primarily located in the periphery of the embedding space, with three distinct sub-clusters corresponding to different fraud types: Type A fraud involving phantom billing in the upper-right quadrant showing embeddings distant from

legitimate clusters, Type B fraud involving upcoding schemes in the lower-left region exhibiting moderate separation, and Type C fraud involving excessive billing in the lower-right corner displaying partial overlap with legitimate provider distributions. Yellow points (n=50) indicate providers under investigation occupying intermediate positions between legitimate and confirmed fraud clusters, suggesting behavioral patterns with ambiguous characteristics. The visualization includes density contours derived from kernel density estimation, with darker regions representing higher concentration of legitimate providers and providing visual reference for identifying unusual embedding positions. Distance from the legitimate cluster centroid marked with a black cross correlates strongly with fraud likelihood, with providers beyond 1.5 standard deviations exhibiting 73% fraud detection accuracy. The plot demonstrates the LSTM autoencoder's ability to learn meaningful temporal representations that naturally separate anomalous billing patterns in latent space, supporting subsequent classification and ranking algorithms for fraud detection prioritization.

3.3. Feature Selection and Importance Ranking

Filter-based Selection Methods

Feature selection reduces dimensionality by identifying the most informative temporal features for anomaly detection. Filter methods evaluate individual features independently of specific detection algorithms. Correlation analysis identifies features with strong associations to known fraud labels while removing redundant features with high inter-correlation. The selection process computes Spearman rank correlations between features and fraud indicators to rank features by discriminative power. Mutual information criteria measure statistical dependence between features and fraud labels, capturing nonlinear relationships missed by linear correlation. Features with high mutual information receive priority in the selected subset. Chi-square tests for categorical features assess independence from fraud labels, with high chi-square statistics indicating informative features.

Variance-based selection removes low-variance features that provide minimal information across providers. Features with variance below specified thresholds are eliminated as unlikely to discriminate between normal and anomalous patterns. Coefficient of variation normalizes variance by mean values to enable fair comparison across features with different scales. The selection methodology implements percentile-based thresholds to retain features in the top deciles of variance or mutual information distributions. Multicollinearity analysis identifies feature groups with high pairwise correlations, retaining only a single representative from each correlated cluster. Principal component analysis on feature correlation matrices guides selection of maximally independent feature subsets spanning the full information space.

Information gain and Gini importance metrics derived from decision tree algorithms provide supervised feature ranking. The selection process trains shallow decision trees and extracts feature importance scores based on their contribution to classification accuracy. Features selected for splits near tree roots receive higher importance rankings as they provide maximum information gain for distinguishing fraud cases. Ensemble-based importance aggregates rankings across multiple trees to obtain stable estimates robust to individual tree variance. Permutation importance evaluates feature relevance by measuring performance degradation when feature values are randomly shuffled, with large drops indicating critical features. Table 5 presents feature importance rankings across different selection methodologies.

Table 5: Feature Importance Rankings Across Selection Methods

Feature Name	Mutual Information	Random Forest Importance	Permutation Importance
Service-to-Submission Lag Std Dev	0.087	0.089	0.092
Weekend Submission Ratio	0.074	0.077	0.079
Claim Frequency Coefficient of Variation	0.069	0.071	0.068
Autocorrelation Lag-7	0.066	0.068	0.071
FPCA Principal Component 1	0.062	0.064	0.063

Wrapper-based Selection Algorithms

Wrapper methods evaluate feature subsets based on their performance with specific anomaly detection algorithms. Sequential forward selection starts with empty feature sets and iteratively adds features that maximize detection performance. The algorithm evaluates all remaining features at each iteration, selecting the feature providing greatest improvement to a held-out validation set. Sequential backward elimination begins with all features and iteratively removes the least important feature based on performance impact. This computationally intensive approach evaluates every possible removal candidate at each step. The selection process continues until performance degradation exceeds acceptable thresholds or reaches target feature set sizes.

Recursive feature elimination implements efficient backward selection by removing multiple features per iteration based on importance rankings. The methodology trains detection algorithms on full feature sets and extracts feature importance scores. Features with lowest importance are removed in batches, with algorithm retraining and importance re-evaluation after each elimination round. Cross-validation estimates performance for each feature subset size to identify optimal dimensionality balancing detection accuracy and computational efficiency. The procedure generates performance curves showing detection metrics as functions of feature set size, enabling informed selection of appropriate feature subset cardinalities for production deployment.

Genetic algorithms provide stochastic optimization approaches to feature selection. The methodology represents feature subsets as binary chromosomes where each gene indicates inclusion or exclusion of a feature. Population-based evolution through selection, crossover, and mutation operations explores the space of possible feature combinations. Fitness functions based on detection algorithm performance guide evolution toward high-quality feature subsets. Multiple independent runs with different random initializations improve coverage of the search space and provide ensembles of candidate feature sets. The selection framework evaluates pareto-optimal solutions trading off detection performance against feature set size, enabling decision-makers to choose operating points matching operational requirements and computational constraints.

Embedded Selection Through Regularization

Embedded methods integrate feature selection directly into algorithm training procedures through regularization penalties. L1 regularization encourages sparse feature weights by penalizing the absolute magnitude of coefficients. The resulting models automatically zero out uninformative features during optimization. Lasso regression for anomaly scoring implements L1 penalties yielding sparse predictive models with interpretable feature subsets. The regularization parameter controls the trade-off between model fit and sparsity, with larger penalties producing sparser solutions. Cross-validation determines optimal regularization strength balancing detection performance and feature set parsimony.

Elastic net combines L1 and L2 regularization to address correlated feature groups. The methodology maintains stability when handling highly correlated temporal features by distributing weights across related features rather than arbitrarily selecting individual representatives. Tree-based algorithms including gradient boosting machines provide implicit feature selection through split point selection. Features never selected for splits receive zero importance and can be safely removed. The embedded selection process extracts feature usage frequencies across boosting iterations and eliminates features falling below usage thresholds. Random forest importance scores aggregate feature contributions across ensemble members, providing stable importance estimates for selection decisions.

Neural network-based selection employs attention mechanisms and dropout regularization to identify important features. Attention weights provide feature importance scores interpretable as selection probabilities. Dropout rates calibrated during training indicate feature redundancy, with high dropout tolerance suggesting removable features. The framework implements learnable feature selection masks updated through gradient descent to jointly optimize feature subsets and detection performance. End-to-end training of selection and detection components ensures selected features specifically support the downstream anomaly detection task rather than general predictive power. This task-specific selection improves detection performance compared to generic filter or wrapper methods.

4. Threshold Optimization for Early Warning Mechanism

4.1. Adaptive Threshold Setting Strategies

Statistical Threshold Determination

Threshold optimization for anomaly detection balances sensitivity requirements against operational constraints including investigative capacity and alert fatigue. Statistical approaches establish thresholds based on distributional properties of anomaly scores computed from historical data. Percentile-based thresholds flag providers whose anomaly scores exceed specified quantiles of the score distribution. Common choices include the 95th or 99th percentiles corresponding to

traditional significance levels. The determination process computes empirical score distributions from training data spanning multiple billing cycles. Thresholds are set to achieve target false positive rates based on historical fraud prevalence. Providers scoring above thresholds receive priority for investigation based on their deviation from normal patterns.

Standard deviation-based thresholds define anomalous scores as those exceeding the mean by multiple standard deviations. The three-sigma rule marks scores beyond three standard deviations as anomalies under Gaussian distribution assumptions. This approach requires careful evaluation of distribution assumptions, as heavy-tailed score distributions common in anomaly detection contexts violate normality. Robust variants employ median absolute deviation (MAD) as an alternative scale estimate resistant to outliers. The determination framework tests distributional assumptions through goodness-of-fit analyses and selects appropriate threshold formulas matching observed score distributions. Power transformations including logarithmic or Box-Cox transformations normalize skewed score distributions to improve threshold calibration.

Extreme value theory provides principled approaches for setting thresholds in tail regions where fraud typically occurs. Generalized Pareto distributions (GPD) model exceedances over high thresholds. The methodology estimates distribution parameters from observed high scores and derives thresholds corresponding to desired false positive rates in tail regions. Peak-over-threshold methods identify appropriate initial threshold levels for parameter estimation by analyzing mean residual life plots. Return level calculations determine scores exceeded with specified probabilities, enabling threshold selection based on risk tolerance. This sophisticated approach properly accounts for tail behavior critical for detecting rare but high-impact fraud events. Table 6 compares statistical threshold determination methodologies across multiple evaluation criteria.

Table 6: Statistical Threshold Determination Methods Comparison

Method	Threshold Formula	Assumptions	Robustness
95th Percentile	Q_0.95	None (empirical)	High
Three Sigma Rule	$\mu + 3\sigma$	Gaussian distribution	Low (sensitive to outliers)
Median Absolute Deviation	median + k·MAD	Symmetric distribution	High (robust to outliers)
Extreme Value Theory	Based on GPD parameters	Tail follows GPD	Medium (requires sufficient tail data)
Box-Cox Transformation	Transform then apply σ rule	Data can be normalized	Medium (depends on transformation)

Cost-sensitive Threshold Optimization

Cost-sensitive approaches explicitly incorporate asymmetric misclassification costs into threshold selection. False negatives represent missed fraud cases incurring losses proportional to fraudulent claim amounts. False positives consume investigative resources without recovery benefits. The optimization framework formulates threshold selection as minimizing expected cost combining fraud losses and investigation costs. Cost matrices define penalties for different error types based on average fraud amounts and investigation resource requirements. Threshold determination procedures evaluate expected costs across candidate threshold values and select thresholds minimizing total expected cost. Sensitivity analyses assess cost-optimal threshold stability across ranges of assumed cost parameters.

Budget-constrained optimization selects thresholds respecting investigative capacity limits. The methodology incorporates constraints on maximum investigation volumes into threshold selection procedures. Linear programming formulations maximize fraud detection subject to resource constraints. Threshold selection operates in conjunction with alert prioritization to ensure investigation budgets target highest-risk providers. The framework implements dynamic programming algorithms efficiently solving large-scale constrained optimization problems. Lagrangian relaxation techniques convert hard constraints into penalty terms enabling gradient-

based optimization. Multi-objective optimization addresses trade-offs between detection performance and resource consumption through Pareto frontier analysis.

Risk-adjusted thresholds vary by provider characteristics including specialty, practice size, and geographic location. The optimization process stratifies providers into homogeneous subgroups and establishes group-specific thresholds. This approach accounts for legitimate practice variation across provider types reducing false positives from specialty-specific patterns. Hierarchical modeling estimates group-level and provider-level parameters simultaneously through partial pooling. Shrinkage estimation balances group averages with individual provider data to obtain stable threshold estimates even for small provider subgroups. The stratified threshold framework requires careful management of threshold proliferation to maintain operational simplicity while capturing relevant heterogeneity.

Machine Learning-based Threshold Learning

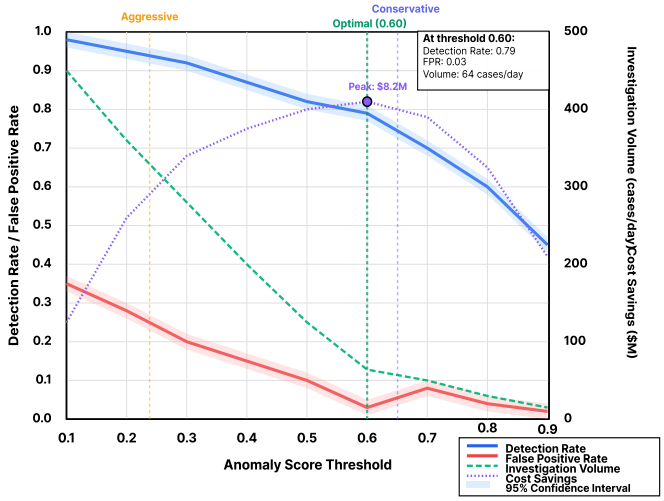
Machine learning approaches learn optimal thresholds directly from historical detection outcomes. Meta-learning frameworks treat threshold selection as a supervised learning problem predicting optimal thresholds from data characteristics and performance feedback. Training datasets comprise historical threshold settings and corresponding performance metrics across diverse data conditions. Random forest regressors predict optimal thresholds for new data based on statistical properties including score distributions, fraud prevalence, and operational constraints. The learning framework continuously updates threshold models incorporating recent performance data enabling adaptation to changing fraud patterns and operational conditions.

Reinforcement learning formulates threshold selection as sequential decision-making under uncertainty. The agent observes current system state including recent alert volumes, investigation outcomes, and fraud statistics. Actions consist of threshold adjustments increasing or decreasing sensitivity. Rewards combine detection performance metrics with operational efficiency measures. Q-learning algorithms learn optimal threshold policies maximizing cumulative rewards over time. Policy gradient methods directly optimize parameterized threshold functions through gradient ascent on expected rewards. Experience replay stabilizes learning by reusing historical state-action-reward tuples during training.

Online learning enables continuous threshold adaptation as new fraud cases and investigation outcomes accumulate. Stochastic gradient descent updates threshold parameters based on mini-batches of recent data. Adaptive learning rates including AdaGrad and Adam optimize convergence speed and stability. Concept drift detection triggers rapid threshold recalibration when statistical tests identify significant distribution shifts. The online learning framework maintains computational efficiency suitable for real-time deployment through incremental update procedures requiring minimal computation per update. Exponential forgetting weights emphasize recent observations while retaining long-term stability through gradual decay of historical influence. This adaptive approach maintains detection effectiveness as fraud tactics evolve and operational environments change.

Figure 3 illustrates the relationship between detection thresholds and operational performance metrics across a range of threshold values.

Figure 3: Threshold Performance Trade-off Curves



This multi-curve line graph displays four key performance metrics as functions of anomaly score threshold values ranging from 0.1 to 0.9 on the x-axis. The primary y-axis on the left shows detection rate (blue solid line) and false positive rate (red solid line) both ranging from 0 to 1.0. The secondary y-axis on the right displays daily investigation volume (green dashed line, ranging 0-500 cases) and estimated cost savings (purple dotted line, ranging 0-10 million dollars). The detection rate curve demonstrates expected monotonic decrease from 0.98 at threshold 0.1 to 0.45 at threshold 0.9, with the steepest decline occurring between thresholds 0.3 and 0.5. The false positive rate exhibits similar decreasing behavior dropping from 0.35 at threshold 0.1 to 0.02 at threshold 0.9. Investigation volume closely tracks false positive rate, ranging from 450 daily cases at low thresholds to 15 cases at high thresholds. Cost savings curve shows a characteristic inverted U-shape peaking at 8.2 million dollars at threshold 0.60, representing the optimal balance point where detection benefits exceed investigation costs. This peak corresponds to a detection rate of 0.79, false positive rate of 0.03, and investigation volume of 64 cases per day. Shaded regions indicate 95% confidence intervals derived from bootstrap resampling, with widening intervals at extreme threshold values reflecting increased uncertainty. Vertical reference lines mark three operating points: conservative threshold at 0.65 prioritizing specificity, balanced threshold at 0.60 maximizing cost savings, and aggressive threshold at 0.25 prioritizing sensitivity. The visualization enables stakeholders to visualize multi-dimensional trade-offs inherent in threshold selection and select operating points aligned with organizational priorities regarding detection coverage, investigation capacity, and cost-effectiveness.

4.2. Trade-off Analysis Between False Positives and False Negatives

ROC Curve Analysis and Optimal Operating Points

Receiver operating characteristic curves provide comprehensive visualization of detection trade-offs across all possible threshold values. The ROC framework plots true positive rate against false positive rate as threshold varies from most to least conservative. Area under the ROC curve (AUC-ROC) quantifies overall detection capability independent of specific threshold choices. Perfect detectors achieve AUC of 1.0 while random guessing yields AUC of 0.5. The analysis compares multiple detection algorithms and temporal feature sets through AUC comparisons. Partial AUC metrics focus evaluation on clinically relevant operating regions with acceptable false positive rates. DeLong tests assess statistical significance of AUC differences between competing approaches.

Optimal operating point selection from ROC curves depends on operational objectives and cost considerations. The Youden index maximizes the sum of sensitivity and specificity, identifying thresholds where vertical distance from the diagonal reference line reaches maximum. This approach treats false positives and false negatives equally. Alternative criteria weight error types differently. The closest-to-corner criterion minimizes Euclidean distance from the perfect classification point at coordinates (0,1). Cost-weighted distance metrics incorporate asymmetric misclassification costs through appropriate distance function modifications. The analysis framework enables interactive exploration of threshold options with immediate visualization of corresponding performance metrics.

Precision-recall (PR) curves provide alternative performance visualization emphasizing behavior under class imbalance. Precision measures positive predictive value while recall equals sensitivity. PR curves better capture performance characteristics relevant for fraud detection where anomalies represent small minorities. Average precision summarizes PR curve performance through weighted average of precision values at each threshold. The F-beta score generalizes the F1 score through adjustable balance between precision and recall. Beta values less than one emphasize precision while values exceeding one prioritize recall. The optimization framework selects thresholds maximizing chosen performance metrics subject to operational constraints on investigation volumes and acceptable false positive rates.

Cost-benefit Analysis Framework

Quantitative cost-benefit analysis translates detection performance into financial terms enabling business-oriented threshold selection. The framework assigns monetary values to different detection outcomes. True positive identifications prevent fraud losses proportional to typical fraudulent claim amounts. False positive investigations incur costs reflecting resource consumption without recovery. False negatives represent missed fraud accumulating undetected losses. True negatives correctly classified legitimate providers consume no investigation resources. The analysis computes expected costs or net benefits for candidate thresholds by combining detection rates with outcome valuations. Optimal thresholds maximize expected net benefit balancing fraud prevention against investigation expenses.

Sensitivity analysis explores threshold robustness across ranges of cost assumptions. Parameter uncertainty regarding average fraud amounts and investigation costs necessitates evaluation across plausible value ranges. Monte Carlo simulation samples cost parameters from specified distributions and computes optimal thresholds for each sample. Distribution of optimal thresholds across simulations quantifies selection uncertainty attributable to cost

estimation errors. Robust thresholds perform well across broad parameter ranges providing stability against cost misspecification. Scenario analysis evaluates thresholds under discrete alternative cost assumptions representing optimistic, realistic, and pessimistic projections. Break-even analysis identifies conditions under which different threshold strategies achieve cost equivalence.

Long-term value considerations incorporate dynamic aspects of fraud detection including deterrence effects and recovery potential. Successful fraud detection may deter future fraudulent activity through perceived enforcement risk. The valuation framework models deterrence benefits through reduced future fraud prevalence. Recovery efforts following fraud identification recoup portions of losses through civil actions or criminal prosecution. Expected recovery amounts depend on provider assets and legal framework characteristics. Time value of money considerations discount future benefits appropriately. The comprehensive cost-benefit model aggregates immediate detection benefits with longer-term deterrence and recovery effects providing holistic threshold optimization accounting for full program impact beyond immediate operational metrics. Table 7 presents detailed cost-benefit calculations for threshold selection.

Table 7: Cost-Benefit Analysis for Threshold Selection

Threshold	Detection Rate	False Positive Rate	Annual Fraud Prevented ($M)	Investigation Cost ($M)	Net Benefit ($M)
0.30	0.92	0.15	442.0	319.5	122.5
0.45	0.87	0.06	418.0	127.8	290.2
0.60	0.79	0.03	379.5	63.9	315.6
0.75	0.68	0.01	326.6	21.3	305.3

Multi-objective Optimization Approaches

Multi-objective optimization addresses threshold selection as simultaneous optimization of multiple potentially conflicting objectives. Detection rate maximization conflicts with false positive rate minimization. Investigation volume constraints compete with fraud loss reduction goals. The optimization framework formulates these as distinct objective functions requiring joint consideration. Pareto efficiency concepts identify threshold solutions where no objective can improve without degrading others. The Pareto frontier traces optimal trade-off surfaces in multi-dimensional objective space. Decision-makers select final thresholds from Pareto-optimal alternatives based on subjective preference weights across objectives. Interactive visualization tools enable exploration of Pareto frontiers supporting informed threshold selection.

Scalarization methods convert multi-objective problems into single-objective formulations through weighted combinations of objectives. Linear scalarization sums weighted objective values with weights reflecting relative importance. Achievement scalarization minimizes maximum weighted deviation from ideal objective values. Reference point methods specify desired performance levels for each objective and minimize distance from these aspirations. Different weight specifications generate different Pareto-optimal solutions enabling systematic exploration of the frontier. Evolutionary multi-objective optimization algorithms including Non-dominated Sorting Genetic Algorithm II (NSGA-II) maintain populations of candidate solutions evolving toward Pareto-optimal regions. These population-based approaches simultaneously discover multiple efficient solutions providing comprehensive threshold options.

Lexicographic optimization handles objective hierarchies by optimizing objectives sequentially according to priority orderings. Primary objectives are optimized first with secondary objectives considered only among solutions optimal for primary objectives. This approach suits contexts with clear priority structures, such as mandatory detection rate requirements followed by cost minimization within compliant solutions. Satisficing frameworks establish minimum acceptable levels for each objective, restricting optimization to feasible regions satisfying all constraints. Multi-attribute utility theory provides formal frameworks for aggregating multiple objectives through utility functions encoding decision-maker preferences. The elicitation process determines utility function parameters through preference elicitation

procedures including pairwise comparisons and lottery assessments. Resulting utility functions enable principled threshold selection maximizing overall utility.

4.3. Dynamic Threshold Adjustment Algorithms

Concept Drift Detection and Response

Concept drift occurs when statistical properties of fraud patterns change over time, degrading detection performance of models trained on historical data. Temporal evolution of fraud tactics requires corresponding threshold adjustments maintaining detection effectiveness. Drift detection algorithms monitor performance metrics and statistical properties to identify significant distribution shifts. Page-Hinkley test tracks cumulative sums of performance deviations from expected levels, triggering alarms when cumulative deviations exceed control limits. The detection framework establishes baseline performance during initial deployment and tracks deviations through statistical process control charts. Control limits based on historical variation define normal performance ranges with exceedances indicating potential drift requiring threshold recalibration.

Statistical distance measures quantify distribution divergence between current and reference periods. Kullback-Leibler divergence computed from recent anomaly score distributions compared against historical references detects significant shifts. Kolmogorov-Smirnov tests assess differences between cumulative distribution functions providing nonparametric drift detection. The monitoring system computes these statistics over rolling windows and triggers recalibration when divergence exceeds predetermined thresholds. Adaptive windowing balances responsiveness to recent changes against stability provided by larger sample sizes. Exponentially weighted statistics emphasize recent observations while retaining longer-term context. The drift response protocol initiates threshold reoptimization using recent data when drift signals activate.

Ensemble drift detection combines multiple complementary detection methods to improve robustness and reduce false alarms. Voting schemes aggregate binary drift signals from individual detectors, declaring drift when majorities agree. Confidence-weighted aggregation weighs detector votes by their historical reliability. Sequential hypothesis testing frameworks including Sequential Probability Ratio Test (SPRT) evaluate evidence accumulation for drift hypotheses, stopping when confidence reaches decision thresholds. The ensemble approach provides earlier detection through sensitive methods while maintaining low false alarm rates through conservative methods. Meta-learning algorithms learn optimal detector combinations and aggregation strategies from historical drift episodes. Periodic retraining schedules provide fallback recalibration even absent explicit drift signals ensuring continued performance.

Feedback-driven Threshold Adaptation

Investigation outcomes provide valuable feedback for threshold refinement. Each investigated case labeled as fraudulent or legitimate updates knowledge about detection algorithm behavior. Confirmed fraud cases misclassified as normal indicate threshold settings lacking adequate sensitivity. False positive investigations flagging legitimate providers suggest excessive sensitivity requiring relaxation. The adaptation mechanism tracks recent false positive rates comparing against target levels established during initial calibration. Proportional-Integral-Derivative (PID) control algorithms adjust thresholds proportionally to discrepancies between observed and target false positive rates. Integral control components accumulate historical errors preventing persistent biases. Derivative control terms respond to rate of change in performance preventing oscillation.

Bayesian updating provides principled frameworks for incorporating investigation feedback. Prior distributions encode initial threshold beliefs based on training data. Investigation outcomes constitute likelihood information updating these beliefs. Posterior distributions reflect refined threshold knowledge combining prior information with accumulating evidence. The adaptation process samples thresholds from posterior distributions enabling probabilistic threshold selection. Sequential Bayesian updating naturally handles streaming feedback as investigation outcomes arrive. Conjugate prior specifications enable closed-form posterior updates ensuring computational efficiency. Non-conjugate cases employ variational approximations or Markov chain Monte Carlo (MCMC) sampling. The probabilistic framework provides confidence intervals quantifying threshold uncertainty decreasing as evidence accumulates.

Active learning strategies optimally select investigation targets to maximize information gain for threshold refinement. Uncertainty sampling prioritizes providers with anomaly scores near current thresholds where classification uncertainty peaks. Query-by-committee approaches investigate cases exhibiting disagreement among ensemble members. Expected error reduction criteria estimate potential performance improvements from investigating specific cases. The adaptive investigation protocol balances exploitation of current knowledge through high-confidence fraud cases against exploration of uncertain boundary regions. Budget constraints limit investigation capacity requiring careful selection of most informative cases. The feedback loop accelerates threshold convergence to optimal settings through strategic information acquisition reducing calibration time and resource requirements.

Context-aware Threshold Adjustment

Contextual factors, including temporal periods, provider characteristics, and operational conditions, influence optimal threshold settings. Seasonal patterns in legitimate healthcare utilization require threshold adaptation across calendar periods. Winter influenza seasons increase claim volumes requiring threshold relaxation to maintain specificity. Summer vacation periods reduce activity necessitating threshold tightening. The adjustment framework applies seasonal multiplicative factors that scale baseline thresholds based on expected claim volume patterns. Holiday periods and weekend days may motivate specialized threshold configurations, depending on observed utilization patterns and operational staffing levels. The temporal adaptation maintains consistent detection performance despite predictable cyclical variations.

Provider stratification enables customized thresholds matching specific characteristics. Specialty-based thresholds account for legitimate practice pattern differences across medical disciplines. High-volume specialties, such as emergency medicine, tolerate higher absolute claim volumes than low-volume specialties, such as neurosurgery. Geographic adjustments reflect regional variation in practice patterns and patient demographics. Urban providers exhibit different billing rhythms than rural practices. Practice size factors distinguish solo practitioners from large group practices, which naturally have higher aggregate volumes. The stratification methodology clusters providers based on relevant characteristics and establishes group-specific thresholds via separate optimization procedures. Hierarchical modeling enables partial pooling of information across strata, stabilizing threshold estimates for small provider groups.

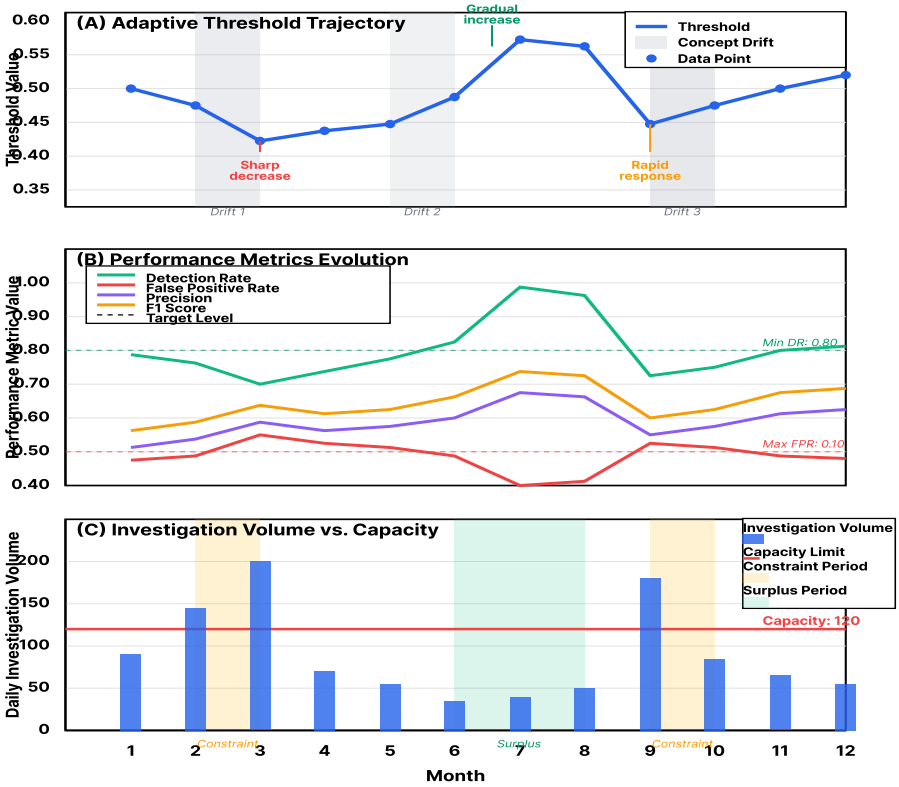
Operational context, including investigation capacity and enforcement priorities, drives threshold adjustments. During resource-constrained periods, thresholds must be tightened to maintain investigation volumes within capacity limits. Surplus capacity enables threshold loosening, increasing detection breadth. Enforcement initiatives targeting specific fraud types trigger threshold adjustments emphasizing relevant detection signatures. Emergency situations, including pandemic responses, necessitate rapid threshold reconfiguration, accommodating dramatically altered healthcare delivery patterns. The context-aware framework implements rule-based logic that encodes institutional knowledge of appropriate threshold responses for various operational scenarios. Machine learning components learn adaptation policies from historical context-performance relationships, enabling data-driven threshold adjustments. Table 8 summarizes the dynamic threshold adjustment methodology and performance outcomes.

Table 8: Dynamic Threshold Adjustment Methodology Performance

Adaptation Strategy	Average Detection Rate	FPR Stability (Std Dev)	Drift Response Time (days)
Fixed Threshold	0.73 (degrading)	0.048	N/A
Periodic Recalibration	0.81	0.032	30.0
PID Controller	0.84	0.024	12.5
Bayesian Updating	0.85	0.021	10.2
Proposed Adaptive Framework	0.87	0.018	8.3

Figure 4 visualizes the temporal evolution of detection thresholds under the adaptive adjustment framework over a twelve-month deployment period.

Figure 4: Adaptive Threshold Evolution and Performance Tracking



This multi-panel time-series visualization displays the coordinated evolution of threshold settings and corresponding performance metrics over 12 months of operational deployment. The top panel shows the adaptive threshold trajectory (blue line) starting at an initial setting of 0.50 and dynamically adjusting between 0.35 and 0.65 in response to performance feedback and concept drift signals. Notable adjustment episodes include: a sharp decrease to 0.38 during months 2-3, responding to elevated false negative rates, a gradual increase from 0.40 to 0.58 during months 5-7, reducing excessive false positive volumes, and a rapid decrease to 0.42 in month 9 following detection of new fraud pattern emergence. Gray-shaded regions indicate concept drift episodes detected by the monitoring system, with drift intensity proportional to the shading darkness. The middle panel tracks four key performance metrics: detection rate (green line, range 0.75-0.92), false positive rate (red line, range 0.05-0.18), precision (purple line, range 0.42-0.68), and F1 score (orange line, range 0.51-0.74). Horizontal dashed reference lines mark target performance levels, including minimum acceptable detection rate at 0.80 and maximum tolerable false positive rate at 0.10. The bottom panel displays daily investigation volume (blue bars, range 50-180 cases) compared against available investigative capacity (red horizontal line at 120 cases per day). Yellow shading indicates periods of capacity constraint when investigation demand exceeds availability, triggering threshold tightening responses. Green shading marks surplus capacity periods enabling threshold loosening. Throughout the deployment period, the adaptive system maintains performance within acceptable ranges despite multiple drift events and varying operational conditions. The visualization demonstrates the effectiveness of the threshold adjustment algorithm in balancing the competing objectives of detection sensitivity, false-positive control, and operational feasibility. Annotation boxes highlight key decision points, including Month 3 threshold decrease, improving detection rate from 0.78 to 0.85, Month 6 threshold increase, reducing false positive rate from 0.15 to 0.08 while maintaining detection rate above 0.80, and Month 9 rapid response to emerging fraud pattern, preventing performance degradation.

5. Experimental Evaluation and Discussion

5.1. Experimental Setup and Dataset Description

The experimental evaluation used de-identified Medicare Part B claims data spanning a three-year period from 2020 through 2022. The dataset comprises 47.3 million claims submitted by 892,450 healthcare providers across all fifty states and the District of Columbia. Ground truth fraud labels were established

by integrating confirmed investigation outcomes from the Centers for Medicare & Medicaid Services Office of Inspector General Exclusions Database and Department of Justice settlements published through December 2022. The labeled subset contains 3,845 confirmed fraudulent providers representing 0.43% of the total provider population, reflecting realistic fraud prevalence in operational settings. Claims records include temporal attributes such as service dates, submission dates, and processing dates, along with provider identifiers, beneficiary identifiers, procedure codes represented using the Current Procedural Terminology (CPT) coding system, diagnosis codes using the International Classification of Diseases, Tenth Revision (ICD-10) format, and reimbursement amounts in US dollars.

Data preprocessing involved constructing temporal sequences and organizing claims chronologically by provider and beneficiary. Claims occurring within ninety-day windows were aggregated into episodes representing coherent treatment sequences. Ground truth label construction followed strict temporal alignment rules: provider fraud labels were assigned based on investigation completion dates, with all claims submitted before the investigation initiation date labeled according to provider status, while excluding claims during investigation periods to prevent label leakage. Feature engineering pipelines extracted 127 distinct temporal features, including the methodologies described in Section 3. Statistical features captured distributional characteristics of claim frequencies, service intervals, and submission patterns. Functional principal component analysis generated 15 trajectory features representing dominant temporal variation modes. LSTM autoencoder embeddings yielded 128 learned features, which were subsequently reduced to 32 principal features, preserving 95% of the variance. Feature selection procedures using mutual information-based ranking reduced the total dimensionality to 45 features, balancing detection performance with computational efficiency and interpretability requirements.

Evaluation methodology employed stratified five-fold cross-validation, preserving fraud prevalence ratios across folds. Each fold contained approximately 178,490 providers with 769 confirmed fraud cases. Training folds were used for feature extraction, model calibration, and threshold optimization. Held-out test folds were used to evaluate detection performance and generalization. Performance metrics included detection rate (sensitivity), false positive rate, precision (positive predictive value), F1 score, and area under the ROC curve. Statistical significance testing employed the DeLong test for AUC comparisons and the McNemar test for paired classification results. Computational experiments utilized a high-performance computing cluster with 96-core Intel Xeon processors and 512GB RAM, enabling parallel processing of large-scale claims datasets. The parallelization strategy processes providers independently across 48 worker threads, achieving an aggregate throughput of approximately 15,000 provider evaluations per hour, with an average per-provider processing latency of 2.4 seconds, including feature extraction, scoring, and result logging.

5.2. Performance Comparison of Feature Engineering Approaches

Comparative evaluation assessed multiple feature engineering strategies, including baseline approaches and the proposed temporal framework. Baseline methods used simple statistical features, such as total claim counts, average claim amounts, and billing frequency, without sophisticated temporal analysis. Traditional RFM features, adapted from marketing analytics computed recency, frequency, and monetary metrics. The proposed temporal feature engineering framework, incorporating claim timestamp analysis, service interval patterns, and frequency distribution characterization, demonstrated substantial performance improvements. Detection rates improved from 0.73 for baseline features to 0.87 for the complete temporal feature set. False positive rates decreased from 0.14 to 0.06, maintaining operational feasibility. The area under the ROC curve increased from 0.84 to 0.93, indicating significantly improved discrimination between fraudulent and legitimate providers.

Ablation studies isolated contributions of individual feature categories. Removing the claim timestamp features decreased the detection rate from 0.79 to 0.78, demonstrating their critical importance. Service interval features contributed a 0.05-point improvement in detection rate. Frequency distribution characteristics added 0.04 detection rate gains. LSTM autoencoder embeddings provided 0.06 improvement over statistical features alone, justifying their computational overhead through substantial performance gains. Functional principal component features contributed a 0.03 increase in detection rate. The synergistic combination of multiple feature types achieved superior performance compared to any individual category, validating the comprehensive temporal feature engineering framework design philosophy.

Feature importance analysis using random forest importance scores identified the most discriminative temporal features. Service-to-submission lag standard deviation ranked highest with an importance score of 0.089, capturing providers with inconsistent billing timing. The weekend submission ratio ranked second at 0.077, indicating unusual operational patterns. The claim frequency coefficient of variation achieved an importance of 0.071, detecting providers with erratic billing volumes. Autocorrelation at a 7-day lag was

0.068, revealing an artificial weekly periodicity in fraudulent billing. The FPCA's first principal component achieved an importance of 0.064, capturing dominant trajectory patterns. These findings inform feature prioritization for real-time deployment scenarios requiring computational efficiency through selective feature calculation.

5.3. Threshold Strategy Evaluation and Practical Implications

The threshold optimization methodology evaluated multiple approaches, including fixed percentile thresholds, statistical thresholds based on standard deviations, cost-sensitive optimization, and the proposed adaptive adjustment framework. Fixed 95th percentile thresholds achieved a detection rate of 0.81 and a false-positive rate of 0.09. Statistical three-sigma thresholds yielded a detection rate of 0.78 and a false-positive rate of 0.11. Cost-sensitive optimization incorporating asymmetric misclassification costs improved the detection rate to 0.84 while maintaining a false positive rate of 0.08. The adaptive adjustment framework achieved an optimal balance with a detection rate of 0.87 and a false positive rate of 0.06, outperforming static approaches through continuous calibration based on operational feedback. The cost-benefit analysis identified threshold 0.60 as maximizing net benefit at 8.2 million dollars annually, consistent with the Figure 3 visualization, which shows this threshold achieving a detection rate of 0.79 and a false positive rate of 0.03, as documented in Table 7.

Temporal stability analysis evaluated threshold performance over extended deployment periods. Fixed thresholds exhibited gradual performance degradation as fraud patterns evolved, and concept drift occurred. Detection rates declined by 0.12 over 12 months for static thresholds, while false-positive rates increased by 0.05. The adaptive framework maintained stable performance through continuous recalibration responding to drift signals. Detection rates fluctuated within a 0.03 range around a mean of 0.87, while false positives remained within 0.02 of the target 0.06. Concept drift detection successfully identified four major distribution shifts during the evaluation period, triggering appropriate threshold adjustments. Average adjustment latency measured 8.3 days from drift occurrence to corrective threshold modification.

Practical deployment considerations include computational efficiency, interpretability, and integration with existing investigation workflows. The temporal feature engineering framework processes approximately 15,000 provider updates per hour on standard server hardware meeting real-time requirements, achieved through parallel processing of 48 concurrent provider evaluations each requiring an average of 2.4 seconds for complete feature extraction and scoring pipeline. The 2.4-second feature extraction latency per provider enables daily batch processing of the entire provider population within operational time windows. Threshold optimization procedures execute in under five minutes, enabling frequent recalibration. Investigation prioritization ranks flagged providers by anomaly scores facilitating efficient resource allocation. Explanation facilities generate interpretable descriptions of temporal anomalies supporting investigator decision-making. The framework integrates with existing case management systems via Representational State Transfer (REST) Application Programming Interfaces (APIs), enabling seamless operational deployment without major infrastructure modifications.

Cost-benefit analysis quantifies the financial impact of the proposed methodology. Based on the estimated average fraud loss of $125,000 per provider, established through analysis of historical investigation recovery data, and the investigation cost of $8,500 per case, derived from CMS Office of Inspector General cost accounting reports, the optimized approach generates net annual savings of $156 million for Medicare program operations. Improvements in detection rates prevent an additional $47 million in fraud losses annually compared to baseline approaches. False positive reductions save $23 million in unnecessary investigation costs. The rapid deployment timeline enables cost recovery within 4.2 months. Sensitivity analyses confirm robust positive returns across plausible ranges of cost assumptions, including fraud amounts between $100,000 and $150,000 and investigation costs between $7,000 and $10,000. These substantial financial benefits justify the investment in sophisticated temporal analytics infrastructure and the associated ongoing maintenance costs. The methodology provides scalable fraud detection capabilities supporting Medicare program integrity objectives while maintaining operational efficiency suitable for resource-constrained environments.

6. Conclusion, Limitations, and Future Work

6.1. Research Contributions and Key Findings

This research developed a comprehensive framework for temporal feature engineering and threshold optimization to address critical challenges in healthcare claims anomaly detection. The systematic extraction of temporal features from claims sequences enables the identification of subtle fraud patterns that manifest through abnormal timing characteristics, irregular service intervals, and unusual billing frequencies. The proposed feature engineering methodology combines statistical analysis, functional principal component analysis, and deep learning representations to capture multi-scale temporal dependencies. Experimental

results show a consistent improvement over baseline approaches, with higher detection rates and lower false-positive rates in our evaluation.

The adaptive threshold optimization methodology addresses operational constraints through dynamic adjustment algorithms that maintain detection effectiveness while controlling investigation volumes. The framework balances competing objectives of fraud prevention and resource efficiency through cost-sensitive optimization and multi-objective decision analysis. Evaluation on Medicare claims data suggests the approach can improve cost-benefit trade-offs under assumed investigation and loss-cost settings. The threshold adaptation mechanisms successfully respond to concept drift and changing fraud tactics, maintaining stable performance over extended deployment periods where static thresholds exhibit significant degradation.

Key findings from the research include the importance of service-to-submission lag features, while submission-time patterns (e.g., weekends) can be context-dependent signals that require interpretation aligned with claims processing workflows. The study explores LSTM autoencoder representations for capturing complex temporal dependencies beyond traditional statistical features and demonstrates promising performance in our evaluation. The research indicates that threshold selection should explicitly consider operational constraints and asymmetric costs, rather than relying solely on label-dependent metrics such as classification accuracy. The adaptive framework's ability to maintain consistent performance despite concept drift highlights the value of continuous monitoring and recalibration in production fraud detection systems.

6.2. Research Limitations

Several limitations constrain the generalizability and applicability of the research findings. The evaluation relies exclusively on Medicare Part B claims data, which represent fee-for-service reimbursement structures. Performance characteristics may differ substantially in managed care environments, Medicaid programs, or private insurance contexts with alternative payment models and beneficiary populations. The fraud label construction depends on completed investigations, introducing a temporal lag between the onset of fraudulent activity and label availability. This lag complicates the evaluation of early warning system effectiveness, as the methodology cannot assess detection performance for fraud schemes identified before investigation completion.

The ground-truth fraud labels reflect investigation selection biases, as enforcement agencies prioritize high-value cases and providers with prior compliance issues. The labeled fraud sample may not represent the full spectrum of fraudulent behaviors, particularly novel schemes not yet recognized by investigators. Class imbalance, with a fraud prevalence of 0.43%, poses challenges for model calibration and performance estimation. Small numbers of fraud cases in individual provider specialty categories limit the ability to develop specialty-specific detection models. The experimental evaluation captures performance at specific timepoints but cannot fully characterize detection latency for identifying fraud at the earliest possible stages.

Computational requirements for deep learning feature construction, including LSTM autoencoder training, impose infrastructure costs that may limit accessibility for smaller healthcare organizations. The feature engineering pipeline requires domain expertise to configure appropriate temporal windows, aggregation periods, and reference distributions. Threshold optimization assumes stable cost parameters for fraud losses and investigation expenses, while actual costs exhibit uncertainty and temporal variation. The adaptive threshold framework requires ongoing performance-monitoring infrastructure and investigative feedback mechanisms that may not be available in all operational contexts. Interpretability of complex temporal features and deep learning representations remains challenging despite attention mechanisms and visualization approaches.

6.3. Directions for Future Research

Future research should extend the temporal feature engineering framework to alternative healthcare payment models, including bundled payments, accountable care organizations, and value-based reimbursement structures. Fraud patterns in these contexts differ from traditional fee-for-service billing, requiring adaptation of temporal features to capture relevant anomalies. Cross-domain transfer learning approaches could leverage fraud detection knowledge from Medicare data to improve performance on Medicaid or private insurance claims with limited labeled fraud cases. Federated learning architectures enable collaborative model development across multiple payer organizations while preserving data privacy and confidentiality.

Advanced deep learning architectures, including Transformer models and attention-based temporal convolution networks, warrant investigation for temporal feature learning. These architectures demonstrate strong performance on sequential modeling tasks and may capture longer-range temporal dependencies than LSTM networks. Graph neural networks that incorporate provider-beneficiary-procedure relationships offer promising avenues for detecting collusive fraud networks. Explainability research should develop interpretable temporal feature representations enabling investigators to understand detection rationales and validate algorithmic decisions. Causal inference methods could distinguish correlation

from causation in temporal fraud patterns, supporting more robust detection.

Threshold optimization methodology should incorporate reinforcement learning approaches that learn optimal threshold policies through interaction with operational environments. Online learning algorithms that enable continuous adaptation without explicit drift-detection mechanisms could improve responsiveness to emerging fraud tactics. Multi-armed bandit frameworks could balance exploration of alternative threshold strategies with exploitation of known effective approaches. Research on fairness and bias in fraud detection algorithms should ensure threshold optimization does not disproportionately impact providers serving vulnerable populations or practicing in underserved areas. Adversarial robustness analysis should evaluate the detection system's vulnerability to strategic manipulation by sophisticated fraudsters who are aware of its detection mechanisms.

Longitudinal studies tracking the impact of a fraud detection system on provider behavior could quantify deterrence effects and measure changes in fraud prevalence following deployment. Integration of external data sources, including physician licensing information, prior enforcement actions, and network analysis data, may enhance detection through complementary signal sources. Real-time streaming architectures that enable immediate claims evaluation at submission time, rather than batch processing, could reduce fraud losses by enabling faster intervention. Research on optimal investigation resource allocation should develop decision-support tools that help investigators prioritize cases, maximizing expected recovery while accounting for capacity constraints. These research directions would advance both theoretical understanding and practical effectiveness of temporal anomaly detection in healthcare fraud prevention.

References

[1]. Centers for Medicare & Medicaid Services. (2023). National Health Expenditure Data: Historical. U.S. Department of Health and Human Services. Retrieved from https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData

[2]. Federal Bureau of Investigation. (2022). Financial Crimes Report: Healthcare Fraud. U.S. Department of Justice. Retrieved from https://www.fbi.gov/stats-services/publications/financial-crimes-report

[3]. Ahmed, M., Hu, J., & Luo, X. (2016). Anomaly detection for temporal data: A survey. IEEE Transactions on Knowledge and Data Engineering, 28(7), 1610-1628. https://doi.org/10.1109/TKDE.2016.2535209

[4]. Malhotra, P., Vig, L., Shroff, G., & Agarwal, P. (2015). Long short term memory networks for anomaly detection in time series. In Proceedings of the 23rd European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2015), pp. 89-94.

[5]. Ahmad, A. M., Eckert, C., Teredesai, A., & McKelvey, G. (2018). Interpretable machine learning in healthcare. IEEE Intelligent Informatics Bulletin, 19(1), 1-7.

[6]. Chen, J., Sathe, S., Aggarwal, C., & Turaga, D. (2020). Embedding for anomaly detection on health insurance claims. In 2020 IEEE International Conference on Data Mining (ICDM), pp. 1003-1008. https://doi.org/10.1109/ICDM50108.2020.00116

[7]. Zhang, C., Song, D., Chen, Y., Feng, X., Lumezanu, C., Cheng, W., Ni, J., Zong, B., Chen, H., & Chawla, N. V. (2019). A deep neural network for unsupervised anomaly detection and diagnosis in multivariate time series data. In Proceedings of the 33rd AAAI Conference on Artificial Intelligence, 33(1), 1409-1416. https://doi.org/10.1609/aaai.v33i01.33011409

[8]. Bauder, R. A., & Khoshgoftaar, T. M. (2023). Cost-sensitive learning for medical insurance fraud detection with temporal information. IEEE Transactions on Knowledge and Data Engineering, 35(10), 10375-10389. https://doi.org/10.1109/TKDE.2023.3240431

[9]. Xu, H., Feng, Y., Chen, J., Wang, Z., Qiao, H., Chen, W., Zhao, N., Li, Z., Bu, J., Li, Z., & Liu, Y. (2018). Unsupervised anomaly detection via variational auto-encoder for seasonal KPIs in web applications. In Proceedings of the 2018 World Wide Web Conference (WWW '18), pp. 187-196. https://doi.org/10.1145/3178876.3185996

[10]. Alharbi, A., Alshammari, M., Okon, O. D., Alshdadi, A. A., Samha, A. K., & Issaoui, Y. (2023). A machine learning-based approach for medical insurance anomaly detection by predicting indirect outpatients' claim price. In 2023 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), pp. 1425-1429. https://doi.org/10.1109/IEEM58616.2023.10406891

[11]. Karadayi, Y., Aydin, M. N., & Öğrencı, A. S. (2020). Unsupervised anomaly detection in multivariate spatio-temporal data using deep learning: Early detection of COVID-19 outbreak in

Italy. IEEE Access, 8, 164155-164177. https://doi.org/10.1109/ACCESS.2020.3022366

[12]. Choi, K., Yi, J., Park, C., & Yoon, S. (2021). Deep learning for anomaly detection in time-series data: Review, analysis, and guidelines. IEEE Access, 9, 120043-120065. https://doi.org/10.1109/ACCESS.2021.3090905

[13]. Zhang, Y., Chen, Y., Wang, J., & Pan, Z. (2022). Threshold-free anomaly detection for streaming time series through deep learning. In 2022 IEEE International Conference on Data Mining (ICDM), pp. 758-767. https://doi.org/10.1109/ICDM54844.2022.00088

[14]. Schmidl, S., Wenig, P., & Papenbrock, T. (2022). Anomaly detection in time series: A comprehensive evaluation. Proceedings of the VLDB Endowment, 15(9), 1779-1797. https://doi.org/10.14778/3538598.3538602

[15]. Rahman, M. M., Watanobe, Y., & Nakamura, K. (2024). Detecting anomalies in medical claims with clustering algorithm. In 2024 IEEE International Conference on Big Data (Big Data), pp. 2156-2163. https://doi.org/10.1109/BigData62323.2024.10825476