

Data-Driven Approaches to Fraud Detection in Health Insurance Claims: A Systematic Review of Medical and Pharmaceutical Services

Sri Rukayah^{1*}, Susi Ari Kristina²

¹²Universitas Gadjah Mada, Yogyakarta, Indonesia

Abstract

Fraud in health insurance claims continues to impose significant financial and operational burdens on healthcare systems, especially as the volume and complexity of claims increase. Conventional rule-based detection mechanisms, although widely used, have limited adaptability to evolving fraud patterns and high-dimensional data environments. This limitation has driven a shift toward data-driven analytical approaches capable of identifying suspicious patterns more effectively. This systematic review synthesizes peer-reviewed, open-access studies published between 2020 and 2025 that applied rule-based, supervised, unsupervised, or hybrid methods for fraud detection in health insurance claims. A comprehensive search across major databases yielded fourteen eligible studies representing diverse systems, datasets, and methodological designs. The findings indicate a clear transition from traditional rule-based systems to machine learning approaches, particularly in addressing challenges such as label scarcity, class imbalance, and complex fraud patterns. Most studies focused on integrated medical claims, where pharmaceutical fraud was embedded rather than analyzed independently, highlighting a gap in service-specific research. Significant heterogeneity was observed in fraud definitions, preprocessing techniques, labeling strategies, and evaluation metrics, limiting cross-study comparability and emphasizing the need for greater methodological transparency. Across the literature, data-driven approaches are consistently positioned as decision-support tools rather than definitive solutions, reinforcing their role in complementing expert judgment and regulatory oversight. Overall, effective implementation requires context-aware design, reliable labeling, and rigorous real-world validation. Future research should prioritize domain-specific analyses, particularly in pharmaceutical fraud, and improve transparency to support scalable and responsible deployment.

KEYWORDS

health insurance fraud detection; machine learning approaches; anomaly detection; healthcare claims analytics; medical and pharmaceutical services; systematic literature review.

Introduction

Fraud detection has therefore become a critical priority not only in healthcare reimbursement systems but also in broader financial and administrative domains, where intelligent analytical methods are increasingly required (1,2). The challenge is further compounded by the increasing scale, complexity, and imbalanced class distributions in claims datasets, which limit the effectiveness of conventional auditing approaches and motivate the adoption of advanced analytical methods (3). In claims-based reimbursement systems, fraudulent practices may occur across

both medical and pharmaceutical services, including upcoding, billing for non-rendered services, duplicate claims, and inappropriate prescription practices. The increasing volume, velocity, and complexity of health insurance claims data have further intensified the difficulty of identifying such irregularities through conventional manual audits alone (4,5).

In the pharmaceutical domain, fraud-related risks are further compounded by the persistent circulation of substandard and counterfeit medicines, which undermine treatment effectiveness and patient safety, particularly in low- and middle-income settings (6). In addition, pharmaceutical fraud risks extend beyond inappropriate billing practices, as the circulation of falsified and substandard medicines continues to represent a major threat to patient safety and health system sustainability (7). The World Health Organization has also emphasized that substandard and falsified medicines constitute a serious global public health concern, threatening health system integrity and increasing the risk of inappropriate pharmaceutical reimbursement and misuse within insurance systems (8).

Within this broader context of healthcare fraud, traditional fraud detection in health insurance has relied primarily on rule-based approaches, expert judgment, and post-payment review mechanisms. While these approaches remain valuable for identifying well-defined and previously known fraud patterns, they are often constrained by limited flexibility, delayed detection, and reduced adaptability to emerging fraudulent behaviors. As claims datasets become larger and more heterogeneous, these limitations have driven a gradual shift toward data-driven analytical approaches capable of processing high-dimensional information and identifying suspicious patterns at scale (9).

Recent advances in data analytics, machine learning, and data mining have accelerated the adoption of data-driven approaches for fraud detection in health insurance claims. These approaches include supervised learning models trained on labeled fraud cases, unsupervised and semi-supervised methods designed to detect anomalous patterns in the absence of reliable ground truth, as well as hybrid frameworks that integrate rule-based logic with statistical or machine learning techniques. Such methods have been applied across diverse datasets, including national health insurance claims, hospital billing records, and simulated claims data, reflecting variations in data availability, regulatory environments, and system maturity across different healthcare settings (10,11).

Despite the growing body of literature, studies on fraud detection in health insurance claims remain highly heterogeneous with respect to data sources, analytical methods, fraud definitions, and evaluation metrics. Moreover, many studies examine medical and pharmaceutical claims jointly, with pharmaceutical fraud frequently embedded within broader medical billing analyses rather than addressed as a distinct analytical domain. This heterogeneity poses substantial challenges for drawing direct comparisons of algorithmic performance or determining the relative effectiveness of specific analytical techniques across contexts (5,10).

Systematic literature reviews (SLRs) play a critical role in synthesizing fragmented evidence, identifying methodological trends, and highlighting gaps in existing research. By applying transparent and reproducible review procedures, SLRs can provide a comprehensive overview of how data-driven fraud detection approaches have been conceptualized and implemented within

health insurance systems. Accordingly, the present review was conducted in accordance with established methodological standards for systematic reviews to ensure rigor and transparency throughout the processes of study identification, selection, and synthesis (12).

This systematic literature review aims to synthesize evidence on data-driven approaches to fraud detection in health insurance claims, with particular attention to applications in medical and pharmaceutical services. Rather than ranking analytical methods based on heterogeneous and non-comparable performance metrics, this review focuses on mapping analytical approaches, data characteristics, and implementation contexts. By doing so, the study seeks to inform researchers, policymakers, and insurance administrators regarding current methodological practices, implementation readiness, and key research gaps that must be addressed to support effective, transparent, and responsible fraud detection in health insurance systems.

Methods

Study design

Across the unsupervised subset, studies consistently identified provider-level anomalies without requiring labeled fraud outcomes. For example, a two-step clustering pipeline grouped comparable providers and then flagged outliers whose DRG coding patterns deviated from peer norms, enabling targeted auditor review of hospitals with atypical behavior (13). At the national scale, fully label-free methods on Medicare claims detected overbilling signals and returned human-interpretable rationales at the provider level, which were cross-validated against enforcement actions, demonstrating the viability of unsupervised detection as a front-end triage layer for investigators (14).

This study was conducted as a systematic literature review (SLR) to synthesize evidence on data-driven approaches to fraud detection in health insurance claims related to medical and pharmaceutical services. The review was designed and reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines to ensure transparency, reproducibility, and methodological rigor throughout the review process (12).

Search Strategy

Study selection and data extraction emphasized method transparency (preprocessing, feature engineering, evaluation metrics) and reproducibility of analytical steps. In the unsupervised stream, we looked for peer group or clustering based strategies capable of flagging outlying provider behavior such as two step clustering (k means plus peer comparison) that has been shown to surface anomalous DRG coding patterns for auditor follow up (13). We likewise tracked hybrid architectures combining association rule mining with anomaly detectors because they operationalize domain rules while preserving flexibility to detect emergent patterns; recent work on CMS DE SynPUF exemplifies this two stage design and reports competitive efficiency with clear unsupervised scoring outputs (5). Collectively, these methodological criteria allowed us to summarize approaches that balance interpretability, scalability, and governance readiness across heterogeneous data conditions (15).

A comprehensive literature search was performed across multiple electronic databases to identify relevant peer-reviewed studies on fraud detection in health insurance claims. The databases included Scopus, PubMed, Google Scholar, Semantic Scholar, and

ScienceDirect, supplemented by hand searching.

Searches were conducted using combinations of keywords and Boolean operators related to health insurance, claims, fraud detection, medical services, pharmaceutical services, data mining, and machine learning. The search strategy was designed to capture studies employing both rule-based and data-driven analytical approaches.

To ensure the relevance and currency of the evidence, the search was limited to articles published between 2020 and 2025. Only articles published in English and available in full-text open-access format were considered eligible for inclusion.

Eligibility Criteria

Studies were included in the review if they met the following criteria:

1. addressed fraud detection or identification in health insurance or healthcare claims;
2. applied data-driven, analytical, or computational methods, including rule-based systems, data mining techniques, or machine learning models;
3. utilized data derived from medical claims, pharmaceutical claims, or integrated healthcare billing datasets, including both real-world and simulated claims data; and
4. were original research articles published in peer-reviewed journals.

Studies were excluded if they focused solely on non-insurance healthcare fraud, did not employ analytical or data-driven approaches, were review articles, editorials, conference abstracts, or lacked sufficient methodological detail. Studies using simulated datasets were retained in the review because they provide methodological insights and proof-of-concept evaluations for fraud detection models; however, their findings were interpreted separately from studies based on real-world insurance claims to maintain contextual relevance.

Study Selection

All records identified through the database searches were exported to reference management software, and duplicate entries were removed. No eligible records were retrieved from PubMed, likely due to differences in indexing scope and search term sensitivity for insurance fraud-related studies. The study selection process was conducted in two stages. First, titles and abstracts were screened to assess initial relevance based on the predefined eligibility criteria. Second, full-text articles were reviewed to confirm eligibility for inclusion in the final synthesis. Any discrepancies during the screening process were resolved through discussion and consensus.

The study selection process is summarized using a PRISMA flow diagram, detailing the number of records identified, screened, excluded, and included at each stage of the review (12).

Data Extraction

Data were systematically extracted from each included study using a standardized extraction framework. The extracted variables included author and year of publication, country or setting, type of insurance system, data source (real-world or simulated claims), type of healthcare service (medical, pharmaceutical, or both), fraud type, analytical approach or algorithm used, and key findings. This structured extraction enabled consistent comparison and thematic synthesis across studies with heterogeneous designs and objectives.

Quality Assessment

The methodological quality of the included studies was assessed using the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) checklist to support interpretation of findings rather than as exclusion criteria. The STROBE framework was selected because the majority of included studies employed observational designs or secondary analyses of claims data. Each study was evaluated based on clarity of objectives, data description, analytical methods, and reporting transparency. The quality assessment was used to inform the interpretation of findings rather than to exclude studies, thereby preserving the breadth of evidence relevant to data-driven fraud detection (16).

Data Synthesis

Given the heterogeneity of study designs, data sources, fraud definitions, and evaluation metrics, a quantitative meta-analysis was not performed. Instead, a narrative synthesis approach was adopted to summarize and integrate findings across studies. The synthesis focused on mapping data-driven approaches, analytical tools, application contexts, and levels of implementation readiness. This approach allowed for a structured comparison of methodological trends and research gaps without imposing inappropriate quantitative comparisons across non-comparable studies.

Result and Discussion

Study Selection

The systematic search across selected databases identified a total number of records relevant to fraud detection in health insurance claims. After removing duplicate records, titles and abstracts were screened based on the predefined eligibility criteria. Studies that did not address fraud detection in insurance claims, did not apply data-driven or analytical approaches, or lacked sufficient methodological detail were excluded. Following full-text screening, a final set of studies was included in the review. The detailed study selection process is presented in the PRISMA flow diagram.

The study selection process adhered to the PRISMA 2020 guidelines. A systematic search across five electronic databases: Google Scholar ($n = 621$), Scopus ($n = 101$), PubMed ($n = 0$), Semantic Scholar ($n = 316$), ScienceDirect ($n = 15$), and Hand search ($n = 9$) yielded a total of 1,062 records. After removing 188 duplicate records using automation tools, 874 records remained for title and abstract screening.

During the initial screening phase, 115 records were excluded due to document type: books ($n = 4$), book sections ($n = 27$), conference papers ($n = 72$), documents ($n = 1$), preprints ($n = 5$), and theses ($n = 6$). This resulted in 759 reports being sought for full-text retrieval.

Of these, 741 reports were excluded for the following reasons: publication year prior to 2020 ($n = 119$), literature reviews ($n = 175$), and titles not relevant to the research topic ($n = 447$). Eighteen reports were assessed for eligibility, but four were excluded due to a lack of full-text access. Ultimately, fourteen studies met the inclusion criteria and were included in the final systematic review. No additional reports were identified that corresponded to the included studies (see [Figure 1](#)).

Characteristics of Included Studies

Fourteen studies met the inclusion criteria and were synthesized in this review, focused on fraud detection in health insurance claims related to medical and

pharmaceutical services. The analysis reveals several key characteristics across geographical, methodological, and thematic dimensions (See [Table 1](#)).

The included studies exhibited substantial heterogeneity in terms of geographical setting, insurance system characteristics, data sources, and analytical approaches. Most studies were conducted in contexts with national or large-scale health insurance systems, reflecting the availability of centralized claims data suitable for large-scale analytical processing. The datasets analyzed across studies consisted of medical claims, pharmaceutical claims, or integrated healthcare billing records, with varying levels of data granularity.

Regarding data sources, the majority of studies utilized real-world insurance claims data obtained from national health insurance agencies, hospitals, or insurance providers. A smaller subset of studies employed simulated or semi-synthetic datasets to evaluate methodological feasibility, particularly in settings where access to labeled fraud data was limited. As summarized in the extraction table, studies based on real-world data tended to focus on operational fraud detection challenges, whereas simulation-based studies emphasized methodological development and algorithm testing.

Overall STROBE assessment indicated adequate reporting of objectives, data sources, and analytical methods across studies, with variations in reporting of bias and generalisability.

Types of Fraud Addressed

Across the reviewed literature, fraud types were defined and operationalized in diverse ways. Commonly addressed fraudulent behaviors included upcoding, billing for non-rendered services, duplicate claims, excessive service utilization, and inappropriate prescription practices. In many studies, pharmaceutical fraud was not treated as a standalone category but was embedded within broader analyses of medical billing fraud. This integration reflects real-world claims environments, where pharmaceutical and medical services are often processed within unified reimbursement systems.

Several studies explicitly acknowledged the difficulty of establishing standardized fraud labels due to variability in regulatory definitions, investigation outcomes, and institutional practices. Consequently, fraud identification was frequently framed as an anomaly detection or risk scoring problem rather than a definitive classification task.

Data-Driven Analytical Approaches

The reviewed studies applied a wide range of data-driven analytical approaches to detect potential fraud in health insurance claims. As shown in the extraction [table 2](#), these approaches can be broadly categorized into rule-based systems, supervised machine learning models, unsupervised or semi-supervised methods, and hybrid frameworks.

Across the unsupervised subset, studies consistently identified provider level anomalies without requiring labeled fraud outcomes. For example, a two step clustering pipeline grouped comparable providers and then flagged outliers whose DRG coding patterns deviated from peer norms, enabling targeted auditor review of hospitals with atypical behavior (Massi et al., 2020). At national scale, fully label free methods on Medicare claims detected overbilling signals and returned human interpretable rationales at the provider level, which were cross validated against enforcement actions demonstrating the viability of unsupervised detection as a front end triage layer for investigators (14)

Hybrid pipelines that integrate association rule mining

with anomaly detectors reported favorable computational efficiency and clear scoring outputs; on CMS DE SynPUF data, combining frequent pattern rules with algorithms such as CBLOF achieved the best silhouette score among tested detectors, illustrating how domain rules can be fused with data driven ranking for scalable triage (5). Meanwhile, supervised models trained on labeled insurance data random forests, logistic regression, and neural networks achieved high accuracy and F1 scores in operational datasets, yet authors emphasized the need for validation on larger, more diverse cohorts to ensure generalizability across payers and regions (17).

Evaluation Strategies and Outcome Reporting

Evaluation strategies differed markedly across studies, as reflected in the extraction table. Some studies reported conventional performance metrics such as accuracy, precision, recall, or area under the curve, while others emphasized qualitative assessments, expert validation, or case-based analysis. Due to differences in datasets, fraud definitions, and evaluation frameworks, direct comparison of model performance across studies was not feasible.

Importantly, several studies highlighted that fraud detection systems are primarily intended as decision-support tools rather than definitive adjudication mechanisms. As such, the outputs of data-driven models were often used to prioritize claims for further investigation rather than to directly classify claims as fraudulent or non-fraudulent.

Summary of Key Findings

The STROBE-based quality assessment indicated that most included studies adequately reported study objectives, data sources, and analytical methods; however, variations were observed in the reporting of data preprocessing procedures, fraud labeling strategies, and model validation approaches. These variations were taken into account during the interpretation of findings rather than used as exclusion criteria. Overall, the results show that data-driven approaches to fraud detection in health insurance claims are widely applied across medical and pharmaceutical services, with increasing methodological sophistication over time. The extracted evidence highlights a clear shift from purely rule-based systems toward more flexible analytical frameworks capable of handling large and complex claims datasets, although the heterogeneity of data sources, fraud definitions, and evaluation methods limits the ability to draw generalized conclusions regarding the relative effectiveness of specific analytical techniques.

This systematic literature review provides a structured synthesis of data-driven approaches applied to fraud detection in health insurance claims for medical and pharmaceutical services. In line with the issues outlined in the Introduction, this review does not attempt to rank analytical methods based on heterogeneous and non-comparable performance metrics. Instead, it emphasizes the mapping of methodological trends, data characteristics, and application contexts across the included studies. This approach aligns with prior work showing that the effectiveness of fraud detection analytics is highly context dependent and shaped by data availability, system design, and governance/interpretability requirements in regulated settings (18,19). Contemporary open access reviews likewise stress the absence of robust cross technique benchmarks and the variability of model behavior across settings, reinforcing the need for cautious, context-aware interpretation (15,20).

A prominent trend identified in this review is the gradual transition from traditional rule-based detection mechanisms toward more flexible data-driven and hybrid analytical

frameworks. Rule-based systems continue to serve as important initial screening tools due to their transparency, interpretability, and alignment with regulatory requirements. However, consistent with findings reported in prior methodological and empirical studies, such approaches alone are insufficient to address evolving fraud patterns and high-dimensional claims data. This limitation has motivated the development of hybrid frameworks that integrate expert-defined rules with machine learning techniques to enhance adaptability while preserving interpretability (5,21). Unsupervised, provider level anomaly detection has also been shown to flag up coding behavior in administrative hospital data, providing auditor ready shortlists that complement rule based screens (13). At payer scale, label free pipelines have identified overbilling signals in Medicare claims and returned interpretable rationales at the provider level, illustrating how unsupervised components strengthen hybrid detection stacks.

Empirical applications of supervised machine learning models further demonstrate how algorithm-driven classification techniques can improve the detection of complex fraud patterns in medical insurance claims, particularly when combined with structured feature engineering and validation strategies ((22). Consequently, supervised, unsupervised, and hybrid machine learning methods have been increasingly adopted to enhance scalability, pattern recognition, and analytical capability in health insurance fraud detection (23,24).

Persistent challenges related to data availability and fraud labeling were also evident across the reviewed studies. Supervised learning approaches frequently relied on historical investigation outcomes or expert judgment to label fraudulent claims; however, the quality, consistency, and completeness of such labels varied substantially across settings. In response to limited or unreliable labeled data, several studies employed unsupervised or semi-supervised techniques, reframing fraud detection as an anomaly detection problem. Recent methods tackle label scarcity by generating proxy fraud labels in a fully unsupervised manner and by selecting informative features prior to labelling approaches that directly address severe class imbalance typical of claims data (25,26). While these approaches offer greater flexibility in data-scarce environments, they also introduce challenges related to model interpretability, validation, and operational trust, particularly when deployed within real-world insurance systems (24).

With respect to service domains, pharmaceutical fraud was rarely examined as a standalone analytical focus and was more commonly embedded within integrated medical

claims analyses. This pattern reflects real-world reimbursement practices, where pharmaceutical and medical services are processed within unified claims systems. However, similar to observations reported in previous studies, such integration may obscure service-specific fraud patterns, particularly those related to prescribing behavior, medication utilization, and pharmacy-level practices (5). The limited number of studies explicitly disaggregating pharmaceutical fraud, therefore, represents an important gap in the existing literature. Recent evidence on recalls within the legitimate supply chain further underscores domain-specific risks in pharmaceuticals, strengthening the case for dedicated analytics beyond integrated medical claims (27). This observation aligns with medicine quality scholarship that documents persistent substandard and falsified products and substantial methodological heterogeneity in sampling studies, underscoring the need for domain-specific analytics in pharmaceutical claims (28).

From a practical and policy perspective, the review supports the positioning of data-driven fraud detection systems as decision-support tools that complement, rather than replace, human investigation and regulatory oversight. In regulated health insurance environments, hybrid analytical systems that combine rule-based logic with adaptive machine learning models appear to offer a pragmatic balance between interpretability, regulatory compliance, and analytical performance. Concurrently, a 2025 governance framework emphasizes algorithmic oversight, risk stewardship, and equity safeguards as preconditions for deploying fraud analytics in regulated insurance environments (29). In regulated settings, explainability remains a prerequisite for deployment; medical AI surveys recommend governance-aware, interpretable pipelines for adjudication workflows (19). Effective implementation of such systems requires not only technical capacity but also robust governance frameworks addressing data quality, transparency, accountability, and ethical considerations, issues that were also emphasized in the Introduction.

Overall, the Discussion reinforces the central argument of this review: that data-driven approaches hold substantial promise for strengthening fraud detection in health insurance systems, yet their effectiveness remains highly dependent on contextual, methodological, and governance-related factors. By systematically mapping existing approaches rather than prioritizing performance comparisons across heterogeneous datasets, this review provides a coherent synthesis that aligns methodological innovation with practical implementation needs across medical and pharmaceutical services.

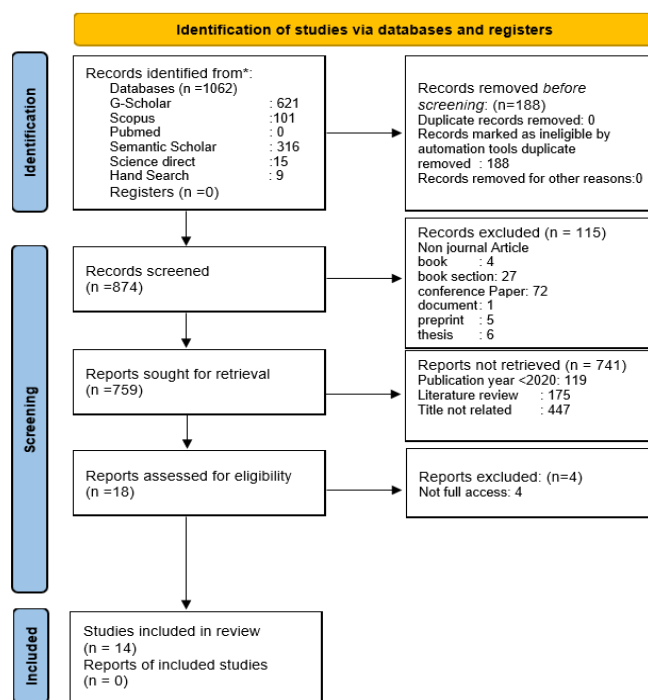


Figure 1. PRISMA Flow Diagram

Table 1. Characteristics of Included Studies (n = 14)

Characteristic	Number of Studies	Percentage
Study Setting		
High-income countries	4	29%
Upper-middle-income countries	1	7%
Lower-middle-income countries	9	64%
First Author Affiliation		
Academic institutions	13	93%
Independent researcher	1	7%
Funding Source		
Not stated in the article	12	86%
Rabdan Academy, Abu Dhabi	1	7%
Princess Nourah bint Abdulrahman University	1	7%
Data Source		
Healthcare institution claims data	8	57%
Hospital insurance claims data	3	21%
Simulated claims data	3	21%
Analytical Approach for Fraud Detection		
Rule-based or non-machine learning analytical methods	5	36%
Supervised machine learning methods	5	36%
Unsupervised or anomaly detection methods	2	14%
Hybrid or combined analytical approaches	2	14%
Year of Publication		
2020	3	21%
2021	0	0%
2022	1	7%
2023	1	7%
2024	5	36%
2025	4	29%

Table 2. Summary of Analytical Approaches Used in Included Studies

No.	Author (Year)	Article Title	Data Source	Service Domain	Analytical Approach	Tools / Techniques	Data Type
1	Kotekani & Velchamy (2020)	<i>An Effective Data Sampling Procedure for Imbalanced Data Learning on Health Insurance Fraud Detection</i>	Health insurance claims database	Medical services	Supervised machine learning (imbalanced learning)	Effective data sampling procedures with classification models	Real-world claims
2	Matloob et al., (2020)	<i>A Framework to Detect Healthcare Frauds Using Association Rules and a Threshold-Based Rule Engine</i>	Hospital insurance claims	Medical & pharmaceutical services	Rule-based and data mining approach	Association rules and threshold-based rule engine	Real-world claims
3	Matloob et al. (2020)	<i>Medical Health Benefit Management System for Real-Time Notification of Fraud Using Historical Medical Records</i>	Historical medical records	Medical services	Predictive and rule-based framework	Real-time fraud notification system	Real-world claims
4	Matloob et al. (2022)	<i>A Sequence Mining-Based Novel Architecture for Detecting Fraudulent Transactions in Healthcare Systems</i>	Hospital insurance claims	Medical & pharmaceutical services	Hybrid fraud detection framework	Sequence mining integrated with machine learning	Real-world claims
5	Matloob et al. (2025)	<i>Healthcare Fraud Detection Using Adaptive Learning and Deep Learning Techniques</i>	Healthcare insurance database	Medical services	Adaptive and deep learning approach	Deep learning with adaptive learning mechanisms	Real-world claims
6	Khanizadeh et al., (2025)	<i>Smart Data-Driven Medical Decisions Through Collective and Individual Anomaly Detection in Healthcare Time Series</i>	Health insurance claims database	Medical services	Unsupervised and semi-supervised learning	Anomaly detection and clustering techniques	Real-world claims
7	Agarwal, (2023)	<i>Detection of Fraudulent Activities in Health Insurance Using Heterogeneous Information Network</i>	Health insurance claims dataset	Medical & pharmaceutical services	Supervised machine learning	Heterogeneous information network-based modeling	Simulated data
8	Hamid et al. (2024)	<i>Healthcare Insurance Fraud Detection Using Data Mining</i>	Healthcare insurance claims	Medical services	Data mining approach	Classification and pattern recognition techniques	Real-world claims

No.	Author (Year)	Article Title	Data Source	Service Domain	Analytical Approach	Tools / Techniques	Data Type
9	Sunilram et al., (2024)	<i>Identifying Health Insurance Claims Frauds Using Machine Learning Concepts</i>	Insurance claims data	Medical services	Machine learning–based detection	Decision tree, random forest, and logistic regression	Simulated data
10	Aarthi et al., (2025)	<i>Leveraging Machine Learning for Improved Detection of Medicare Fraud</i>	Medicare insurance claims	Medical services	Supervised machine learning	Machine learning models for Medicare fraud detection	Real-world claims
11	Anwer et al., (2024)	<i>A Comprehensive Study of Healthcare Fraud Detection Using Machine Learning</i>	Healthcare insurance claims	Medical services	Supervised machine learning	Comparative machine learning algorithms	Real-world claims
12	Subbarayudu et al., (2024)	<i>The Evaluation of Distributed Topic Modeling Paradigms for Detection of Fraudulent Insurance Claims in Healthcare Forum</i>	Health insurance claims forum data	Medical services	Data mining and topic modeling	Distributed topic modeling paradigms	Simulated data
13	Wang et al., (2025)	<i>A Robust and Interpretable Ensemble Machine Learning Model for Predicting Healthcare Insurance Fraud</i>	Healthcare insurance claims	Medical services	Interpretable ensemble machine learning	Robust and interpretable ensemble ML models	Real-world claims
14	Cherkaoui et al., (2024)	<i>Artificial Intelligence–Based Approach for Healthcare Insurance Fraud Detection</i>	Health insurance claims data	Medical services	Artificial intelligence–based fraud detection	AI-driven classification and anomaly detection models	Real-world claims

At the same time, the collective evidence highlights persistent challenges that shape real-world applicability, including inconsistent fraud definitions, limited and uneven labeling practices, variability in preprocessing and validation strategies, and the scarcity of service-specific analyses for pharmaceutical claims. These factors underscore the importance of context-aware analytical design and transparent reporting standards. Hybrid frameworks that integrate expert-defined rules with machine learning based anomaly or classification models appear particularly pragmatic for regulated insurance environments, combining interpretability with adaptive detection capacity.

In operational settings, data-driven systems are best positioned as decision support tools rather than automated adjudication mechanisms. Their responsible deployment requires minimum governance safeguards such as explainability, auditability, bias monitoring, and performance validation that reflects investigator relevant outcomes. Strengthening fraud labeling through human-in-the-loop workflows, adopting standardized reporting checklists, and conducting external or temporal validation

are essential to improving comparability and trustworthiness. With these considerations, hybrid and context-aware analytical approaches can support more scalable, accountable, and effective fraud detection capabilities within diverse health insurance systems.

Comparison with Previous Studies

The findings of this review are consistent with earlier studies and reviews that emphasize the limitations of purely rule-based fraud detection systems and the growing adoption of data-driven analytical approaches in health insurance contexts. Similar to previous literature, this review identifies a shift toward machine learning-based methods to address increasing data complexity and evolving fraud schemes.

Compared with prior reviews that focused on specific algorithms or isolated healthcare settings, this review offers a broader synthesis encompassing both medical and pharmaceutical claims across diverse insurance systems. Nevertheless, as reported in earlier studies, substantial heterogeneity in data sources, fraud definitions, and evaluation metrics persists, limiting cross-study

comparability and reinforcing the context-dependent nature of analytical performance.

Limitations and Cautions

Several limitations of this systematic literature review should be acknowledged. First, substantial heterogeneity across the included studies in terms of data sources, analytical methods, fraud definitions, and evaluation strategies limited the feasibility of quantitative synthesis or meta-analysis. Similar challenges have also been reported in previous reviews of health insurance fraud detection, where differences in datasets and validation frameworks constrained direct comparison of analytical performance across studies (10,24). Consequently, the findings of this review are primarily descriptive and methodological, focusing on mapping approaches rather than evaluating comparative model effectiveness.

Second, variations in fraud definitions and labeling practices may affect the generalizability of reported findings. In many studies, fraud identification relied on proxy indicators, historical investigation outcomes, or expert judgment rather than consistently verified fraudulent cases. Prior literature has noted that such inconsistencies in labeling can influence model training, validation, and interpretability, particularly in supervised learning approaches (5,30). This limitation underscores the importance of cautious interpretation when extrapolating findings across different insurance systems and regulatory contexts.

Third, although a comprehensive search strategy was applied, this review was restricted to English-language, open-access publications, which may have resulted in the exclusion of relevant studies published in other languages or indexed in non-open-access sources. Similar constraints have been acknowledged in previous systematic reviews within the healthcare analytics domain (24). These limitations were considered during the interpretation of results and do not diminish the overall methodological insights generated by the review, but they highlight areas where future evidence synthesis may be further expanded.

Recommendations for Future Research

Future research should prioritize the development and adoption of standardized definitions of fraud, along with transparent reporting of data preprocessing procedures, labeling strategies, and validation frameworks, to improve comparability across studies. Previous literature has emphasized that inconsistencies in fraud definitions and data preparation practices remain a major barrier to cumulative knowledge building in health insurance fraud detection research (5,10). Greater use of real-world insurance claims data, supported by rigorous and context-appropriate validation strategies, is also essential to enhance the practical relevance and generalizability of data-driven fraud detection models (24,30).

In addition, future studies should place greater emphasis on model explainability and interpretability, particularly in policy-driven insurance environments where transparency, accountability, and regulatory compliance are critical. Recent studies have highlighted that the adoption of complex machine learning models without sufficient explainability may hinder trust and operational uptake by insurers and regulators (23,24). Comparative evaluations of multiple analytical approaches using shared datasets or benchmark frameworks would further support evidence-based decision-making and methodological advancement in this field.

From a policy and implementation perspective, closer collaboration between researchers, insurance administrators, and regulatory bodies is strongly

recommended. Such collaboration can help ensure that methodological innovations align with operational realities, legal frameworks, and ethical considerations governing health insurance systems. Strengthening these cross-sector partnerships may also facilitate access to high-quality real-world data and support the responsible integration of data-driven fraud detection tools into routine insurance governance and decision-making processes (5).

Conclusion

This systematic review synthesized recent applications of data-driven approaches for detecting fraud in health insurance claims across medical and pharmaceutical services. Evidence shows a consistent transition from purely rule-based screening to supervised, unsupervised, and hybrid frameworks that better accommodate large scale, high dimensional claims data. Rather than supporting a single best technique, the findings indicate that analytical choices must be shaped by data availability, fraud labeling practices, regulatory requirements, and operational constraints.

Pharmaceutical fraud remained mostly embedded within integrated medical claims analyses, underscoring a gap for service-specific modeling that can capture prescribing and dispensing patterns. Substantial heterogeneity in fraud definitions, preprocessing, labeling, and evaluation strategies limited cross-study comparability and points to the need for stronger reporting discipline.

Actionable recommendations: (i) adopt transparent reporting checklists (data provenance, preprocessing, label sources, imbalance handling, and validation design); (ii) strengthen labeling via human in the loop workflows that combine investigation outcomes with expert adjudication; (iii) prefer context aware validation (temporal splits, payer/region external validation) and report decision support metrics relevant to operations (investigator workload lift, alert level PPV, time to detection); (iv) develop pharmaceutical service specific models; and (v) ensure governance with explainability, audit trails, and bias monitoring. In practice, data-driven systems are best positioned as decision support tools that augment expert review and regulatory oversight, enabling more accountable and scalable fraud detection within health insurance systems.

Author contributions

Sri Rukayah conceived the study, designed the systematic review protocol, and conducted the literature search, study selection, data extraction, and synthesis. She drafted the manuscript. Susi Ari Kristina provided methodological oversight, contributed to the interpretation of findings, and critically revised the manuscript for important intellectual content. Both authors approved the final version of the manuscript and accept full responsibility for the integrity and accuracy of the work.

Acknowledgements

The authors gratefully acknowledge the Master's Program in Pharmacy Management, Faculty of Pharmacy, Universitas Gadjah Mada, for the academic guidance and institutional support throughout this research. Appreciation is also extended to all individuals and organizations who contributed to the study,

particularly for their valuable time, insights, and assistance. The authors also recognize the professional experience contributed by one of the co-authors in their role as a claim verifier at BPJS Kesehatan, Indonesia's

national health insurance institution, which provided practical perspectives that enriched the research process without involving any confidential or proprietary data

References

- West J, Bhattacharya M, Islam R. Intelligent Financial Fraud Detection Practices: An Investigation. In: Tian J, Jing J, Srivatsa M, editors. International Conference on Security and Privacy in Communication Networks. Springer; 2015. p. 186–203. (Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering). doi:10.1007/978-3-319-23802-9_16
- Ngai EWT, Hu Y, Wong YH, Chen Y, Sun X. The Application of Data Mining Techniques in Financial Fraud Detection. *Decis Support Syst*. 2011;50(3):559–69. doi:10.1016/j.dss.2010.08.006
- Leevy JL, Khoshgoftaar TM, Bauder RA, Seliya N. A Survey on Addressing High-Class Imbalance in Big Data. *J Big Data*. 2018;5(1):42. doi:10.1186/s40537-018-0151-6
- Agarwal A, Nene MJ. A five-layer framework for AI governance: integrating regulation, standards, and certification. *Transforming Government: People, Process and Policy*. 2025. doi:10.1108/TG-03-2025-0065
- Hamid Z, Khalique F, Mahmood S, Daud A, Bukhari A, Alshemaimri B. Healthcare Insurance Fraud Detection Using Data Mining. *BMC Med Inform Decis Mak*. 2024;24(1):112. doi:10.1186/s12911-024-02512-4
- Almuzaini T, Choonara I, Sammons H. Substandard and Counterfeit Medicines: A Systematic Review of the Literature. *BMJ Open*. 2013;3(8):e002923. doi:10.1136/bmjopen-2013-002923
- Ozawa S, Evans DR, Bessias S, Haynie DG, Yemeke TT, Laing SK, et al. Prevalence and Estimated Economic Burden of Substandard and Falsified Medicines in Low- and Middle-Income Countries. *JAMA Netw Open*. 2018;1(4):e181662. doi:10.1001/jamanetworkopen.2018.1662
- World Health Organization. WHO Global Surveillance and Monitoring System for Substandard and Falsified Medical Products [Internet]. 2017. Available from: <https://iris.who.int/bitstream/handle/10665/339295/WHO-EMP-RHT-SAV-2017.01-eng.pdf>
- Agarwal. Detection of Fraudulent Activities in Health Insurance Using Heterogeneous Information Network. *Comput Intell Neurosci*. 2023;2706928. doi:10.1155/2023/2706928
- Khanzadeh F, Etefaghian A, Wilson G, Shirazibeheshti A, Radwan T, Luca C. Smart Data-Driven Medical Decisions Through Collective and Individual Anomaly Detection in Healthcare Time Series. *Int J Med Inform*. 2025;194. doi:10.1016/j.ijmedinf.2024.105696
- Matloob I, Khan SA, Rukaiya R, Khattak MAK, Munir A. A Sequence Mining-Based Novel Architecture for Detecting Fraudulent Transactions in Healthcare Systems. *IEEE Access*. 2022;10:48447–63. doi:10.1109/ACCESS.2022.3170888
- Page MJ, McKenzie JE, Bossuyt PM, others. The PRISMA 2020 Statement. *Syst Rev*. 2021;10(1):89. doi:10.1186/s13643-021-01626-4
- Massi MC, Ieva F, Lettieri E. Data Mining Application to Healthcare Fraud Detection. *BMC Med Inform Decis Mak*. 2020;20(1):160. doi:10.1186/s12911-020-01143-9
- Shekhar S, Leder-Luis J, Akoglu L. Unsupervised Machine Learning for Explainable Health Care Fraud Detection. *ArXiv*. 2023. doi:10.48550/arXiv.2211.02927
- Razzaq K, Shah M. Next-Generation Machine Learning in Healthcare Fraud Detection. *Information*. 2025;16(9):730. doi:10.3390/info16090730
- Von Elm E et al. The STROBE Statement: Guidelines for Reporting Observational Studies. *PLoS Med*. 2007;4(10):e296. doi:10.1371/journal.pmed.0040296
- Nabrawi E, Alanazi A. Fraud Detection in Healthcare Insurance Claims Using Machine Learning. *Risks*. 2023;11(9):160. doi:10.3390/risks11090160
- Muspratt R, Mammadov M. Anomaly Detection with Sub-Extreme Values: Health Provider Billing. *Data Sci Eng*. 2024;9(1):62–72. doi:10.1007/s41019-023-00234-7
- Tjoa E, Guan C. A Survey on Explainable Artificial Intelligence (XAI). *IEEE Trans Neural Netw Learn Syst*. 2021;32(11):4793–813. doi:10.1109/TNNLS.2020.3027314
- Sabic E, Keeley D, Henderson B, Nannemann S. Healthcare and Anomaly Detection: Using Machine Learning to Predict Anomalies in Heart Rate Data. *AI Soc*. 2021;36(1):149–58. doi:10.1007/s00146-020-00985-1
- Kotekani SS, Velchamy I. An Effective Data Sampling Procedure for Imbalanced Data Learning on Health Insurance Fraud Detection. *Journal of Computing and Information Technology*. 2020;28(4):269–85. doi:10.20532/cit.2020.1005216
- Agarwal. An Intelligent Machine Learning Approach for Fraud Detection in Medical Claim Insurance: A Comprehensive Study. *SJET*. 2023;11(9):191–200. doi:10.36347/sjet.2023.v11i09.003
- Lu et al. Health Insurance Fraud Detection Using an Attributed Heterogeneous Information Network with a Hierarchical Attention Mechanism. *BMC Med Inform Decis Mak*. 2023;23:62. doi:10.1186/s12911-023-02152-0
- Curtis ED, Billion-Polak P, Khoshgoftaar TM, Furht B. A Review of Distinct Machine Learning Classifiers for Healthcare Fraud Detection. *J Big Data*. 2025;12(1):238. doi:10.1186/s40537-025-01295-3
- Walauski MA, Khoshgoftaar TM. Unsupervised Label Generation for Severely Imbalanced Fraud Data. *J Big Data*. 2025;12(1):63. doi:10.1186/s40537-025-01120-x
- Kennedy RKL, Villanustre F, Khoshgoftaar TM. Unsupervised Feature Selection and Class Labeling for Credit Card Fraud. *J Big Data*. 2025;12(1):111. doi:10.1186/s40537-025-01154-1
- Fryze I, Naughton BD. Substandard and Falsified Medicine Recalls in the Legitimate Supply Chain: A Systematic Review of Evidence. *BMJ Open*. 2025;15(10):e103672. doi:10.1136/bmjopen-2025-103672
- McManus D, Naughton BD. A Systematic Review of Substandard, Falsified, Unlicensed and Unregistered Medicine Sampling Studies. *BMJ Glob Health*. 2020;5(8):e002393. doi:10.1136/bmjgh-2020-002393
- Kumar R, Sporn K, Waisberg E, Ong J, Paladugu P, Vadhera AS, et al. Navigating Healthcare AI Governance: The Comprehensive Algorithmic Oversight and Stewardship Framework for Risk and Equity. *Health Care Analysis*. 2025. doi:10.1007/s10728-025-00537-y
- Matloob I, Khan S, ur Rahman H, Hussain F. Medical Health Benefit Management System for Real-Time Notification of Fraud Using Historical Medical Records. *Applied Sciences*. 2020;10(15):5144. doi:10.3390/app10155144
- Sunilram C, Bhavana D, Abhisiraj K, Gayatri M. Identifying Health Insurance Claim Frauds Using Machine Learning Concepts. *International Journal of Engineering Research and Science & Technology*. 2024;20(1):223–8.
- Aarhi V, Raghavendra VS, Rao VD, Birudaraju H. Leveraging Machine Learning for Improved Detection of Medicare Fraud. *International Journal of Scientific Research in Engineering and Management*. 2025;09(06):1–8. doi:10.55041/IJSREM.NCFT031
- Anwer S, Faisal F, Qureshi MA. A Comprehensive Study of Healthcare Fraud Detection Using Machine Learning. *International Journal of Advanced Research in Science, Communication and Technology*. 2024;5(1):20–5. doi:10.5281/zenodo.10473530
- Subbarayudu Y, Vijendar Reddy G, Sandhya M, Bhargavi J, Abhilash PK, Pushkarna G. Evaluation of Distributed Topic Modeling Paradigms for Detection of Fraudulent Insurance Claims. In: *MATEC Web of Conferences*. 2024. p. 1111. doi:10.1051/mateconf/202439201111
- Wang Z, Chen X, Wu Y, Jiang L, Lin S, Qiu G. A Robust and Interpretable Ensemble Machine Learning Model for Predicting Healthcare Insurance Fraud. *Sci Rep*. 2025;15(1):218. doi:10.1038/s41598-024-82062-x
- Cherkaoui O, Anoun H, Maizate A. A Benchmark of Health Insurance Fraud Detection Using Machine Learning Techniques. *IAES International Journal of Artificial Intelligence*. 2024;13(2):1925–34.