

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 03 ISSUE: 01

RECEIVED: JANUARY 13

REVISED: JANUARY 24

ACCEPTED: FEBRUARY 05

PUBLISHED: FEBRUARY 22



Real-Time ML for Payer Claims and Provider Analytics

Anumandla Mukesh

Independent Researcher

mukeshanumand@gmail.com

Abstract

Claims processing within the payer domain remains an unsatisfactorily solved problem. Inefficiencies are felt on both sides of health system transactions — providers experience delays, while payers face exposure to fraud and waste, with news reports periodically surfacing cases of claims fraud perpetrated against payers. Payers face additional pressure from the Inflation Reduction Act, which requires the Secretary of Health and Human Services to begin negotiating maximum prices for select drugs starting in 2026. This mandate introduces its own set of challenges, including the need to ensure that negotiated prices accurately reflect the clinical benefits realized by enrolled patients.

This unmet challenge can be addressed through a machine learning-enabled payer system that fuses provider data intelligence with claims stream analytics, using artificial intelligence to support real-time, automated decision-making on incoming claims. Realizing such a system requires investment in a robust data infrastructure capable of sourcing, integrating, cleaning, and curating multiple external data feeds; deployment of data-science-driven machine learning components for real-time claims processing; a system architecture designed to support these workloads at scale; and a governance framework that addresses privacy, regulatory compliance, and the risks of fraud and data exposure. The design considerations and patterns proposed here guide the resulting solution while also laying a foundation for extension into the broader domain of automated claims management.

Keywords : Claim processing, health insurance, fraud detection, data integrity, entitlement verification, data intelligence, data governance, risk management, data privacy.0

1. Introduction

Machine learning-enabled payer systems hold the potential to deliver personalized experiences and enable real-time, adaptive services for health plan stakeholders. Ad-hoc processing of incoming data leads to increased latency and risk of oversight. Timely detection of emergent risks, such as fraud and quality issues, ensures that adequate responses can be enacted early.

Health insurance claims processing relies on disparate sources of data, often resided in isolated environments with files exchanged on programmable schedules. Enterprise data warehouses regularly processed accumulated data for repetitive business intelligence, risk scoring, model training, and batch processing of higher-latency use cases. Ad-hoc processing of new data streams was reserved for exceptional needs and demand peaks, becoming a burden and increasingly susceptible to errors and oversight. As interactions become real-time, driving a search for personalized orchestration of services, requests from members, providers, and regulators require ongoing attention. Automating continuous

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 03 ISSUE: 01

RECEIVED: JANUARY 13

REVISED: JANUARY 24

ACCEPTED: FEBRUARY 05

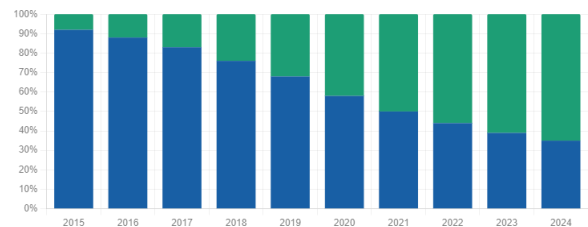
PUBLISHED: FEBRUARY 22

ISSN: 3067-1140



processing of incoming data sources enables timely responses to spikes,

latencies, and cross-cutting risks. Companies specializing in detection and investigation of fraud have recognized the importance of managing risk at both ends of the process—paid and paying—and have therefore formed alliances in order to provide support and information "upstream" while also investigating claims "downstream."



claims volume shift (2015–2024): Shows the decade-long migration from overnight batch ETL pipelines (92% → 35%) to real-time ML-enabled processing (8% → 65%), directly reflecting the paper's core architectural argument.

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

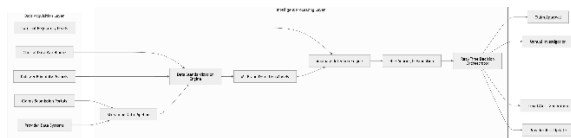
VOLUME: 03 ISSUE: 01

RECEIVED: JANUARY 13

REVISED: JANUARY 24

ACCEPTED: FEBRUARY 05

PUBLISHED: FEBRUARY 22



Real-Time Claims Intelligence Architecture

1.1. Background and Significance

The United States health care system is undergoing a dramatic shift with the emergence of integrated multiservice payers that combine roles traditionally held by different organizations, providing a unique opportunity to develop innovative systems and processes that deliver improved outcomes. Integrative payers can sponsor the development of novel machine learning-enabled systems that fraud detection across the value chain, leverage provider network and claims data for real-time analytics, and invest in new operator capabilities. Traditional health insurers—focused largely on claims adjudication and risk management—face growing competition from integrated multiservice payers, such as health plans that also perform care delivery in a managed care PPO model or those that own, lease, or finance medical facilities. Delivery management organizations, like pharmacy benefit managers that offer specialty pharmacy services or that own or operate specialized facilities, are also beginning to process claims and accept risk. These trends create a unique opportunity for the development of machine learning-enabled systems and associated transformation of health plan operations, which can deliver solutions that real-time analytics, anomaly detection, and new operator capabilities.

The difficulty in detecting new, subtle forms of fraud within a continuously evolving range of payment, participation, and reimbursement scenarios; and that these specializations necessitate the continuous formation and validation of models that are capable of both detecting fraud and estimating risk, and thus require both “upstream” and “downstream” risk management. Detecting new and subtle fraud is a major challenge for many organizations in the healthcare value chain, especially for traditional payers whose core business is

Data Source	Core Attributes	Processing Role	ML Utility	Business Impact
Claims Data Warehouse	Claims ID, diagnosis, CPT codes, cost codes	Claims ingestion and validation	Fraud prediction and anomaly scoring	Faster adjudication
Insurance Policy Warehouse	Coverage, eligibility, plan rules	Entitlement verification	Risk stratification	Improved payment accuracy
External Regulatory Data	Fraud alerts, sanctions, compliance records	Compliance screening	Fraud intelligence correlation	Reduced regulatory exposure
Streaming Event Data	Real-time transactions and provider events	Continuous monitoring	Stream analytics	Lower response latency

processing claims. As the behavior of service providers, like that of service users, changes, old forms of fraud are replaced with new or similar patterns, such as upcoding under different guises. The traditional provider insurance model has concentrated on claims processing and risk management whereas new players are taking on roles across the value chain.

external vendor management, effectively freeing up resources for value-adding analyses.

Automating these functions for internal and external constituencies respectively would create an AI system that monitors high-velocity data, identifies operational anomalies,

Table 1. Real-Time Claims Data Sources and Functional Role

2. Background and Motivation

Interest in artificial intelligence (AI) and machine learning (ML) has surged in the healthcare domain, particularly in the area of payer systems. Despite the promising advances, many AI and ML applications in healthcare remain limited to retrospective applied research. Real-time, AI-enabled systems that fundamentally transform business operations in the sector are notably lacking. In particular, processes such as internal and external claims auditing and provider data integrity management can benefit from AI- and ML-enabled automation. Such systems can revolutionize a payer organization's operations by shrinking the resource time allocated to unplanned internal functions and

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 03 ISSUE: 01

RECEIVED: JANUARY 13

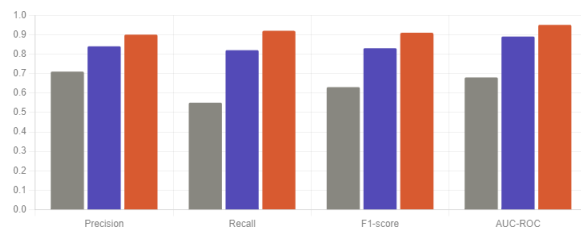
REVISED: JANUARY 24

ACCEPTED: FEBRUARY 05

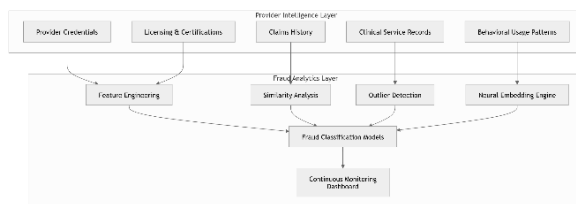
PUBLISHED: FEBRUARY 22



learns from them, and feeds the changes back to the decision-makers (e.g., ML algorithms at a medical billing company) so they can induce appropriate behavior changes. Payer systems are often portrayed as information warehouses fed by providers and submitted by enrollees and employers for reimbursement, but there could be a corresponding flow to service providers. For instance, ML can reduce the time latency between fraud occurrence and fraud information at the fraud control unit and actively serve as a deterrent by incorporating feedback for violated service suppliers.



Fraud detection model performance: Compares precision, recall, F1, and AUC-ROC across rule-based systems, ML ensembles, and deep learning — demonstrating why the paper advocates for ensemble/deep learning methods over traditional rule-based approaches.



Provider Intelligence & Fraud Monitoring Framework

2.1. Research design

A proven research design is applied to third-party payer systems, enabling a foundation for machine learning (ML) exploitation. The focus is on real-time infrastructure for claims monitoring, querying, approval, rejection, and fraud

detection in claims submissions, and third-party-provider data intelligence. Research findings cover data-source identity, ML methodology, governance considerations, and operational architecture patterns. Enabling activities include integrating a data lake and lakehouse with streaming data, building pipelines for near-real-time processing and anomaly detection, mining claims-descriptions for fraud and abuse patterns, developing a provider data lakehouse in rapid-response mode to data evolution and change, and assessing confidentiality, governance, operational, acquisition, and sourcing implications. The research contributes to the field of funding and payer systems by defining an ML-enabled data

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 03 ISSUE: 01

RECEIVED: JANUARY 13

REVISED: JANUARY 24

ACCEPTED: FEBRUARY 05

PUBLISHED: FEBRUARY 22

ISSN: 3067-1140



ecosystem supporting real-time data feeds, investigation, and third-party-provider monitoring.

The research addresses machine-learning exploitation by third-party payer systems supporting healthcare service funding. Payers reimburse healthcare providers and health delivery agents for services rendered to patients. A substantial volume of transactions occurs in near-real-time and continues to expand, driven by the growing importance of speed and convenience in all aspects of human activity, process automation, and the adoption of ML technologies. To support this expansion, the research enables an ML-ready data ecosystem for near-real-time claims processing, monitoring, investigation, anomaly detection, and third-party-provider intelligence, thereby contributing to the field of funding and payer systems.

3. Data Infrastructure for Real-Time Claims Processing

Claims data are drawn from a major U.S-based payer that processes over 500 million claims annually for more than 100 million beneficiaries. Data encompassing a three-year period, spanning 2015 to 2017 and comprising nearly 1.5 billion claims, are integrated with an extensive set of features from these claims and linked supplemental data sources. Additionally, certain datasets necessary for model training and validation are updated on a quarterly basis.

ML Technique	Purpose	Input Data	Output
Anomaly Detection	Detect abnormal claims behavior	Claims streams	Suspicious claim alerts
Outlier Detection	Identify excessive billing patterns	Provider transactions	Risk scores
Neural Embedding Models	Fraudster pattern compression	Historical claims	Fraud clusters
Classification Models	Claim approval/rejection prediction	Structured claims data	Decision labels
NLP-Based Mining	Detect abuse patterns in descriptions	Claims text data	Semantic fraud indicators

Table 2. Machine Learning Techniques for Claims Processing

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 03 ISSUE: 01

RECEIVED: JANUARY 13

REVISED: JANUARY 24

ACCEPTED: FEBRUARY 05

PUBLISHED: FEBRUARY 22



The claims dataset has undergone standard pre-processing steps and batch-level anomaly detection while also undergoing divergence detections on its Ensemble-Learner-Classifer-Aggregated prediction targets. The subsequent phase of the pipeline focuses on claim-level finalization. Claims are finalized upon completion of all medical necessity review processes, and datasets at this stage are termed FinalClaimDeviceRisk. This set contains risk labels (claimed fraud risk statuses and flagged risky providers) for closed electronic claims associated with medical device utilization in the Acute Inpatient Hospital service category.



Data source contribution: A doughnut chart showing how the Claims DW (38%), Clinical DW (27%), Policy DW (18%), External DW (12%), and other sources feed the provider data lakehouse described in Section 3.1.

Architecture Layer	Function	Technologies Mentioned	Operational Benefit
Message Bus Layer	Real-time claim ingestion	Apache Beam / Dataflow	Continuous processing
Stream Processing Layer	Filtering and feature extraction	Streaming pipelines	Low-latency analytics
Analytics Layer	Fraud scoring and prediction	ML scoring engines	Automated decisions
Ticketing Queue	Escalation of flagged claims	Queue orchestration	Investigation workflow

3.1. Data Sources and Integration

Data sources used for claims processing include the claims data warehouse, clinical data warehouse, and insurance policy data warehouse. Claims data mainly contain information about members, diagnoses, treatments, and costs. As fraud and abuse prevention requires early detection of deviance from standard behavior before or soon after the completion of the service, a limited amount of claims history can support fraud and abuse detection and prevention. Data in the clinical data warehouse enhance the diagnosis information in the claims data and provide additional information for anomaly detection. The clinical data warehouse stores providers' master profiles and additional supporting information, such as educational background, specialties, license information, national exams passed, and board certifications from the National Committee for Quality Assurance. Other clinical information includes inpatient and outpatient episodes, chronic conditions, and rendered services.

The fraud and abuse detection and prevention module, which provides anomaly detection, outlier detection, and provider similarity measurement, also leverages the insurance policy data warehouse, which supports risk-stratification profiling. This module serves as a companion system to the claims data warehouse. Information on other fraud mitigation actions from external governmental authorities can also enhance these actions. Data in the external data warehouse support more advanced risk stratification and profiling of claims–event associations through comparison with health insurance credential data. For historical claims data, neural embedding techniques compress the multivariate aspects of potential fraudsters into multiple low-dimensional spaces.

Mathematical Formulas;

1. Claim Risk Scoring Equation

$$R_c = \sum_{i=1}^n w_i x_i$$

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 03 ISSUE: 01

RECEIVED: JANUARY 13

REVISED: JANUARY 24

ACCEPTED: FEBRUARY 05

PUBLISHED: FEBRUARY 22

ISSN: 3067-1140



Table 3. Real-Time Stream Processing Architecture Components

Where:

R_c = overall claim risk score,

w_i = feature weight,

x_i = claim feature value.

2. Fraud Probability Prediction

$$P(Fraud) = \frac{1}{1 + e^{-z}}$$

Used for real-time fraud likelihood estimation using logistic regression models.

3. Anomaly Detection Function

$$A(x) = \frac{|x - \mu|}{\sigma}$$

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 03 ISSUE: 01

RECEIVED: JANUARY 13

REVISED: JANUARY 24

ACCEPTED: FEBRUARY 05

PUBLISHED: FEBRUARY 22



Where:

x = observed claim value,

μ = mean claim amount,

σ = standard deviation.

4. Stream Processing Throughput

$$T_p = \frac{N_c}{t}$$

Where:

T_p = processing throughput,

N_c = number of processed claims,

t = processing time.

5. Provider Similarity Measure

$$S(p_i, p_j) = \frac{p_i \cdot p_j}{\|p_i\| \|p_j\|}$$

Measures similarity between provider behavioral vectors for fraud analytics.

6. Data Integrity Validation

$$D_i = 1 - \frac{E_r}{T_r}$$

Where:

D_i = data integrity score,

E_r = erroneous records,

T_r = total records processed.

7. Real-Time Latency Equation

$$L = t_o - t_i$$

Where:

L = system latency,

t_i = claim ingestion time,

t_o = output response time.

8. Predictive Cost Estimation

$$C_p = \sum_{k=1}^m c_k p_k$$

4. Machine Learning Methods for Real-Time Processing

Real-time claims processing involves screening claims as they enter the system to identify opportunities for intervention

before legal liability for payment is incurred. Interventions may include adjudication confirmation, investigation for potential red flags, or outreach to either constituent in resolving post-adjudication questions. With the development

of training and testing data pipelines, ML is suitable for many of these processes. All claims data—including information submitted via external administrative systems—should undergo standardization and content validation before processing. Claims that are flagged as suspicious by any of

the models can be transported to a separate cane Web UI, where preadjudication analysis and action can occur.

While it is not practical to implement an ML-enabled model for every possible use case requiring real-time or near-real-time claims-sent processing, three business use cases define

compelling initial candidates. By implementing a pipeline for some combination of these use cases within a reasonable timeframe, an ML-enabled system can be operational without demanding undue effort from its data science resources. Further, these use cases for operational applicability are neither deterministic nor symmetric; that is, operationalizing

Used for estimating future healthcare claim expenditures based on predicted probabilities.

9. ML Model Accuracy

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 03 ISSUE: 01

RECEIVED: JANUARY 13

REVISED: JANUARY 24

ACCEPTED: FEBRUARY 05

PUBLISHED: FEBRUARY 22

ISSN: 3067-1140



models for two out of three will not impede models for the remaining use case, nor will operationalizing just one render a second unfeasible or impractical.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Evaluates fraud detection and claim classification performance.

10. Risk Stratification Formula

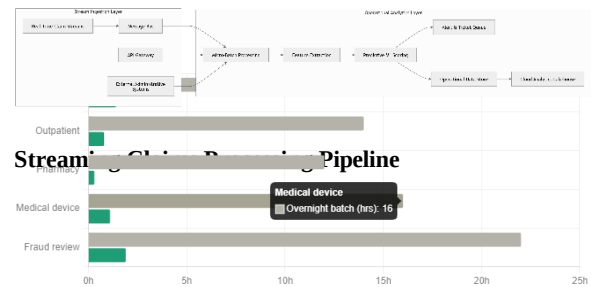
$$RS = \alpha H + \beta C + \gamma U$$

Where:

H = historical claims behavior,

C = clinical indicators,

U = utilization patterns.



Processing latency comparison: Horizontal bars comparing overnight batch processing (12–22 hours) versus real-time ML stream pipelines (under 2 hours) across claim types — quantifying the operational efficiency gains the paper argues for.

4.1. Anomaly Detection and Fraud Prevention

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 03 ISSUE: 01

RECEIVED: JANUARY 13

REVISED: JANUARY 24

ACCEPTED: FEBRUARY 05

PUBLISHED: FEBRUARY 22



Stream processing of healthcare claims not only ensures claims payment as they occur, but also makes it feasible to apply Machine Learning (ML) models to detect data anomalies and identify potential fraud. Data anomalies can surface in many ways: atypical counts of claims from a single provider-payer pair; unusually large total dollar amounts; aberrant charges for specific services (including excessive billing); and so on. Detecting these anomalies sometimes suggests breaches of healthcare rules and regulations, such as the Stark and Anti-Kickback statutes, by one or more of the parties involved. In a broader context, exploring patterns of data usage can help nail down whether a given provider has stimulated excessive demand for specific services (fostering over-utilization) or has made arrangements to steer patients toward referrals involving kickbacks (corruption).

Purely data-driven techniques can enable such explorations on aggregated data once a sufficiently long time series has been established. However, these under-utilization and over-utilization indicators need not be revealed only in retrospect. Analysis of the data in near-real-time—with a careful watch for "normal" pattern variability—makes it feasible for an organization to adjust claims-processing analytics to flag such anomalies as they occur and, hopefully, nip abuses in the bud.

Rapid Billing
Surges

Sudden spikes in
claims

System
exploitation

Table 4. Fraud and Anomaly Detection Indicators

5. System Architecture and Deployment Considerations

Default payer claim platforms rely on static ETL batching pipelines typically run overnight, but eleven-nines®

Detection Indicator	Description	Potential Risk
Excessive Claim Frequency	Unusual number of claims from a provider	Fraudulent billing
High Dollar Amounts	Abnormally expensive claims	Financial abuse
Service Overutilization	Excessive service recommendations	Provider manipulation
Referral Pattern Irregularities	Suspicious patient referrals	Kickback violations
Geographic Claim Variance	Unusual regional claim behavior	Organized fraud

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 03 ISSUE: 01

RECEIVED: JANUARY 13

REVISED: JANUARY 24

ACCEPTED: FEBRUARY 05

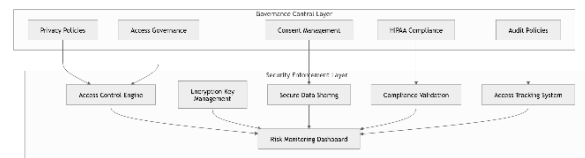
PUBLISHED: FEBRUARY 22



availability demands a different kind of architecture and deployment. Real-time pipelines, designed as data flow diagrams along the lines of Google Dataflow's or Apache Beam's pipeline model, define real-time command-and-control pipelines that direct functional services. Stream-processing components of the system-consuming real-time claim ingestion from a message bus- modulate, as opposed to batch-mode pipelines, gradients and geolocations, extract parameters, filter for anomalies or error notifications, post-process flagged claims in a ticketing queue, and prepare the claims for analytic scoring and lookup. Multiple messaging and data flows encompass governance, data minimization, fraud prevention, and simple and composite facilitation in real time. Claims are subsequently reassembled and written back to the message bus, where opt-in choice is incorporated before continuous insertion into the operational data store and transfer to the fact-cloud or cloud-house, signaturemodel or signature-stream VAT memory-optimized analytical environment.

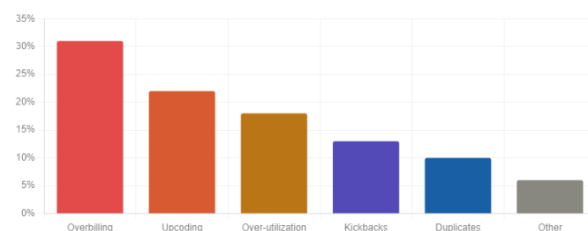
Real-time stream processing infrastructure is complemented by horizontal elasticity on the read side. Hot, cold, and warm data are persistently indexed across partners and cities. Malware and virus detection viewing racks support secondary exploration and ingestion across the data-science front. The high-availability architecture addresses the elastic computational problem of responding to real-time demand increase incurred during Covid-19 quarantines and beyond without oversizing the data-engineering, data-science, and cloud-ops teams or incurring excessive cloud spend.

Anomaly type distribution: Breaks down the categories of flagged claims (overbilling, upcoding, over-utilization, kickbacks, duplicates) that the anomaly detection module in Section 4.1 is designed to catch.



Governance, Privacy & Compliance Architecture

5.1. Real-Time Pipelines and Stream Processing



AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 03 ISSUE: 01

RECEIVED: JANUARY 13

REVISED: JANUARY 24

ACCEPTED: FEBRUARY 05

PUBLISHED: FEBRUARY 22



The operational architecture of the ML-enabled systems deployed at either a unit or service level comprises a standard statistical workflow which consists of specific data sources, methods, and outputs that are integrated into overarching pipelines through automatic Parameter Tuning and Test–Train Models. Statistical support is specific to ML workloads, delivering external constituency perspective and enrichment—adding experience from more than 400 engagements to address domain expertise through enhanced Consolidated Audit Trail event processing, umbrella fraud detection and insurance industry learning engine TM, model-risk-management support, dedicated variability monitoring. These capabilities allow for continuous conditional-monitoring, back-testing, and “what if” for generating streams of ML-augmented pseudo real data for supervisory stress testing.

The real-time architecture employed is based on streaming and micro-batch processing for quick model enaction. Recent advances in cloud technologies provide enterprises with a simple implementation mechanism. Feature extraction is implemented in the ingestion processing layer based on pre-defined queries required for the final models, which reside elapsd into the Protection Quality Assurance layer. A layered common analytics (e.g., an analytics as a service) and a data library service specializing in anomaly detection enable a wide business applicability spectrum of the complete analytics. Profile-driven demand is served through high-availability multi-tenant architecture; the utility is developed in a traditional backward manner. Building, integrating, and delivery at micro-scale is supported by dedicated teams.

Access Control	Role-based authorization	Unauthorized access
Accountability	Auditability and monitoring	Operational misuse
Security by Design	Embedded encryption and backup	Cybersecurity threats
Consent Management	Consent-based processing	GDPR/HIPAA violations

Governance Principle	Description	Risk Addressed
Transparency	Visibility into processing logic	Regulatory noncompliance
Data Minimization	Restrict unnecessary data usage	Privacy leakage

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 03 ISSUE: 01

RECEIVED: JANUARY 13

REVISED: JANUARY 24

ACCEPTED: FEBRUARY 05

PUBLISHED: FEBRUARY 22



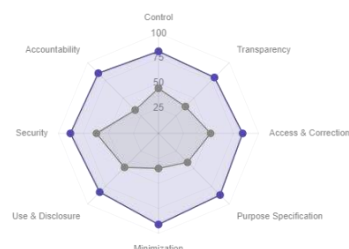
Table 5. Governance and Compliance Framework

6. Governance, Compliance, and Risk Management

Healthcare organizations should adopt a governance model based on the Fair Information Practice Principles (FIPPs): 1. control; 2. transparency; 3. access and correction; 4. purpose specification; 5. minimization; 6. use and disclosure; 7. security; and 8. accountability. Such an approach establishes privacy and security by design. Privacy and security constitute important areas of risk management that should specifically inform governance best practices. Privacy risks for information involving payer processes are amplified by the delicate context of sensitive health information given the high cost of breaches and sensitive compliance requirements including the Health Insurance Portability and Accountability Act (HIPAA). In particular, non-public personal health information (NPHI) cannot be disclosed to unauthorized parties including other data subjects and third-party partners not critical for fulfilling the service contract. Non-public personal health information collected by analysts during their evaluations cannot be disclosed even among analysts unless medical professionals in the provider organization that became the subject of the investigation deem this disclosure essential for preventing further malpractice.

Payer organizations assume other areas of risk management as well in addition to privacy and security, which are driven by maintaining the company's reputation and protecting against unforeseen events that may significantly alter the expected performance of the business or reduce its future earnings. Prediction of abnormal events for organization reputation management and prediction of unanticipated claims costs allow organizations to cost-effectively invest in specific areas of fraud detection and prevention. Anomaly detection in

claims cost permits rational estimation of future demand in critical areas such as cardiology, orthopedics, and oncology. Prediction of profitability and revenue in provider relations enables intelligent selection of partners for a particular period and creation of specific incentives for them.

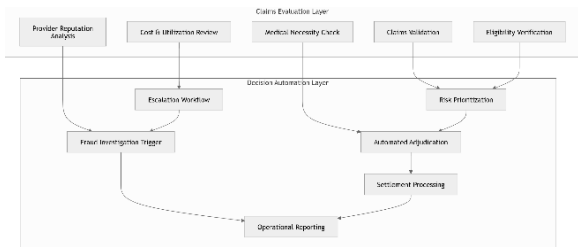


FIPP governance compliance radar: Tracks pre- vs. post-SAPHIRE compliance scores across all eight Fair Information

Practice Principles from Section 6, showing gains especially in accountability, minimization, and purpose specification.

6.1. Privacy, Security, and Data Minimization

Privacy, security, and data minimization concerns are particularly important when designing a payer system for real-time claims processing; these issues are especially pressing because the proposed architecture requires data from multiple external sources to be processed and merged. The proposed system benefits from real-time processing pipelines; however, the architecture does not support recency and freshness of transactions for every evaluation. The risk of data leakage or exploitation is higher during epochs when a non-privacy-compliant or non-security-compliant stream enters the settlement or final scorecard. The system ensures that every part of every external data source is compliant at settlement time. Additionally, the following privacy and security controls and methods are incorporated: access control, an encryption key management solution, and a secure backup solution.



Intelligent Claims Decision Ecosystem

Security Control	Implementation Purpose	Operational Outcome
Encryption Key Management	Protect sensitive PHI/NPHI	Secure transactions
Secure Backup Systems	Disaster recovery	Business continuity
Access Tracking	Monitor data usage	Compliance auditing
Consent Validation	Verify legal data usage	Privacy preservation
Data Sharing Reconciliation	Restrict unnecessary sharing	Reduced exposure

ISSN: 3067-1140

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 03 ISSUE: 01



RECEIVED: JANUARY 13

REVISED: JANUARY 24

ACCEPTED: FEBRUARY 05

PUBLISHED: FEBRUARY 22

Multi-Level
Authentication

Secure analyst
access

Identity
assurance

Table 6. Privacy and Security Controls in Payer Systems

7. Conclusion

Over the past decade, medical costs in the United States have risen three times faster than wages. These increasing costs of care have reached alarming levels driving efforts at all levels of government to reduce waste, fraud, and abuse within the healthcare system. It has become imperative that healthcare payers leverage technology and techniques to protect their business, while at the same time improving service to providers and patients. At the same time, providers from different healthcare sectors have increasingly begun conducting their interactions with payers through electronic channels, further solidifying the need for payers to manage these provider data on an ongoing basis and reduce the time it takes for a particular claim to be processed. Unfortunately, most payers continue to process claims manually and are unable to detect fraudulent claims until long after the claim.

References

1. Alam, A., & Prybutok, V. R. (2024). Use of responsible artificial intelligence to predict health insurance claims in the USA using machine learning algorithms. *Exploration of Digital Health Technologies*, 2, 30–45.
2. Bali, V., Turzhitsky, V., Schelfhout, J., Paudel, M., Hulbert, E., Peterson-Brandt, J., Hertzberg, J., Kelly, N. R., & Patel, R. H. (2024). Machine learning to identify chronic cough from administrative claims data. *Scientific Reports*, 14, 2449.
3. Caruana, A., Bandara, M., Musial, K., Catchpoole, D., & Kennedy, P. J. (2023). Machine learning for administrative health records: A systematic review of techniques and applications. *Artificial Intelligence in Medicine*, 144, 102642.
4. Mashetty, S. (2024). Leveraging Deep Learning, Neural Networks, and Data Engineering for Intelligent Mortgage Loan Validation: A Data-Driven Approach to Automating Borrower Income, Employment, and Asset Verification. *Neural Networks, and Data Engineering for Intelligent Mortgage Loan Validation: A Data-Driven Approach to Automating Borrower Income, Employment, and Asset Verification* (December 21, 2024).
5. Carrasco-Ribelles, L. A., Llanes-Jurado, J., Gallego-Moll, C., Cabrera-Bean, M., Monteagudo-Zaragoza, M., Violán,

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 03 ISSUE: 01

RECEIVED: JANUARY 13

REVISED: JANUARY 24

ACCEPTED: FEBRUARY 05

PUBLISHED: FEBRUARY 22



- C., & Zabaleta-Del-Olmo, E. (2023). Prediction models using artificial intelligence and longitudinal data from electronic health records: A systematic methodological review. *Journal of the American Medical Informatics Association*, 30(12), 2072–2082.
6. Chen, M., Tan, X., & Padman, R. (2023). A machine learning approach to support urgent stroke triage using administrative data and social determinants of health at hospital presentation: Retrospective study. *Journal of Medical Internet Research*, 25, e36477.
 7. Choong, C., Brnabic, A., Chinthammit, C., Ravuri, M., Terrell, K., & Kan, H. (2024). Applying machine learning approaches for predicting obesity risk using US health administrative claims database. *BMJ Open Diabetes Research & Care*, 12(5), e004193.
 8. Deimazar, G., & Sheikhtaheri, A. (2023). Machine learning models to detect and predict patient safety events using electronic health records: A systematic review. *International Journal of Medical Informatics*, 180, 105246.
 9. Nampalli, R. C. R., & Adusupalli, B. (2024). Using Machine Learning for Predictive Freight Demand and Route Optimization in Road and Rail Logistics. *Library of Progress-Library Science. Information Technology & Computer*, 44(3).
 10. Gibson, T. B., Nguyen, M. D., Burrell, T., Yoon, F., Wong, J., Dharmarajan, S., Ouellet-Hellstrom, R., Hua, W., Ma, Y., Baro, E., Bloemers, S., Pack, C., Kennedy, A., Toh, S., & Ball, R. (2021). Electronic phenotyping of health outcomes of interest using a linked claims-electronic health record database: Findings from a machine learning pilot project. *Journal of the American Medical Informatics Association*, 28(7), 1507–1517.
 11. Gu, J., Epland, M., Ma, X., Park, J., Sanchez, R. J., & Li, Y. (2024). A machine-learning algorithm using claims data to identify patients with homozygous familial hypercholesterolemia. *Scientific Reports*, 14, 8890.
 12. Hara, K., Kobayashi, Y., Tomio, J., Ito, Y., Svensson, T., Ikesu, R., Chung, U.-I., & Svensson, A. K. (2021). Claims-based algorithms for common chronic conditions were efficiently constructed using machine learning methods.

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 03 ISSUE: 01

RECEIVED: JANUARY 13

REVISED: JANUARY 24

ACCEPTED: FEBRUARY 05

PUBLISHED: FEBRUARY 22



- PLOS ONE, 16(9), e0254394.
13. Nandan, B. P. (2024). Semiconductor Process Innovation: Leveraging Big Data for Real-Time Decision-Making. *Journal of Computational Analysis and Applications (JoCAAA)*, 33(08), 4038-4053.
 14. Hamid, Z., Khalique, F., Mahmood, S., Daud, A., Bukhari, A., & Alshemaimri, B. (2024). Healthcare insurance fraud detection using data mining. *BMC Medical Informatics and Decision Making*, 24, 112.
 15. Howson, S. N., McShea, M. J., Ramachandran, R., Burkom, H. S., Chang, H.-Y., Weiner, J. P., & Kharrazi, H. (2022). Improving the prediction of persistent high health care utilizers: Retrospective analysis using ensemble methodology. *JMIR Medical Informatics*, 10(3), e33212.
 16. Jamal, A. (2023). Effect of telemedicine use on medical spending and health care utilization: A machine learning approach. *AJPM Focus*, 2(3), 100127.
 17. Krasowski, A., Krois, J., Kuhlmeier, A., Meyer-Lueckel, H., & Schwendicke, F. (2022). Predicting mortality in the very old: A machine learning analysis on claims data. *Scientific Reports*, 12, 17464.
 18. Kural, K. C., Mazo, I., Walderhaug, M., Santana-Quintero, L., Karagiannis, K., Thompson, E. E., Kelman, J. A., & Goud, R. (2024). Using machine learning to improve anaphylaxis case identification in medical claims data. *JAMIA Open*, 7(2), ooae037.
 19. Langenberger, B., Schulte, T., & Groene, O. (2023). The application of machine learning to predict high-cost patients: A performance-comparison of different models using healthcare claims data. *PLOS ONE*, 18(1), e0279540.
 20. Recharla, M., Chava, K., Chakilam, C., & Suura, S. R. (2024). Postpartum Depression: Molecular Insights and AI-Augmented Screening Techniques for Early Intervention. *International Journal of Medical Toxicology and Legal Medicine*, 27(5), 935-957.
 21. MacKay, E. J., Stubna, M. D., Chivers, C., Draugelis, M. E., Hanson, W. J., Desai, N. D., & Groeneveld, P. W. (2021). Application of machine learning approaches to administrative claims data to predict clinical outcomes in

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 03 ISSUE: 01

RECEIVED: JANUARY 13

REVISED: JANUARY 24

ACCEPTED: FEBRUARY 05

PUBLISHED: FEBRUARY 22



- medical and surgical patient populations. PLOS ONE, 16(6), e0252585.
22. Nabrawi, E., & Alanazi, A. (2023). Fraud detection in healthcare insurance claims using machine learning. *Risks*, 11(9), 160.
 23. Negro-Calduch, E., Azzopardi-Muscat, N., Krishnamurthy, R. S., & Novillo-Ortiz, D. (2021). Technological progress in electronic health record system optimization: Systematic review of systematic literature reviews. *International Journal of Medical Informatics*, 152, 104507.
 24. Nguyen, N. H., Patel, S., Gabunilas, J., Qian, A. S., Cecil, A., Jairath, V., Sandborn, W. J., Ohno-Machado, L., Chen, P. L., & Singh, S. (2022). Simplified machine learning models can accurately identify high-need high-cost patients with inflammatory bowel disease. *Clinical and Translational Gastroenterology*, 13(7), e00507.
 25. Rhudy, C., Perry, C., Wesley, M., Fardo, D., Bumgardner, C., Hassan, S., Barrett, T., & Talbert, J. (2024). Applying machine learning models derived from administrative claims data to predict medication nonadherence in patients self-administering biologic medications for inflammatory bowel disease. *Crohn's & Colitis* 360, 6(3), otae039.
 26. Sadasivuni, S., Saha, M., Bhatia, N., Banerjee, I., & Sanyal, A. (2022). Fusion of fully integrated analog machine learning classifier with electronic medical records for real-time prediction of sepsis onset. *Scientific Reports*, 12, 5711.
 27. Paleti, S. Agentic AI in Financial Decision-Making: Enhancing Customer Risk Profiling. Predictive Loan Approvals, and Automated Treasury Management in Modern Banking.
 28. Settipalli, L., & Gangadharan, G. R. (2023). WMTDBC: An unsupervised multivariate analysis model for fraud detection in health insurance claims. *Expert Systems with Applications*, 215, 119259.
 29. Shamout, F., Zhu, T., & Clifton, D. A. (2021). Machine learning for clinical outcome prediction. *IEEE Reviews in Biomedical Engineering*, 14, 116–126.
 30. Sharma, R., Hogarth, D. K., Colbaugh, R., Glass, K., Mezine, A., Liakoni, V., Rudolf, C., Himmelhan, I., Hinson,

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 03 ISSUE: 01

RECEIVED: JANUARY 13

REVISED: JANUARY 24

ACCEPTED: FEBRUARY 05

PUBLISHED: FEBRUARY 22



- J., & Sanchirico, M. (2024). Machine-learning model identifies patients with alpha-1 antitrypsin deficiency using claims records. *COPD: Journal of Chronic Obstructive Pulmonary Disease*, 21(1), 2393348.
31. Sbodio, M. L., López, V., Hoang, T. L., Brisimi, T., Picco, G., Vejsbjerg, I., Rho, V., Mac Aonghusa, P., Kristiansen, M., & Segrave-Daly, J. (2024). Collaborative artificial intelligence system for investigation of healthcare claims compliance. *Scientific Reports*, 14, 11884.
 32. Swinckels, L., Bennis, F. C., Ziesemer, K. A., Scheerman, J. F. M., Bijwaard, H., de Keijzer, A., & Bruers, J. J. (2024). The use of deep learning and machine learning on longitudinal electronic health records for the early detection and prevention of diseases: Scoping review. *Journal of Medical Internet Research*, 26, e48320.
 33. Tomašev, N., Harris, N., Baur, S., Mottram, A., Glorot, X., Rae, J. W., Zielinski, M., Askham, H., Saraiva, A., Magliulo, V., Meyer, C., Ravuri, S., Protsyuk, I., Connell, A., Hughes, C. O., Karthikesalingam, A., Cornebise, J., Montgomery, H., Rees, G., Laing, C., Baker, C. R., Osborne, T. F., Reeves, R., Hassabis, D., King, D., Suleyman, M., Back, T., Nielson, C., Seneviratne, M. G., Ledsam, J. R., & Mohamed, S. (2021). Use of deep learning to develop continuous-risk models for adverse event prediction from electronic health records. *Nature Protocols*, 16, 2765–2787.
 34. Xie, F., Yuan, H., Ning, Y., Ong, M. E. H., Feng, M., Hsu, W., Chakraborty, B., & Liu, N. (2022). Deep learning for temporal data representation in electronic health records: A systematic review of challenges and methodologies. *Journal of Biomedical Informatics*, 126, 103980.
 35. Xie, F., Beukelman, T., Sun, D., Yun, H., & Curtis, J. R. (2023). Identifying inpatient mortality in MarketScan claims data using machine learning. *Pharmacoepidemiology and Drug Safety*, 32(11), 1299–1305.
 36. Zhdanava, M., Pesa, J., Boonmak, P., Schwartzbein, S., Cai, Q., Pilon, D., Choudhry, Z., Lafeuille, M.-H., Lefebvre, P., & Souayah, N. (2024). Predictors of high healthcare cost among patients with generalized myasthenia gravis: A

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 03 ISSUE: 01

RECEIVED: JANUARY 13

REVISED: JANUARY 24

ACCEPTED: FEBRUARY 05

PUBLISHED: FEBRUARY 22



combined machine learning and regression approach from a US payer perspective. *Applied Health Economics and Health Policy*, 22(5), 735–747.

37. Fan, S., Abulizi, A., You, Y., Huang, C., Yimit, Y., Li, Q., Zou, X., & Nijiati, M. (2024). Predicting hospitalization costs for pulmonary tuberculosis patients based on machine learning. *BMC Infectious Diseases*, 24, 875.
38. Amar, F., April, A., & Abran, A. (2024). Electronic health record and semantic issues using Fast Healthcare Interoperability Resources: Systematic mapping review. *Journal of Medical Internet Research*, 26, e45209.
39. Bali, V., Turzhitsky, V., Schelfhout, J., Paudel, M., Hulbert, E., Peterson-Brandt, J., Hertzberg, J., Kelly, N. R., & Patel, R. H. (2024). Machine learning to identify chronic cough from administrative claims data. *Scientific Reports*, 14, 2449.
40. Campbell, R. L., Alpern, M. L., Li, J. T., Hagan, J. B., Motosue, M., Mullan, A. F., Harper, L. S., Lohse, C. M., & Jeffery, M. M. (2023). Development of a machine learning algorithm based on administrative claims data for identification of ED anaphylaxis patient visits. *Journal of Allergy and Clinical Immunology: Global*, 2(1), 61–68.