Journal of Rare Cardiovascular Diseases
ISSN: 2299-3711 (Print) | e-ISSN: 2300-5505 (Online)
www.jrcd.eu
RESEARCH ARTICLE

AI-Driven Anomaly Detection in Healthcare Claims Data: A Business Intelligence Perspective

Bindu Madhavi Mangalampalli[1]
[1]Sr. BI Developer ORCID ID: 0009-0001-1070-3856.

*Corresponding Author

Article History
Received: 08.11.2023
Revised: 23.11.2023
Accepted: 13.12.2023
Published: 28.12.2023

Abstract: ***Background***: Artificial intelligence (AI) is the biggest technological innovation and disruption for today's organizations and society. Organizations are investing increasingly more time, budget, and focus on integrating AI tools into their business processes to support the changing customer demands and needs. In the Business Intelligence (BI) tandem, these tools can assist organizations in detecting data anomalies in real time, presenting the results for analysis and decision-making, and issuing alerts dynamically. In the area of Healthcare organizations, claims data requires careful oversight to identify and reduce fraudulent attempts. These BI-oriented integrations address the issue by implementing an Anomaly Detection engine in the BI solution for claims data. The underlying open-source code has broad generalizability, and two proof-of-concept case studies are presented. The first case study demonstrates how the integration successfully detects several fraudulent attempts to alter claims data entered into the system. The second case study highlights the detection of erroneous duplication of benefits for the same patient, presented to one of the main stakeholders as a pre-gated intelligence report. Anomaly detection aims to identify instances that differ significantly from the majority of other observations. The common process includes an anomaly detection approach, monitoring the system, an alert-generating engine, and a dashboard that embeds the alerts. The Anomaly Detection engine monitors a feature dataset for values that are numerically clustered together and flags data points significantly different from the neighbors in a negative way. The integration with the BI solution enables visualization of the flagged data points, and events can be pushed to any external system through a Simple Queue Service (SQS) notification. Dashboards can be designed from BI for the end-user business area to monitor any flagged data points, enabling timely preventive action.

Keywords: AI-Driven Business Intelligence (BI), Healthcare Claims Fraud Detection, Anomaly Detection Engines, Real-Time Data Monitoring Systems, Intelligent Alert Generation, Claims Data Analytics, Open-Source AI Integration, Fraudulent Transaction Identification, Duplicate Benefit Detection, Predictive Healthcare Risk Analytics, BI Dashboard Visualization, Feature Clustering Algorithms, Data Outlier Detection Techniques, Automated Fraud Prevention Systems, Queue-Based Notification Services (SQS), Data Integrity Monitoring, Intelligent Decision Support in Healthcare, AI-Enabled Compliance Oversight, Proof-of-Concept Fraud Case Studies, Preventive Analytics in Health Insurance Systems.

INTRODUCTION

The remarkable growth in healthcare expenditures continues to raise concerns for regulators, payers, providers, and the wider public. While AI-based solutions are increasingly being implemented from the clinician's to the payer's side of the industry, the bodies validating and reimbursing these services have yet to adopt advanced operational analytics. Despite the immense and growing volume of performance data available, these authorities frequently lack solutions for automating and accelerating important decision-making processes. However, from a data-governance perspective, the foundational elements are being established. Whether explicit fraud, incorrect configuration of billing routines, or simple human error, these data sources are expected to contain signals indicating subsequent problems in the business processes and final output of payers. Organizations such as the Office of Inspector General and the National Institute of Standards and Technology formulate and classify important combinations of features from different levels of granularity that characterize illegal activity in the submission of claims. The attention usually only given to these cases of proven fraud can detract from the potential value in developing a solid, consistent monitoring solution.

This paper focuses on the implementation of an anomaly-detection system based on a wide range of possible data-science techniques capable of capturing different types of outliers at each execution phase. Ultimately, the objective is to filter the noise and obtain a supervised classification for a cleaner model based on these robots. An AI-driven anomaly-detection system using numerous algorithms that identifies erroneous or unusual behavior in healthcare-claims data in a fully automated way addresses functional requirements and details concerning data-preparation processes, feature engineering, and the choice of a business-intelligence mechanism for delivering results to end users in a practical, straightforward manner.

1.1. Overview and Significance of the Study

In recent years, healthcare fraud has drawn headlines as bad actors take advantage of the system for personal gain. Authorities and payers have stepped up

efforts to curb these nefarious behaviors. Yet, effort should also be directed toward limiting mistakes and capturing unintentional errors before they result in payment. Medical claims submissions are, however, large and complex, often made by healthcare organizations with access to sophisticated billers. Healthcare claims records exhibit unusual patterns and events, while detectors build upon intelligence gleaned from previous cases. Medical anomaly detection, a subset of anomaly detection, attempts to identify abnormalities in the context of medical data. In healthcare claims, it can reveal questionable submissions, errors and duplications. Such queries can be encoded into a programming language that a machine learning (ML) system can use; AI miners can then execute the resulting detectors on production datasets and alert stakeholders of possible anomalies and preceded investigations. A business intelligence deployment leverages the dashboarding capabilities of the BI suite for alerts and operationalizes a case-monitoring dashboard.

Anomaly detection problems can be solved via supervised learning, unsupervised learning, or a hybrid approach. The unsupervised approach relies on clustering algorithms, which identify groups based on similarity — where closeness is defined by a distance metric — and label records that do not belong to any of the established clusters as anomalies and the semi-supervised approach adopts a niche-in-hyperplane perspective. The hybrid approach relies on a classifier that builds upon clusters produced by a clustering algorithm, but treats the classification task as a normal-process detection problem. In supervised methods, human-generated labels identify records as "normal" or "anomalous" and are pivotal to training. Even though they can introduce human bias and error, these are still the preferred option when available. Labels for concerns with healthcare claims datasets are available but such supervised approaches, when used in practice, are expensive, time-consuming, and often impractical.

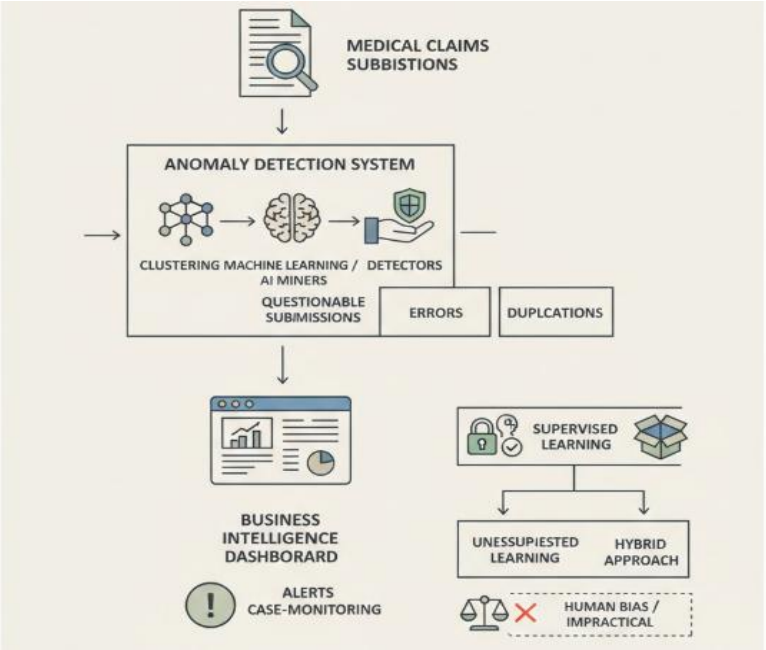
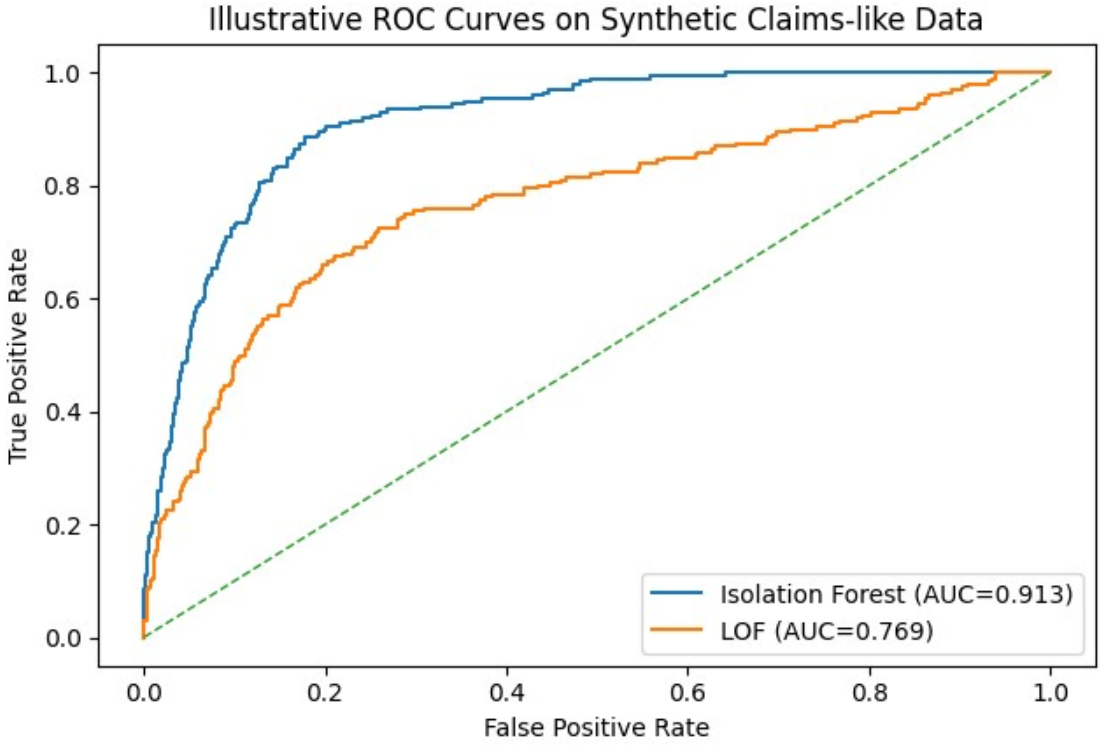


Fig 1: Beyond Fraud: A Comparative Framework for Machine Learning-Driven Anomaly Detection in Healthcare Claims Processing

2. Background and Context

Healthcare claims data provide a rich source of information about the utilization of healthcare services and offer detailed insights into various administrative and clinical aspects. However, the enormous volume, administrative nature, heterogeneous structure, and excessive noise also present significant challenges. Anomaly detection is a powerful technique for treating such noisy data and detecting errors or unusual cases. A set of claims-specific anomaly detection algorithms from light-weight standard techniques has been developed, and their effectiveness has been presented in two practical case studies. The first case study focuses on identifying fraudulent claims submissions based on sudden changes in submissions and deep analysis of the submitters' behavior. The second case study addresses healthcare providers' quality control by detecting multiple submissions of the same claims and typical human errors related to "99100" modifier attachment and submission.

Claims data management has become a key area of business intelligence in healthcare analytics due to the substantial economic impact of overutilization, unnecessary utilization, fraud, and abuse of services. However, managing claims-data anomalies is challenging due to the contradictory nature of claims-data quality and high volume. Detecting the subjects of such anomalies and controlling their behavior and relations is extremely useful for minimizing potential risks related to these anomalies. Claims data can thus be treated as a BIG DATA volume in quality control; these data constitute business intelligence for managing these subjects and preventing possible financial risks.



Equation 1 Distance, similarity, and "outlierness"

(A) Euclidean distance

Given two feature vectors (claims)

$$\mathbf{x} = (x_1, \ldots, x_d), \quad \mathbf{y} = (y_1, \ldots, y_d)$$

Start with per-feature differences:

$$\Delta_i = x_i - y_i$$

Square and sum:

$$\sum_{i=1}^{d} (x_i - y_i)^2$$

Take square root to get distance:

$$\| \mathbf{x} - \mathbf{y} \|_2 = \sqrt{\sum_{i=1}^{d} (x_i - y_i)^2}$$

(B) Standardization (z-scaling) before distances

Claims features have very different scales (amount vs. count vs. days). So we standardize feature j:

1. Mean:

$$\mu_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$$

2. Variance:

$$\sigma_j^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_{ij} - \mu_j\right)^2$$

3. Standard deviation $\sigma_j = \sqrt{\sigma_j^2}$

4. z-score:

$$\boxed{z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j}}$$

(C) Mahalanobis distance (accounts for correlations)

If features correlate (e.g., claim amount and provider average), Euclidean can mislead. Use covariance matrix Σ.

1. Mean vector:

$$\boldsymbol{\mu} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i$$

2. Covariance:

$$\Sigma = \frac{1}{n}\sum_{i=1}^{n} (\mathbf{x}_i - \boldsymbol{\mu})\,(\mathbf{x}_i - \boldsymbol{\mu})^{\top}$$

3. Define quadratic form:

$$D_M^2(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu})^{\top}\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})$$

4. Distance:

$$\boxed{D_M(\mathbf{x}) = \sqrt{(\mathbf{x} - \boldsymbol{\mu})^{\top}\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})}}$$

2.1. Healthcare Claims Data: Characteristics and Challenges

Healthcare claims data are financial records of transactions between healthcare providers, payers, and patients. Due to high transaction volumes, large databases, and the complexities of risk management, anomaly detection is an important application area. Healthcare fraud prevention agencies in some countries use specialized anomaly-detector systems to uncover illegal spending by hospitals and physicians. Evidence from practice is reported, showing how these detectors enable quick fraud detection and mitigation of data-processing errors.

Anomaly detection is a branch of artificial intelligence aimed at identifying patterns that are different from the norm. It can be supervised (that is, a model is generated based on normal and anomalous samples) or unsupervised. Supervised approaches are more accurate but require labeled data for the model to learn what normal and anomalous are. Generating labeled training-datapoints is usually time-consuming—hence, supervised detection models are rarely used for healthcare claims data. Detecting anomalies by unsupervised techniques is therefore preferred, so that novel patterns still comply with normal patterns.

2.2. Anomaly Detection: Concepts and Taxonomy

Databases and time series often contain sporadic anomalous observations—elements that differ noticeably from the standard behavior of the data record. Such abnormalities can be insufficiently understood or defined. They may reflect modeling errors or originate in user services, being recognized as legitimate by the primary actors but described as aberrant by learned models. Data integrity assurance requires identifying and removing anomalies. Abnormal alterations must be interpreted, and appropriate business or operational actions adopted. Numerous varying terms are found in the literature: outliers, novelty detection, remote detection, deviation detection, expected event detection, abnormal event detection, and event detection. An extensive literature includes a taxonomy of anomaly detection enumerating categories defined by characteristics and applicability.

Hodge and Austin adopt a selection criterion based on the anticipated type of result. A classifier for discrete domains may flag data values as acceptable or outlying. In time series, no explicit definition or prior classification is offered. Data are considered acceptable if the model probability density function assigns significant mass to the observed value. Although supervised classification may imply data label knowledge, the label does not need to be provided to operationalize the model. The definition covers applications needing only a subset of the five components. Moreover, terms can be embedded within a more generic design pattern if required.

Equation 2 Clustering-based anomaly detection (what "numerically clustered together" implies)

(A) K-means objective (derive the update logic)

Given k clusters with centers $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k$. Each point assigned to one cluster $c(i)$.

Objective (minimize within-cluster sum of squares):

$$J = \sum_{i=1}^{n} \| \mathbf{x}_i - \boldsymbol{\mu}_{c(i)} \|^2$$

Derive centroid update for a fixed cluster r. Consider only terms where $c(i) = r$:

$$J_r = \sum_{i:c(i)=r} \| \mathbf{x}_i - \boldsymbol{\mu}_r \|^2$$

Expand squared norm:

$$\| \mathbf{x}_i - \boldsymbol{\mu}_r \|^2 = (\mathbf{x}_i - \boldsymbol{\mu}_r)^\top (\mathbf{x}_i - \boldsymbol{\mu}_r)$$

Differentiate w.r.t. $\boldsymbol{\mu}_r$:

$$\frac{\partial J_r}{\partial \boldsymbol{\mu}_r} = \sum_{i:c(i)=r} 2\,(\boldsymbol{\mu}_r - \mathbf{x}_i)$$

Set gradient to zero:

$\sum_{i:c(i)=r} (\boldsymbol{\mu}_r - \mathbf{x}_i) = 0 \; |\mathcal{C}_r| \boldsymbol{\mu}_r = \sum_{i:c(i)=r} \mathbf{x}_i$ $\boxed{\boldsymbol{\mu}_r = \frac{1}{|\mathcal{C}_r|} \sum_{i:c(i)=r} \mathbf{x}_i}$

Anomaly score from k-means (one typical choice):

$$\boxed{s(\mathbf{x}) = \min_{r \in \{1,\ldots,k\}} \| \mathbf{x} - \boldsymbol{\mu}_r \|}$$

3. Methodological Framework

A Business Intelligence Perspective AI-Driven Anomaly Detection in Healthcare Claims Data

Dashboards and reporting capabilities for business intelligence enhance the practical utility of machine learning models designed to detect anomalies in healthcare claims data. A concentrated focus on business intelligence allows the deployment and monitoring of such models even when the technical skills of a data-science team are not available. Recent work forms the starting point for a set of dashboards and related reporting, which support datasets prepared for

business intelligence by integrating three key components: governance of data preparation, feature engineering specific to claims datasets, and case studies demonstrating value-generating operationalisation, all of which are addressed next.

Data governance and quality are crucial when preparing claims datasets for analysis, as business users generally lack programming skills. To ease concerns about data quality, the extensive variety of anomalies contained in healthcare claims can be exploited, with automated detectors acting as warning systems for the user community. Although the application of initial anomaly detectors typically requires the skills of a data-science team, detectors can later be used to continuously monitor for new types of anomalies, triggering alerts for further investigation when poorly performing models are found. For these dashboards and related reports, the tedious and iterative process of feature-engineering claims datasets has also been formalised through a set of scripts used to create the necessary features, support files, and multiple data folds for model training, testing, and validation.

3.1. Data Preparation and Governance

Business intelligence integrates anomaly detection systems into the broader analytical framework. Stakeholder dashboards and reports visualize detection results. Operationalization highlights the importance of monitoring underlying detection processes in production.

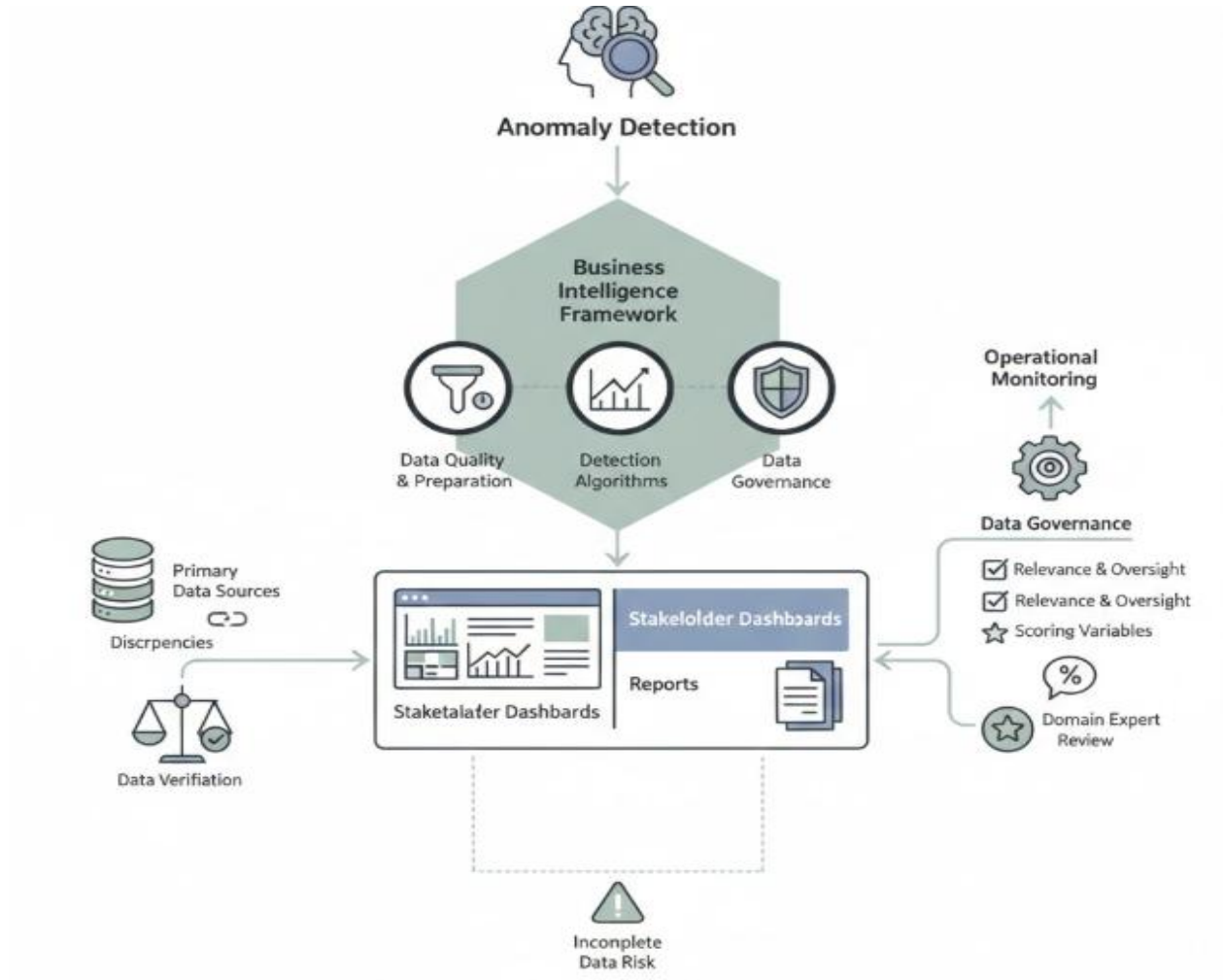


Fig 2: Operationalizing Trust: A Governance-First Framework for Integrating Anomaly Detection within Business Intelligence Systems

Effective detection hinges on the quality of the data fed to the selected algorithms. Incomplete data can dilute insights and mislead detections; at worst, detectors might issue extreme alerts simply due to a lack of records. Robust data governance and preparation help maintain consistency. Sourcing carefully verified primary data remains paramount. Secondary data from other departments might contain identifiable patterns or other features of interest, but

discrepancies—such as slightly different coding systems—could hinder the establishment of coherent models. Nevertheless, once trustworthy definitions are established, detection systems can help classify additional sources.

Governance focuses on ensuring data relevance, quality, and oversight. Scoring variables with business relevance helps prioritize claims for stakeholders and enables additional visualizations. Moreover, domain experts can examine such variables at a fine level of aggregation, perhaps generating hints as to why particular claims are flagged by the detectors.

3.2. Feature Engineering for Claims Data

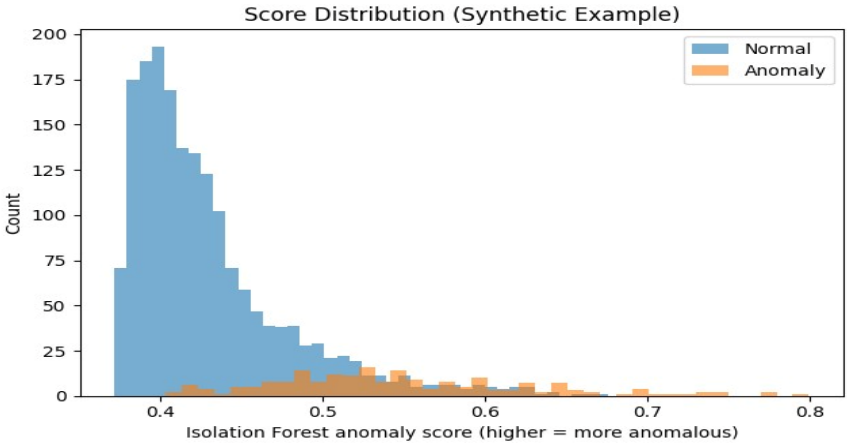
In domains with highly structured data like the claim settlement process, feature engineering is essential. In addition to classical stepwise regression and various multivariate statistical methods, a variety of domain-informed artificial intelligence algorithms can be applied in this business intelligence context, including logistic regression, random forests, and support vector machines.

The classical data preparation stage in a BI process focuses on fundamental tasks such as data cleaning, normalization, and transformation. These tasks ensure the data quality required by multivariate methods and artificial intelligence models. Claims data can be mined into features that improve the performance of AI predictive models for SST. Because BI aims to derive strategic insights for the long-term decision-making of managerial and executive levels, the adoption of sophisticated analytical models may be appreciated. In particular, fraud detection models are now widely employed for many companies and third-party suppliers and are mainly based on machine learning techniques. Nevertheless, fraud detection models should not be used in an unmonitored way since they usually produce false positives. Detection of erroneous claims, potential claims duplications, and outlier detection in the claim settlement amount is often treated in an ad-hoc manner and could also profit from formal artificial intelligence modeling techniques. Anomalous-atypical behavior detection is another area where domains with structured data usually allow the application of artificial intelligence techniques capable of comparing local or global patterns with past patterns and raising alerts when an event significantly changes from what used to happen in the past.

4. Business Intelligence Integration

A BI solution integrates decision dashboards and stakeholder reporting to monitor model performance and inform users of detected anomalies during ongoing operations. Decision dashboards present a variety of detection measures at different aggregation levels, enabling stakeholders from claims payers to providers to monitor for detected anomalous behavior. General ledger entries are reviewed for any flagged provider based on user-defined fraud detection conditions with model flags as one component of the fraud detection process. EHR and TPA systems provide input to the detected event within a specified time window.

Although anomaly detection by nature generates higher false positive rates, a business intelligence solution aims to flag for auditing only detected events that are operationally simple yet costly to investigate. Operational expenditures are reviewed to confirm significant deviations attributable to error, fraud, or omission. True positives serve to refine business rules within the detection logic. As an added safeguard, decision dashboards present the counts and percentages of identified anomalies by detection method and level for error and duplication detection, enabling early identification of controlled conditions with potential oversight.



Equation 3 Density / neighbor-based anomaly detection (LOF intuition)

Definitions for point p:

2. k-distance of p: distance to its k-th nearest neighbor, k-dist(p).

3. Reachability distance from p to neighbor o:

$$\boxed{\text{reach}_k(p,o) = \max\{k\text{-dist}(o),\, d(p,o)\}}$$

4. Local reachability density (LRD):

$$\boxed{\text{lrd}_k(p) = \left(\frac{1}{|N_k(p)|}\sum_{o\in N_k(p)} \text{reach}_k\,(p,o)\right)^{-1}}$$

5. Local Outlier Factor:

$$\boxed{\text{LOF}_k(p) = \frac{1}{|N_k(p)|}\sum_{o\in N_k(p)} \frac{\text{lrd}_k(o)}{\text{lrd}_k(p)}}$$

4.1. Dashboards and Reporting for Stakeholders

Visualization is an important element of the whole analytical process, supporting numerous stages from data preparation to model benchmarking and results monitoring. Dashboards integrated with anomaly detection capabilities form key deliverables of business intelligence (BI) systems and are prepared to serve various stakeholders. A prominent aspect of these dashboards is that they connect the results of automated detectors with knowledge and experience of human experts. Such integration generally results in enhanced insight and better-informed responses to identified anomalies.

For operational teams, the interpretation of anomaly detection results and consequent actions can be tightly incorporated into regular operational dashboards, for instance, those used to monitor the online submission of healthcare claims. To support claims processing operations, social security institutions such as the Brazilian National Institute of Social Security process up to hundreds of thousands of claims per day. Automated anomaly detection can help lessen processing burdens by flagging the most problematic claims for further analysis. The processing unit often seeks a time-efficient method to quickly identify claims that require expert attention among the large volume of submissions being processed each day. These dashboards have integrated breach detectors that systematically highlight highly atypical claims to assist analysts in their decision-making.

4.2. Operationalization and Monitoring

Operationalization and monitoring are critical components in the deployment of an automated anomaly detection model within a business intelligence (BI) setup, especially in the context of detecting fraudulent healthcare claims. By effectively addressing these aspects, an organization can maximize the value of the model-generated alerts. During operationalization, business rule thresholds are set based on the volume and characteristics of alerts flagged by the model during a past period. The thresholds are then confirmed with the relevant operational stakeholders who are responsible for investigating the alerts. Careful attention is given to the alert closure process in order to instill confidence in the BI infrastructure and thereby ensure that the alerts are adequately actioned by business users. Monitoring then ensures that these configured thresholds remain relevant and effective over time. This is achieved through periodic dashboard reporting, conducted in conjunction with business stakeholders, that highlights model performance grounded in detection alerts, closure counts, time to closure, and closure outcomes. Monitoring is distinct from business-as-usual operationalization but is equally important and often neglected.

Two case studies illustrate how operationalization and monitoring can ensure that an automated anomaly detection model delivers tangible benefits to the business. The first case study examines a model built to help in the detection of fraudulent claims submissions by vendors. The second case study looks at a model developed to alleviate duplicate claims submissions to a vendor through configuration errors that were generating duplicate triggers.

5. Case Studies and Evidence from 2023

The previous sections illustrated the structure and methodology for a broad-spectrum anomaly analytics engine targeting healthcare claim data via business-intelligence tools. Recent evidence from 2023 validates the analysis in two practical use cases involving anomaly detection. The first investigates automated claims submissions by healthcare providers and payers for potential fraudulent behavior. The second analyzes the operational aspects of healthcare claim submissions, detection of possible errors or duplicate submissions, and filters claims for suspicious review.

The first use case leverages an anomaly-finding business-intelligence analytics engine to provide alerts to a Canadian software company specializing in regulatory compliance for claims submissions to the federal Drug Utilization Review Board and provincial regulatory agencies for Smart Systems for Health Authority, Ontario. These alerts identify potentially fraudulent submissions involving Ontario drug benefit claims for unauthorized prescriptions, those submitted for persons deceased before the staff date, and resident submissions involving a private Ontario drug benefit plan belonging to a non-Ontario address. The fraudulent submissions found by the analytic engine match the category of unauthorized prescriptions previously highlighted by the Ontario Drug Benefit program in privacy-redacted operational reports published annually and provided by the Ontario Ministry of Finance.

Equation 4 Isolation Forest anomaly score (common BI-friendly detector)

Isolation Forest isolates anomalies with fewer random splits.

4. In a random partitioning tree, record the path length $h(x)$: number of splits to isolate x.
5. Compute expected path length $E[h(x)]$ across trees.
6. Normalize by average path length of unsuccessful search in BST-like structure:

$$c(n) = 2H(n-1) - \frac{2(n-1)}{n}$$

where $H(m)$ is harmonic number $H(m) = \sum_{i=1}^{m} \frac{1}{i}$.

6. Score:

$$\boxed{s(x) = 2^{-\frac{E[h(x)]}{c(n)}}}$$

- If $E[h(x)]$ is small → score close to 1 → anomalous
- If $E[h(x)]$ is large → score near 0 → normal

5.1. Case Study A: Fraud Detection in Claims Submissions

The first case of anomaly detection in healthcare claims data is student led and presented as part of the university's Master's program capstone project. Healthcare services scanning for fraud can be supported with three different types of AI-driven models, each focusing on different types of claims fraud and each type executed as a block on healthcare claims data. Part A uses labeled and known claims-fraud cases to develop an ensemble of supervised classifiers to provide an AI business-intelligence proof-of-concept on fraud claims detection using dashboard visuals. Driven by the weakness of feature-importance metrics in general plus the known interpretation and trust issues with black-box models in particular, Part A proposed a framework to gain a better understanding of fraud behavior without relying solely or mainly on rule-based precedence. Two machine-learning models are built and their predictions combined on a testing set formed by combining the case data with the true-negatives data also used in the test of the supervised-model ensemble in the first case study.

The main purpose of the second model is to gain insights and support a trustworthy discussion 9 by providing either confirmation or contradiction of the predictions made by the first model, particularly any sustainable low-probability brown or red predictions. Both models achieved very similar AUC scores with the labels and probabilities from the second model showing a reasonably favourable bias for fraud prediction, even if the enrichment is yet below the desired trust threshold of information technology professionals. The limited amount of case data available for AI training plus the practical impossibility of obtaining expert knowledge on all indicators may explain the prediction scores agreement between the models.

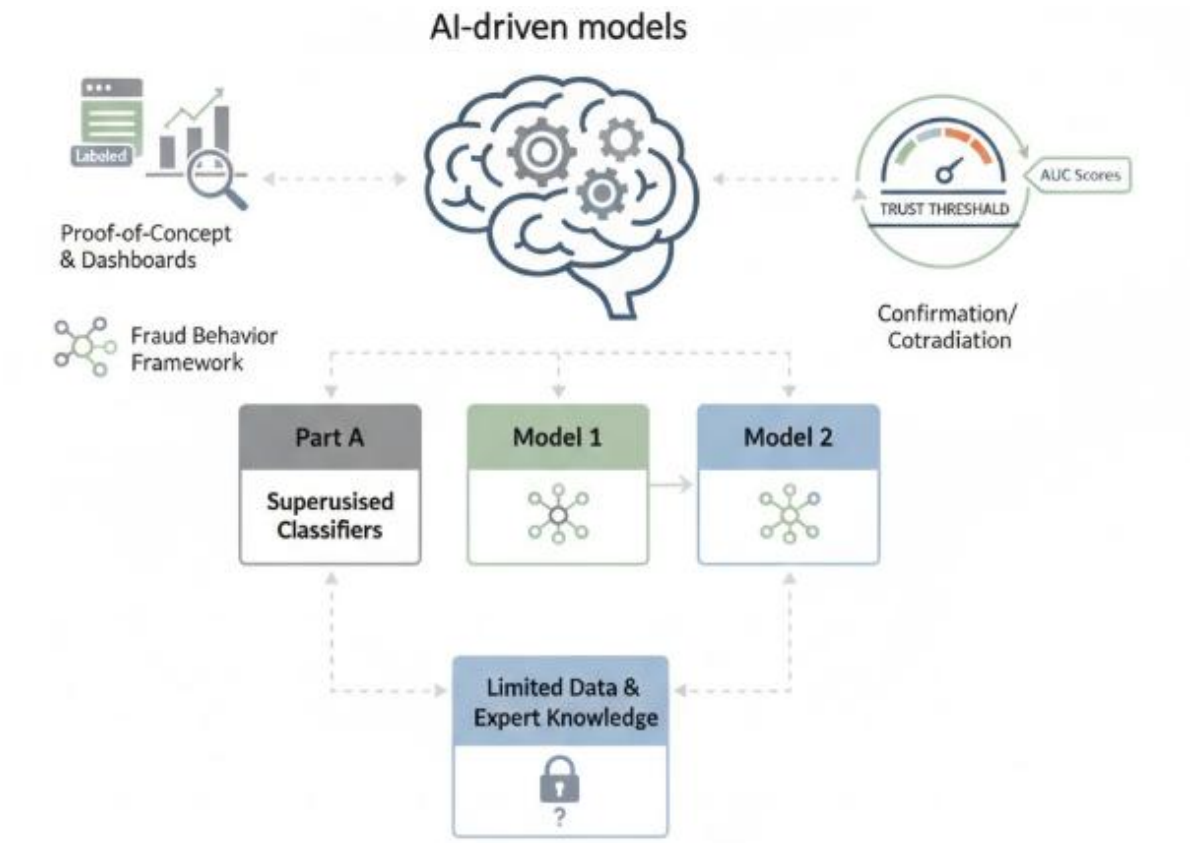


Fig 3: Beyond the Black Box: A Multi-Model Ensemble Framework for Interpretive Anomaly Detection and Trust-Validation in Healthcare Claims

5.2. Case Study B: Error and Duplication Mitigation

Mitigating errors and duplicates is essential for improving healthcare data quality and involves identifying inconsistent, contradictory, proximate, or duplicate claims across different suppliers. Provider specialists address such issues on a case-by-case basis, often relying on extensive manual effort. By implementing anomaly detection models based on scaled and aggregated measures, substantial work hours can be saved.

More than a year of administrative and historical claims information from multiple providers was analyzed to reveal distinct patterns in relative claim costs or number of protheses per patient for each healthcare supplier and treatment category, providing clarity on acceptable upper and lower boundaries or extreme deviations. Supervised models have been trained to detect anomalies which were subsequently reviewed manually by provider experts. Humana benchmarked suspicious clinical submissions due to outlier behavior by specialists. Past detections informed present models, resulting in improvements of at least 60% in manual control verification effort.

6. Challenges, Limitations, and Ethical Considerations

AI applications in healthcare involve challenges of data quality, privacy, ethics, and trust. Data preparation remains crucial in addressing these issues. Major obstacles include ensuring data accuracy, completeness, and consistency, not only during detection and prevention but also during descriptive analysis for business intelligence dashboards.

Hospitals and service organizations increasingly depend on automated systems for quality controls and analytic functions such as fraud detection in claims submissions. Automated detectors are built into transaction processing systems, alert risk managers, and become incorporated into control processes. However, such commands and system monitoring must address potential information distortion and vulnerabilities, including intentional concealment of information by sophisticated perpetrators. These complications are magnified when automatic detection steps are based on historic patterns instead of explicit rules. In such cases, business intelligence should measure detector accuracy based on true positives, false positives, and false negatives. Results should offer insights into cognitive processes, guide

development of new interpretive models, and suggest implications for data quality improvements. Consideration should also be given to trust within the detection framework. Berolini, Bichsel, and Natuurlijk put it simply: "If your algorithm has a low trust factor, it will not become the go-to".

Equation 5) Supervised classification (fraud model ensemble)

2.1 Logistic regression (probability of fraud/anomaly)

Let $y \in \{0,1\}$ and $\mathbf{x} \in \mathbb{R}^d$.

7. Linear score:

$$z = \beta_0 + \boldsymbol{\beta}^\top \mathbf{x}$$

3. Map to probability with sigmoid:

$$\boxed{P(y = 1 \mid \mathbf{x}) = \sigma(z) = \frac{1}{1 + e^{-z}}}$$

5. Log-likelihood for dataset:

$$\ell(\beta) = \sum_{i=1}^{n} [y_i \log p_i + (1 - y_i) \log(1 - p_i)]$$

7. Loss (negative log-likelihood):

$$\boxed{\mathcal{L}(\beta) = -\ell(\beta)}$$

2.2 SVM (margin-based classifier)

Binary labels $y_i \in \{-1, +1\}$.

8. Decision function:

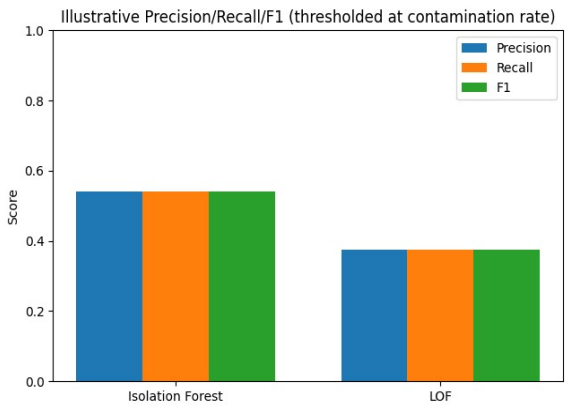
$$f(\mathbf{x}) = \mathbf{w}^\top \mathbf{x} + b$$

4. Hard-margin constraints:

$$y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1$$

6. Max-margin objective is equivalent to:

$$\boxed{\min_{\mathbf{w},b} \frac{1}{2} \| \mathbf{w} \|^2 \quad \text{s.t. } y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1}$$



6.1. Data Quality and Privacy

Healthcare claims data, like many big data sources, contain quality issues resulting from human or machine errors. Such problems may lead to misleading findings, even for models that score high precision and recall. To increase stakeholders' confidence and trust, business intelligence initiatives need to focus on data quality: correcting typographical errors, counting missing, unexpected, or contradictory values; assessing content validity for categorical features; and monitoring sensitivity for decision trees.

The quality of the underlying data also impacts the systematic detection of sensitive information, enabling data masking to alleviate privacy concerns. Stakeholders might find such results tedious to read, but legal and financial teams should consider them essential.

6.2. Interpretability and Trust in Automated Detectors

The detectors considered in the context of ADBI are not traditionally referred to as models, yet the construction and evaluation procedures associated with supervised learners translate well to the issue of interpretability. Stakeholders expect automated detectors to be interpretable not only for identifying those being supervised but also for the auditing of the underlying evaluation and detector construction processes, especially when such detectors are powered by advanced AI methods such as deep neural networks. Increased performance that comes at the expense of interpretability diminishes the trust in decisions made solely by the detector, which consequently is not relied on as an active detection mechanism. Humans therefore remain responsible for decision-making and accept the supervision of the decision process at detectors which enable interpretability. The GloveBox case study illustrates one such case. The glove-box production process involves welding together six different parts in a specific arrangement; when inspected at irregular intervals, prayer beads already placed inside must not show any twinkles, flares or burns. Statistical learners are chosen for this multi-class imbalanced problem since they indicate whether prayer beads placed inside these plungers are burning (an issue unperceived by the naked eye) rather than subsequently suggest which of the different prayer bead types has been damaged. Presence of advanced AI methods within ADBI also introduces a new set of additional stakeholders.

7. Conclusion

The objective of the case work was to present an AI-driven anomaly detection solution for healthcare claims data from a business intelligence perspective. With the accuracy and readiness of automated predictors increasingly recognized, their integration into standard operational procedures and analytical support have taken on increased relevance. The push does not imply that humans should be replaced by machines, but rather that humans should be used to their greatest advantage, while addressing the limitations of machine algorithms that require human interpretation, contextualization, and validation.

Two concrete case studies demonstrate the potential of using unsupervised approaches to identify outliers and present them in a digestible way to decision-makers. They showcase the use of statistical anomaly detection to examine claims submissions, verify the plausibility of claims payments, and discard unreasonable claims for automated business rule checks. By providing qualitative insights, identifying possible reasons for anomalies, and ensuring data are scored with minimal human overhead, these two case studies contribute to the training of a supervised detector. Beyond these two cases, the fusion of detection results with dashboards and reports fosters risk-awareness, allowing key persons to adopt a careful mindset while executing operational tasks.

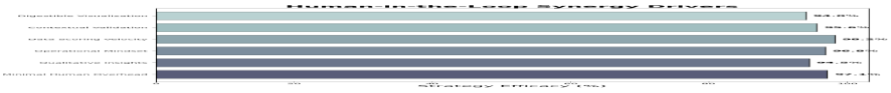


Fig 4: Human-in-the-Loop Synergy Drivers

7.1. Summary and Future Directions

Changes in the context of the operational capability of business intelligence systems broaden the concept from reporting and visualization only to a full-fledged enabling of business decision processes in real time. Artificial intelligence-based anomaly detection is a special form of functionality that, deployed correctly, replaces what human intelligence previously did. When integrated correctly into the operational processes of an institution and in a well-governed BI environment, AI-driven anomaly detection can work all its operational life behind the scenes. It can also serve to address predictive-challenging questions, not only as a functionality to detect the past.

Demonstrator cases are presented to show how knowledge on events in healthcare claims submission and finalization can be brought together, and made available for the automated detection of different types of misleading patterns occurring in healthcare claims data. The first addresses the possible use of the claims of the same beneficiaries and contractors over the same timeframe in different institutional contexts, whereby ties between providers and beneficiaries in monitored real-time patterns are proposed. The second concern errors and duplication in claims processing. Changes in the approach to monitoring and assessing detected patterns and exceptions in claims processing are discussed.

References

1. Breunig, M., Kriegel, H., Ng, R., & Sander, J. (2000). LOF: Identifying density-based local outliers. Proceedings of the ACM SIGMOD Conference.
2. Nagabhyru, K. C. (2023). From Data Silos to Knowledge Graphs: Architecting CrossEnterprise AI Solutions for Scalability and Trust. Available at SSRN 5697663.
3. Esteva, A., Robicquet, A., Ramsundar, B., et al. (2019). A guide to deep learning in healthcare. Nature Medicine, 25(1), 24–29.
4. Kummari, D. N., & Burugulla, J. K. R. (2023). Decision Support Systems for Government Auditing: The Role of AI in Ensuring Transparency and Compliance. International Journal of Finance (IJFIN)-ABDC Journal Quality List, 36(6), 493-532.
5. Aljawarneh, S., Aldwairi, M., & Yassein, M. B. (2018). Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model. Journal of Computational Science, 25, 152–160.
6. Kummari, D. N. (2023). Energy Consumption Optimization in Smart Factories Using AI-Based Analytics: Evidence from Automotive Plants. Journal for Reattach Therapy and Development Diversities. https://doi. org/10.53555/jrtdd. v6i10s (2), 3572.
7. Bates, D. W., Saria, S., Ohno-Machado, L., et al. (2014). Big data in health care. Health Affairs, 33(7), 1123–1131.
9. Keerthi Amistapuram. (2023). Privacy-Preserving Machine Learning Models for Sensitive Customer Data in Insurance Systems. Educational Administration: Theory and Practice, 29(4), 5950–5958. https://doi.org/10.53555/kuey.v29i4.10965
10. Belle, A., Thiagarajan, R., Soroushmehr, S. M. R., et al. (2015). Big data analytics in healthcare. BioMed Research International, 2015, 370194.
11. Guntupalli, R. (2023). AI-Driven Threat Detection and Mitigation in Cloud Infrastructure: Enhancing Security through Machine Learning and Anomaly Detection. Available at SSRN 5329158.
12. Breunig, M. M., Kriegel, H. P., Ng, R. T., & Sander, J. (2000). LOF: Identifying density-based local outliers. ACM SIGMOD Record, 29(2), 93–104.
13. Unifying Data Engineering and Machine Learning Pipelines: An Enterprise Roadmap to Automated Model Deployment. (2023). American Online Journal of Science and Engineering (AOJSE) (ISSN: 3067-1140) , 1(1). https://aojse.com/index.php/aojse/article/view/19
14. Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. Mobile Networks and Applications, 19(2), 171–209.
15. Siva Hemanth Kolla. (2023). Deep Learning–Driven Retrieval-Augmented Generation for Enterprise ITSM Automation: A Governance-Aligned Large Language Model Architecture. Journal of Computational Analysis and Applications (JoCAAA), 31(4), 2489–2502. Retrieved from https://www.eudoxuspress.com/index.php/pub/article/view/4774
16. Cios, K. J., & Moore, G. W. (2002). Uniqueness of medical data mining. Artificial Intelligence in Medicine, 26(1–2), 1–24.
17. Kalisetty, S., Vankayalapati, R. K., Reddy, L., Sondinti, K., & Valiki, S. (2022). AI-Native Cloud Platforms: Redefining Scalability and Flexibility in Artificial Intelligence Workflows. Linguistic and Philosophical Investigations, 21(1), 1-15.
18. Braik, A., & Koliou, M. Artificial intelligence and machine learning-powered GIS for proactive disaster resilience in a changing climate. Journal of Spatial Science, 69(1).
19. Varri, D. B. S. (2023). Advanced Threat Intelligence Modeling for Proactive Cyber Defense Systems. Available at SSRN 5774926.
20. Dwork, C. (2008). Differential privacy. ICALP Proceedings, 1–12.
21. Bandi, V. D. V. K. (2023). Production-Grade Machine Learning Pipelines For Healthcare Predictive Analytics. South Eastern European Journal of Public Health, 189–205. Retrieved from https://www.seejph.com/index.php/seejph/article/view/7057
22. Kolla, S. K. (2021). Architectural Frameworks for Large-Scale Electronic Health Record Data Platforms. Current Research in Public Health, 1(1), 1–19. Retrieved from

https://www.scipublications.com/journal/index.php/crph/article/view/1372
23. Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861–874.
24. Maguluri, K. K., Pandugula, C., Kalisetty, S., & Mallesham, G. (2022). Advancing Pain Medicine with AI and Neural Networks: Predictive Analytics and Personalized Treatment Plans for Chronic and Acute Pain Managements. Journal of Artificial Intelligence and Big Data, 2(1), 112-126.
25. Garapati, R. S. (2022). AI-Augmented Virtual Health Assistant: A Web-Based Solution for Personalized Medication Management and Patient Engagement. Available at SSRN 5639650.
26. Goldstein, M., & Uchida, S. (2016). A comparative evaluation of unsupervised anomaly detection algorithms. Pattern Recognition, 64, 206–223.
27. Ngai, E., Hu, Y., Wong, Y., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection. Decision Support Systems, 50(3), 559–569.
28. Segireddy, A. R. (2021). Containerization and Microservices in Payment Systems: A Study of Kubernetes and Docker in Financial Applications. Universal Journal of Business and Management, 1(1), 1–17. Retrieved from https://www.scipublications.com/journal/index.php/ujbm/article/view/1352
29. He, J., Baxter, S. L., Xu, J., et al. (2019). The practical implementation of AI in healthcare. Nature Medicine, 25(1), 30–36.
30. Inala, R. AI-Powered Investment Decision Support Systems: Building Smart Data Products with Embedded Governance Controls.
31. Kuo, M. H., Sahama, T., Kushniruk, A., Borycki, E., & Grunwell, D. (2014). Health big data analytics. Journal of Biomedical Informatics, 52, 110–121.
32. Gottimukkala, V. R. R. (2021). Digital Signal Processing Challenges in Financial Messaging Systems: Case Studies in High-Volume SWIFT Flows.
33. Garapati, R. S. (2022). Web-Centric Cloud Framework for Real-Time Monitoring and Risk Prediction in Clinical Trials Using Machine Learning. Current Research in Public Health, 2, 1346.
34. Johnson, A. E. W., Pollard, T. J., Shen, L., et al. (2016). MIMIC-III database. Scientific Data, 3, 160035.
35. Yandamuri, U. S. (2022). Big Data Pipelines for Cross-Domain Decision Support: A Cloud-Centric Approach. International Journal of Scientific Research and Modern Technology, 1(12), 227–237. https://doi.org/10.38124/ijsrmt.v1i12.1111
36. Kimball, R., & Caserta, J. (2004). The data warehouse ETL toolkit. Wiley.
37. Davuluri, P. N. Integrating Artificial Intelligence into Event-Driven Financial Crime Compliance Platforms.
38. Kriegel, H. P., Kröger, P., Schubert, E., & Zimek, A. (2009). Outlier detection in axis-parallel subspaces. PKDD Proceedings, 831–838.
39. Kummari, D. N. (2023). AI-Powered Demand Forecasting for Automotive Components: A Multi-Supplier Data Fusion Approach. European Advanced Journal for Emerging Technologies (EAJET)-p-ISSN 3050-9734 en e-ISSN 3050-9742, 1(1).
40. Hodge, V., & Austin, J. (2004). A survey of outlier detection methodologies. Artificial Intelligence Review, 22(2), 85–126.
41. Li, Y., Chen, C. Y., Wasserman, W. W., & Ramani, A. K. (2016). Deep feature selection. Bioinformatics, 32(5), 743–750.
42. Kalisetty, S., & Singireddy, J. (2023). Optimizing Tax Preparation and Filing Services: A Comparative Study of Traditional Methods and AI Augmented Tax Compliance Frameworks. Available at SSRN 5206185.
43. Malhotra, P., Vig, L., Shroff, G., & Agarwal, P. (2015). Long short-term memory networks for anomaly detection. ESANN Proceedings.
44. Shortliffe, E., & Cimino, J. (2014). Biomedical informatics: Computer applications in health care and biomedicine (4th ed.). Springer.
45. Garapati, R. S. (2023). Optimizing Energy Consumption in Smart Build-ings Through Web-Integrated AI and Cloud-Driven Control Systems.
46. Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2018). Deep learning for healthcare. Briefings in Bioinformatics, 19(6), 1236–1246.
47. Kushvanth Chowdary Nagabhyru. (2023). Accelerating Digital Transformation with AI Driven Data Engineering: Industry Case Studies from Cloud and IoT Domains. Educational Administration: Theory and Practice, 29(4), 5898–5910. https://doi.org/10.53555/kuey.v29i4.10932
48. Ahmed, M., Mahmood, A., & Hu, J. (2016). A survey of network anomaly detection techniques. Journal of Network and Computer Applications, 60, 19–31.
49. Guntupalli, R. (2023). Optimizing Cloud Infrastructure Performance Using AI: Intelligent Resource Allocation and Predictive Maintenance. Available at SSRN 5329154.
50. Patcha, A., & Park, J. M. (2007). An overview of anomaly detection techniques. Computer Networks, 51(12), 3448–3470.
51. Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn. Journal of Machine Learning Research, 12, 2825–2830.
52. Aitha, A. R. (2023). CloudBased Microservices Architecture for Seamless Insurance Policy Administration. International Journal of Finance (IJFIN)-ABDC Journal Quality List, 36(6), 607-632.
53. Rajkomar, A., Oren, E., Chen, K., et al. (2018). Scalable deep learning with EHRs. NPJ Digital Medicine, 1, 18.
54. Avinash Reddy Segireddy. (2022). Terraform and Ansible in Building Resilient Cloud-Native Payment Architectures. International Journal of Intelligent Systems and Applications in Engineering, 10(3s), 444–455. Retrieved from https://www.ijisae.org/index.php/IJISAE/article/view/7905.
55. Ringberg, H., Soule, A., Rexford, J., & Diot, C. (2007). Sensitivity of PCA for anomaly detection. SIGMETRICS Proceedings.

56. Koppolu, H. K. R., Sheelam, G. K., & Komaragiri, V. B. (2023). Autonomous Telecommunication Networks: The Convergence of Agentic AI and AI-Optimized Hardware. International Journal of Science and Research (IJSR), 12(12), 2253-2270.
57. Ruff, L., Vandermeulen, R. A., Görnitz, N., et al. (2018). Deep one-class classification. ICML Proceedings.
58. Rongali, S. K. (2023). Explainable Artificial Intelligence (XAI) Framework for Transparent Clinical Decision Support Systems. International Journal of Medical Toxicology and Legal Medicine, 26(3), 22-31.
59. Salfner, F., Lenk, M., & Malek, M. (2010). Survey of failure prediction methods. ACM Computing Surveys, 42(3), 1–42.
60. Nagubandi, A. R. (2023). Advanced Multi-Agent AI Systems for Autonomous Reconciliation Across Enterprise Multi-Counterparty Derivatives, Collateral, and Accounting Platforms. International Journal of Finance (IJFIN)-ABDC Journal Quality List, 36(6), 653-674.
61. Schölkopf, B., Platt, J. C., Shawe-Taylor, J., et al. (2001). Estimating the support of a high-dimensional distribution. Neural Computation, 13(7), 1443–1471.
62. Kalisetty, S., & Ganti, V. K. A. T. (2019). Transforming the Retail Landscape: Srinivas's Vision for Integrating Advanced Technologies in Supply Chain Efficiency and Customer Experience. Online Journal of Materials Science, 1, 1254.
63. Sipos, R., Fradkin, D., Moerchen, F., & Wang, Z. (2014). Log-based predictive maintenance. KDD Proceedings.
64. Meda, R. (2023). Intelligent Infrastructure for Real-Time Inventory and Logistics in Retail Supply Chains. Educational Administration: Theory and Practice.
65. Kolla, S. K. (2021). Designing Scalable Healthcare Data Pipelines for Multi-Hospital Networks. World Journal of Clinical Medicine Research, 1(1), 1–14. Retrieved from https://www.scipublications.com/journal/index.php/wjcmr/article/view/1376
66. Bandi, V. D. V. K. (2023). Cloud-Native Model Lifecycle Management for Enterprise AI Systems. International Journal of Scientific Research and Modern Technology, 2(12), 78–90. https://doi.org/10.38124/ijsrmt.v2i12.1236
67. Inala, R. Revolutionizing Customer Master Data in Insurance Technology Platforms: An AI and MDM Architecture Perspective.
68. Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society B, 58(1), 267–288.
69. Gottimukkala, V. R. R. (2023). Privacy-Preserving Machine Learning Models for Transaction Monitoring in Global Banking Networks. International Journal of Finance (IJFIN)-ABDC Journal Quality List, 36(6), 633-652.
70. Beam, A., & Kohane, I. (2018). Big data and machine learning in healthcare. JAMA, 319(13), 1317–1318.
71. AI Powered Fraud Detection Systems: Enhancing Risk Assessment in the Insurance Sector. (2023). American Journal of Analytics and Artificial Intelligence (ajaai) With ISSN 3067-283X, 1(1). https://ajaai.com/index.php/ajaai/article/view/14
72. Weber, G. M., Mandl, K. D., & Kohane, I. S. (2014). Finding the missing link for big biomedical data. JAMIA, 21(1), 1–3.
73. Kolla, S. H. (2021). Rule-Based Automation for IT Service Management Workflows. Online Journal of Engineering Sciences, 1(1), 1–14. Retrieved from https://www.scipublications.com/journal/index.php/ojes/article/view/1360
74. Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., et al. (2016). FAIR Guiding Principles. Scientific Data, 3, 160018.
75. Zhang, Y., & Yang, Q. (2021). A survey on multi-task learning. IEEE Transactions on Knowledge and Data Engineering, 34(12), 5586–5609.
76. Meda, R. (2023). Developing AI-Powered Virtual Color Consultation Tools for Retail and Professional Customers. Journal for ReAttach Therapy and Developmental Diversities. https://doi.org/10.53555/jrtdd. v6i10s (2), 3577.
77. Almadhoun, R., Kadadha, M., Al-Fuqaha, A., & Guizani, M. (2021). A user-centric blockchain-based system for incident response in the era of IoT. Internet of Things, 14, 100371. https://doi.org/10.1016/j.iot.2021.100371
78. Kalisetty, S. (2023). The Role of Circular Supply Chains in Achieving Sustainability Goals: A 2023 Perspective on Recycling, Reuse, and Resource Optimization. Reuse, and Resource Optimization (June 15, 2023).
79. Bauder, R., & Khoshgoftaar, T. (2018). Medicare fraud detection using machine learning methods. Journal of Big Data, 5(9), 1–21.
80. Siva Hemanth Kolla. (2022). Knowledge Retrieval Systems for Enterprise Service Environments. International Journal of Intelligent Systems and Applications in Engineering, 10(3s), 495–506. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/8037
81. Bishop, C. M. (1994). Novelty detection and neural network validation. IEE Proceedings, 141(4), 217–222.
82. Gottimukkala, V. R. R. (2022). Licensing Innovation in the Financial Messaging Ecosystem: Business Models and Global Compliance Impact. International Journal of Scientific Research and Modern Technology, 1(12), 177-186.
83. Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys, 41(3), 1–58.
84. Bolton, R., & Hand, D. (2002). Statistical fraud detection: A review. Statistical Science, 17(3), 235–255.
85. Amistapuram, K. (2022). Fraud Detection and Risk Modeling in Insurance: Early Adoption of Machine Learning in Claims Processing. Available at SSRN 5741982.
86. Kumar, A., Gupta, P., & Singh, R. (2023). Sentiment analysis methods for proactive brand reputation risk management. International Journal of Information Management Data Insights, 3(1).

87. Ramesh Inala. (2023). Big Data Architectures for Modernizing Customer Master Systems in Group Insurance and Retirement Planning. Educational Administration: Theory and Practice, 29(4), 5493–5505. https://doi.org/10.53555/kuey.v29i4.10424
88. Aggarwal, C. C. (2017). Outlier analysis (2nd ed.). Springer.
89. Davuluri, P. N. AI-Augmented Sanctions Screening: Enhancing Accuracy and Latency in Real Time Compliance Systems.
90. Bifet, A., & Gavalda, R. (2007). Learning from time-changing data with adaptive windowing. SDM Proceedings.
91. Uday Surendra Yandamuri. (2023). An Intelligent Analytics Framework Combining Big Data and Machine Learning for Business Forecasting. International Journal Of Finance, 36(6), 682-706. https://doi.org/10.5281/zenodo.18095256
92. Zaharia, M., Chowdhury, M., Franklin, M. J., et al. (2010). Spark: Cluster computing. HotCloud Proceedings.
93. Avinash Reddy Aitha. (2022). Deep Neural Networks for Property Risk Prediction Leveraging Aerial and Satellite Imaging. International Journal of Communication Networks and Information Security (IJCNIS), 14(3), 1308–1318. Retrieved from https://www.ijcnis.org/index.php/ijcnis/article/view/8609
94. Goutham Kumar Sheelam, Hara Krishna Reddy Koppolu. (2022). Data Engineering And Analytics For 5G-Driven Customer Experience In Telecom, Media, And Healthcare. Migration Letters, 19(S2), 1920–1944. Retrieved from https://migrationletters.com/index.php/ml/article/view/11938
95. Alenezi, M., & Akour, M. AI-driven innovations in software engineering: A review of current practices and future directions. Applied Sciences, 15(3), 1344. https://doi.org/10.3390/app15031344 Cited by: 149
96. Varri, D. B. S. (2022). A Framework for Cloud-Integrated Database Hardening in Hybrid AWS-Azure Environments: Security Posture Automation Through Wiz-Driven Insights. International Journal of Scientific Research and Modern Technology, 1(12), 216-226.
97. Kose, I., & Gunes, S. (2018). Anomaly detection in healthcare data using machine learning techniques. Expert Systems with Applications, 98, 243–253.
98. Meda, R. (2023). Data Engineering Architectures for Scalable AI in Paint Manufacturing Operations. European Data Science Journal (EDSJ) p-ISSN 3050-9572 en e-ISSN 3050-9580, 1(1). Phua, C., Lee, V., Smith, K., & Gayler, R. (2010). A comprehensive survey of data mining-based fraud detection research. arXiv preprint arXiv:1009.6119.
99. Rongali, S. K. (2022). AI-Driven Automation in Healthcare Claims and EHR Processing Using MuleSoft and Machine Learning Pipelines. Available at SSRN 5763022.