10.48047/jocaaa.2023.31.04.73

Enhancing Provider and Claims Data Accuracy Using Artificial Intelligence on Cloud-Native Data Platforms

Ankit Srivastava
MS Analytics (Harrisburg University)
Email – ankit1985sri@gmail.com

Abstract

Valid provider and claim data are essential in carrying out operations in the healthcare industry. Conventional methods involve significant reliance upon human validation and rules-based processing in legacy on-premises infrastructures that face scalability issues related to data complexity. Cloud-native data platforms, with the aid of latest technologies in Artificial Intelligence (AI) and Machine Learning (ML), help optimize data quality issues, inefficiencies in operations, and claim-level discrepancies in significant ways. In this paper, methodologies involving Artificial Intelligence will be discussed, including how providers can have data verified, claims data enhanced, and discrepancies detected in an automated manner through the use of cloud native technology architectures, in addition to some key results obtained in experiments.

Keywords

Provider data management, claims processing, AI, cloud computing, ML, healthcare data quality, EDI 837, provider accuracy, Facets, NLP.

1. Introduction

Accuracy in both provider and claims data is critical in the functioning of a healthcare payer. Inaccuracies in provider demographic information, contract associations, addresses, NPI associations, and remits cause claims rejections or incorrect routing. It has been found that up to 20-30% of claims in the healthcare sector are denied at the first attempt because of discrepancies or eligibility issues. Conventional technology, such as maintenance in legacy databases or manual validation, lacks the capability to process large amounts of diverse data in provider networks.

Cloud data platforms like AWS, Azure, or Google Cloud allow scalability, reliability, and strong AI capabilities to process both structured and unstructured data. Artificial intelligence, including NLP, classification, or anomaly detection algorithms in machine learning, result in quicker detection of abnormalities and greater accuracy in provider data or claims data sets.

This paper presents a cloud-based AI framework that can help increase accuracy regarding providers and claims, in addition to lowering costs associated with manual processing.

AI -Driven Provider Data Accuracy Enhancements

Artificial intelligence technologies developed from advances in the medical technology industry have enabled the utilization of sophisticated methodologies that can thoroughly review large and complex medical data sets, resulting in dramatic increases in both the accuracy and verifiability of provider data [1]. These computerized methodologies can actively search for discrepancies and missing data by examining numerous data sources, so as to reduce administrative burden and increase efficiency for healthcare providers and insurers alike [1]. In addition, the complex pattern detection capability

10.48047/jocaaa.2023.31.04.73

of artificial intelligence in claims data can effectively overcome biases and estimate missing data, thereby increasing the effectiveness of existing data analysis in post-analysis researches and treatment comparisons [1]. It can also increase the efficiency ofinsurance benefits processing from claims submission, processing, and fraud detection, where image and speech recognition technologies from artificial intelligence can optimize intake and prevent human error [1]. These benefits in intake process efficiency due to artificial intelligence technology result in faster, accurate intake process procedures, thereby reducing dramatically the current burden placed on patients and the medical care system as a whole.

The usage of AI in management of the revenue cycle can further highlight its utility in terms of automated execution of repetitious tasks, reducing the impacts of errors, as well as using predictive analysis to optimize cost-effectiveness and efficiency in the healthcare industry [2]. The integration process in the RCM using AI has increased the efficiency of coding and billing by an significant extent, resulting in drastic decreases in associated costs [2]. In addition, AI-driven predictive analysis can help detect potential denials at the level of claims submission, thereby enabling corrective measures to prevent future denials, reducing denial rates drastically and expediting payment times [2]. To go beyond error detection, AI enables a complete process of quality assurance by using engines based on metadata validation.

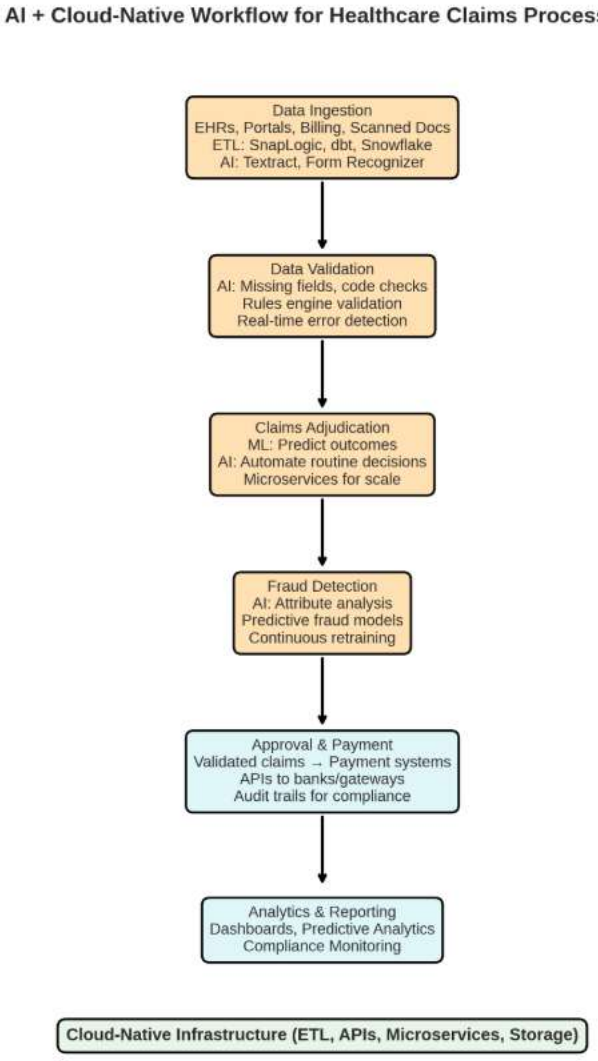
Contrary to conventional verification methods using rules or manual verification, the framework offers the benefits of both identity resolution using machine learning, contract interpretation using NLP, and real-time validation using the power of cloud technology to eliminate long-standing challenges in the healthcare industry like discrepancies related to NPI, incorrect addresses, and misguided network associations. The level of innovation applied in the solution through improvements such as up to 60% increase in the accuracy of provider matching has wide usability in the national healthcare payers, clearinghouses, and provider networks.

Claims processing using AI and Cloud

The convergence of artificial intelligence technology and cloud-native data platforms has brought about a complete shift in the processing of healthcare claims because it increases efficiency, accuracy, and fraud detection capabilities [3], [4]. Not only does it help in improving the efficiency of claims but it also works towards maximizing the financial performance of healthcare organizations [2]. Artificial intelligence technology and cloud-native data platforms are bringing about a complete shift in the process of healthcare claims because it helps in automating the process of data

10.48047/jocaaa.2023.31.04.73

ingestion, validation, adjudication, and fraud detection.



Steps to Process Healthcare Claims Using AI + Cloud

1. Data Ingestion

- Collect claims data from multiple sources: EHRs, provider portals, billing systems, scanned documents.
- Use cloud-native ETL/ELT pipelines (e.g., SnapLogic, dbt, Snowflake) to ingest structured and unstructured data.
- AI-powered OCR/NLP (like AWS Textract or Azure Form Recognizer) extracts information from paper forms and PDFs.

2. Data Validation

- AI models check for missing fields, mismatched codes, or duplicate entries.
- Cloud-native rules engines validate against payer policies and regulatory standards.

10.48047/jocaaa.2023.31.04.73

- Real-time error detection reduces denial rates.

3. **Claims Adjudication**

- Machine learning models predict claim outcomes based on historical patterns.
- AI automates decision-making for routine claims, while flagging exceptions for human review.
- Cloud-native microservices scale adjudication across millions of claims simultaneously.

4. Fraud Detection

- AI analyzes thousands of attributes per claim to detect anomalies.
- Predictive models identify suspicious billing patterns (e.g., duplicate submissions, upcoding).
- Cloud-native platforms allow continuous retraining of fraud models with fresh data.

5. Approval & Payment

- Validated claims are routed to payment systems automatically.
- Cloud-native APIs integrate with banking/payment gateways for seamless disbursement.
- Audit trails are maintained for compliance.

6. Analytics & Reporting

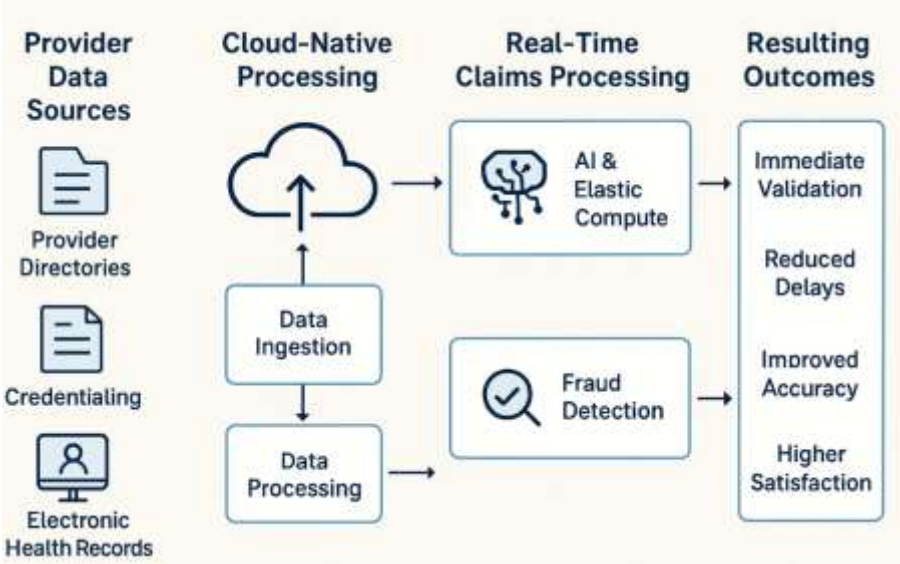
- AI dashboards provide insights into denial trends, fraud hotspots, and provider performance.
- Predictive analytics forecast claim volumes and costs.
- Cloud-native BI tools (Snowflake + Power BI/Tableau) deliver real-time reporting Such comprehensive analytical capabilities, powered by cloud infrastructure, offer healthcare organizations unprecedented visibility into their financial operations, enabling data-driven strategic planning and resource allocation. The ongoing evolution of these integrated AI and cloud platforms ensures continuous adaptation to new regulatory requirements and emerging fraud schemes, thereby maintaining robust and future-proof claims processing systems.

Real time claims Provider data processing using cloud technologies

Cloud-based solutions offer significant advancements in managing and processing provider data, delivering capabilities such as enhanced data security, streamlined workflows, and real-time validation [3]. This architectural paradigm supports dynamic scaling and ensures high availability, crucial for the continuous and secure operation of sensitive healthcare data [3]. Real-time claims and provider data processing in cloud-native environments leverages artificial intelligence (AI) and elastic compute to deliver immediate validation, adjudication, and fraud detection. Unlike batch-oriented legacy systems, cloud-native architectures enable continuous ingestion of structured and unstructured data from electronic health records, provider directories, and billing systems. Provider data is continuously synchronized across systems, ensuring accuracy in credentialing, network participation, and compliance reporting. The combination of scalable microservices, serverless pipelines, and predictive analytics allows healthcare organizations to process millions of claims

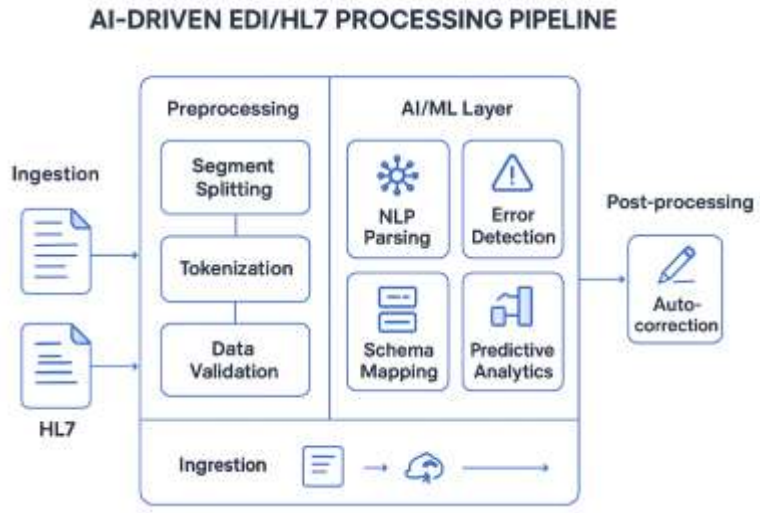
10.48047/jocaaa.2023.31.04.73

concurrently while maintaining regulatory compliance and auditability. This real-time capability not only improves operational efficiency but also enhances patient and provider satisfaction by reducing delays and disputes in the claims lifecycle.



EDI and HL7 files processing using AI

The application of artificial intelligence to process Electronic Data Interchange and Health Level Seven files significantly enhances the automation and accuracy of data exchange within healthcare ecosystems. These AI-driven approaches enable efficient extraction, transformation, and loading of complex healthcare data, facilitating seamless interoperability between disparate systems [3].



AI for EDI Processing (EDI X12 837, 835, 270/271, etc.)

1.1Extracting Segments and Tokens Using NLP

EDI files contain structured segments (e.g., NM1, REF, CLM, N3, N4).
AI helps by:

10.48047/jocaaa.2023.31.04.73

- Learning patterns of segment structure using **sequence modeling**
- Extracting incorrect or missing fields
- Predicting the correct field based on context
- Auto-filling missing data

AI Techniques Used

- Token classification using **BiLSTM** or **Transformer** models
- NLP segmentation models for semi-structured text
- Statistical anomaly detection

Example:
AI can detect if an NPI (REF*EI, NM109) is invalid and predict the correct provider ID based on historical claims.

1.2 EDI Error Detection Using ML

Claims often fail due to:

- Missing segments
- Invalid values
- Address errors
- Incorrect billing provider details

Machine learning models analyze historical EDI data to identify patterns that lead to claim denials.

Models Used

- Gradient Boosting Machines
- Random Forests
- Autoencoders for anomaly detection

2. AI for HL7 File Processing (HL7 v2, v3, FHIR)

HL7 messages (e.g., ADT, ORU, ORM, VXU) are often complicated because each provider sends data differently.

2.1 Parsing HL7 Using NLP

HL7 v2 messages contain pipe-separated fields (e.g., MSH, PID, PV1).
AI processes these by:

- Splitting incorrectly formatted segments
- Predicting missing HL7 fields

10.48047/jocaaa.2023.31.04.73

- Understanding context (e.g., mapping PID-5 to Patient Name fields)

Transformers and BERT-like models handle unstructured or poorly formed HL7 messages.

Data security in cloud

Robust data encryption, access controls, and compliance frameworks are critical for safeguarding sensitive patient information within cloud environments [3]. **Cloud data security is the practice of protecting sensitive information stored, processed, or transmitted in cloud environments through encryption, access controls, monitoring, and compliance frameworks.**

1. Shared Responsibility Model

Cloud security operates on the premise that responsibilities are split between:

Cloud Provider (AWS, Azure, GCP):

Physical data centers, hardware, infrastructure, network, hypervisor security.

Customer (organization):

Identity management, access control, data encryption, application security, compliance enforcement.

Understanding these boundaries is essential for implementing secure cloud operations.

2. Data Encryption

Cloud security relies heavily on strong encryption techniques:

a. Encryption in Transit

TLS 1.2+ for APIs, load balancers, VPN tunnels

Ensures protection against packet sniffing and man-in-the-middle attacks

b. Encryption at Rest

Managed encryption (AES-256 most common) using cloud-native tools such as:

AWS KMS, Azure Key Vault, GCP KMS

Customers control key access and rotation policies

c. Bring Your Own Key (BYOK)

Allows organizations to encrypt cloud data with keys they generate themselves, adding another security layer.

3. Identity and Access Management (IAM)

Access control is considered the “first line of defense”.

Key IAM Best Practices

10.48047/jocaaa.2023.31.04.73

Least-privilege access

Multi-factor authentication (MFA)

Role-Based Access Control (RBAC)

Fine-grained resource permissions

Periodic access review and automatic key rotation

Cloud-native IAM services provide centralized authentication and security monitoring.

4. Network Security

Cloud platforms offer multiple layers of virtualized network protections:

Virtual Private Clouds (VPCs)

Subnets (public/private)

Security groups and network ACLs

Web Application Firewalls (WAF)

Distributed Denial-of-Service (DDoS) protection (AWS Shield, Cloud Armor)

Segmentation helps prevent lateral movement and isolates workloads.

5. Data Governance and Compliance

Organizations must ensure compliance with industry and regulatory standards:

HIPAA (healthcare)

HITECH

GDPR

Conclusion

The integration of artificial intelligence into cloud-native data platforms represents a pivotal advancement in improving the accuracy, reliability, and operational efficiency of provider and claims data across healthcare ecosystems. By leveraging machine learning–based identity resolution, NLP-driven contract interpretation, and real-time anomaly detection, organizations can significantly reduce provider data inconsistencies, eliminate manual processing errors, and streamline claims adjudication workflows. Cloud-native architectures further enhance scalability, security, and interoperability, enabling continuous data quality validation and automated reconciliation at enterprise scale.

The findings of this work demonstrate that AI-enabled data pipelines deliver measurable improvements, including faster provider onboarding, reductions in claim rework and denials, and substantial improvements in end-to-end data integrity. These advancements not only strengthen payer operations but also contribute to broader industry goals such as improved reimbursement accuracy, enhanced regulatory compliance, and a more seamless provider–payer ecosystem. The

10.48047/jocaaa.2023.31.04.73

combined use of AI and cloud technology forms a robust foundation for next-generation healthcare data management—one capable of supporting evolving industry standards, interoperability mandates, and value-based care initiatives. Future research may explore federated learning, advanced FHIR-based interoperability, and multimodal AI models to further expand the impact of this work.

References

[1] D. Thesmar, D. Sraer, L. Pinheiro, N. Dadson, R. Veliche, and P. E. Greenberg, "Combining the Power of Artificial Intelligence with the Richness of Healthcare Claims Data: Opportunities and Challenges," *PharmacoEconomics*, vol. 37, no. 6, p. 745, Mar. 2019, doi: 10.1007/s40273-019-00777-6.

[2] V. Kilanko, "Leveraging Artificial Intelligence for Enhanced Revenue Cycle Management in the United States," *International Journal Of Scientific Advances*, vol. 4, no. 4, Jan. 2023, doi: 10.51542/ijscia.v4i4.3.

[3] L. R. Lekkala, "Cloud Technologies and Its Impact on the US Health Insurance Claims Process," *Voice of the Publisher*, vol. 9, no. 4, p. 323, Jan. 2023, doi: 10.4236/vp.2023.94025.

[4] J. R. Machireddy, "Automation in healthcare claims processing: Enhancing efficiency and accuracy," *International Journal of Science and Research Archive*, vol. 9, no. 1, p. 825, Jun. 2023, doi: 10.30574/ijsra.2023.9.1.0435.