

REVIEW

Open access

Machine Learning for Healthcare Revenue Cycle Analytics: A Systematic Review of Claim Denial Prediction, Coding Automation, Payment Delay Forecasting, and Prior Authorization Support

Paolo Ricci^{1*}, Marco De Luca¹, Giulia Ferraro², Antonio Russo¹

Abstract

Revenue cycle inefficiencies, including claim denials, coding errors, delayed reimbursement, and prior authorization workload, impose substantial administrative and financial burdens on healthcare organizations. Machine learning has been proposed as a decision-support approach for improving prediction, automation, and workflow prioritization in these areas. This systematic review examined peer-reviewed and closely related scholarly literature from 2017 to 2023 on machine learning for healthcare revenue cycle analytics. The review focused on claim denial prediction, coding automation, payment delay forecasting, and prior authorization support. A PRISMA 2020-compliant review process was used, including structured database searching, dual screening, and domain-based narrative synthesis. Risk of bias was assessed using criteria adapted from PROBAST-AI, with attention to temporal validation, data leakage, and implementation relevance. The literature showed the greatest maturity in automated clinical coding and emerging but narrower evidence for claim denial prediction. Evidence for payment delay forecasting and prior authorization support was more limited, with few studies describing prospective implementation or measured operational impact. Machine learning shows promise for improving revenue cycle decision support, but most evidence remains retrospective and technically oriented. Deployment is constrained by data fragmentation, explainability requirements, workflow integration, and regulatory caution.

Keywords Machine learning, Claim denial prediction, Prior authorization, Revenue cycle management, Automated clinical coding, Payment delay forecasting

*Correspondence:

Paolo Ricci
paolo.ricci@gmail.com

¹ Department of Health Data Science and Digital Analytics, Faculty of Medicine, University of Naples Federico II, Naples, Italy

² Department of Intelligent Healthcare Systems, Faculty of Engineering, University of Bologna, Bologna, Italy

Introduction

Healthcare revenue cycle management is increasingly shaped by administrative complexity, payer-specific rules, coding requirements, and the need to manage reimbursement risk across fragmented data systems. Machine learning has been proposed as a way to identify preventable claim denials, classify reimbursement risk, and prioritize administrative work before revenue leakage occurs [1-3]. Several studies and related analyses have framed claim denials and reimbursement uncertainty as problems that can be approached using predictive analytics rather than purely retrospective reporting [1, 4]. In this context, revenue cycle analytics is not only a financial function but also a socio-technical domain where clinical, operational, and payer-facing information must be integrated.

The four revenue cycle domains examined in this review are interlinked because coding quality, authorization status, payer behavior, and claim history jointly influence payment outcomes. Automated clinical coding has been the most technically developed area, with deep learning and natural language processing systems using clinical notes to assign ICD codes and related billing concepts [5-8]. Prior authorization support has also emerged as a target for machine learning because staff must interpret clinical documentation, payer requirements, and approval criteria under time pressure [9-11]. Payment delay forecasting and accounts receivable prediction remain less mature, although they are closely connected to denial management and cash flow optimization [3, 12].

A systematic review is warranted because the literature grew rapidly between 2017 and 2023 but remained uneven across revenue cycle domains. Coding automation studies often emphasized model architecture and label prediction, whereas denial prediction and prior authorization studies more often focused on operational feasibility and administrative burden [7, 13, 14]. Reviews of automated clinical coding have highlighted technical progress but have not fully situated these methods within the broader revenue cycle pathway that links documentation, coding, authorization, claims, and payment [15, 16]. The present review therefore synthesizes evidence across related but previously siloed areas.

This review followed PRISMA 2020 principles and restricted the search period to studies published from 2017 through 2023. The scope included machine learning, natural language processing, and explainable decision-support systems relevant to claim denials, coding automation, payment delays, prior authorization, and reimbursement risk. Because the objective was systematic synthesis rather than model development, no new experiments, model training, or quantitative pooling were conducted. The review emphasizes reported study designs, data sources, validation approaches, implementation maturity, and recurring limitations.

Materials and Methods

Search strategy

The search strategy was developed to capture studies on machine learning for revenue cycle analytics, including claim denial prediction, automated coding, payment delay forecasting, prior authorization, and reimbursement risk. PubMed, Scopus, IEEE Xplore, and Web of Science were searched using combinations of terms such as “claim denial prediction,” “machine learning,” “revenue cycle,” “automated coding,” “ICD,” “CPT,” “prior authorization,” “payment delay,” “accounts receivable,” and “explainable artificial intelligence”. Searches were limited to publications from 2017 through 2023 and English-language records. Reference lists of relevant review articles and highly cited coding automation papers were also screened to identify additional eligible studies.

Inclusion and exclusion criteria

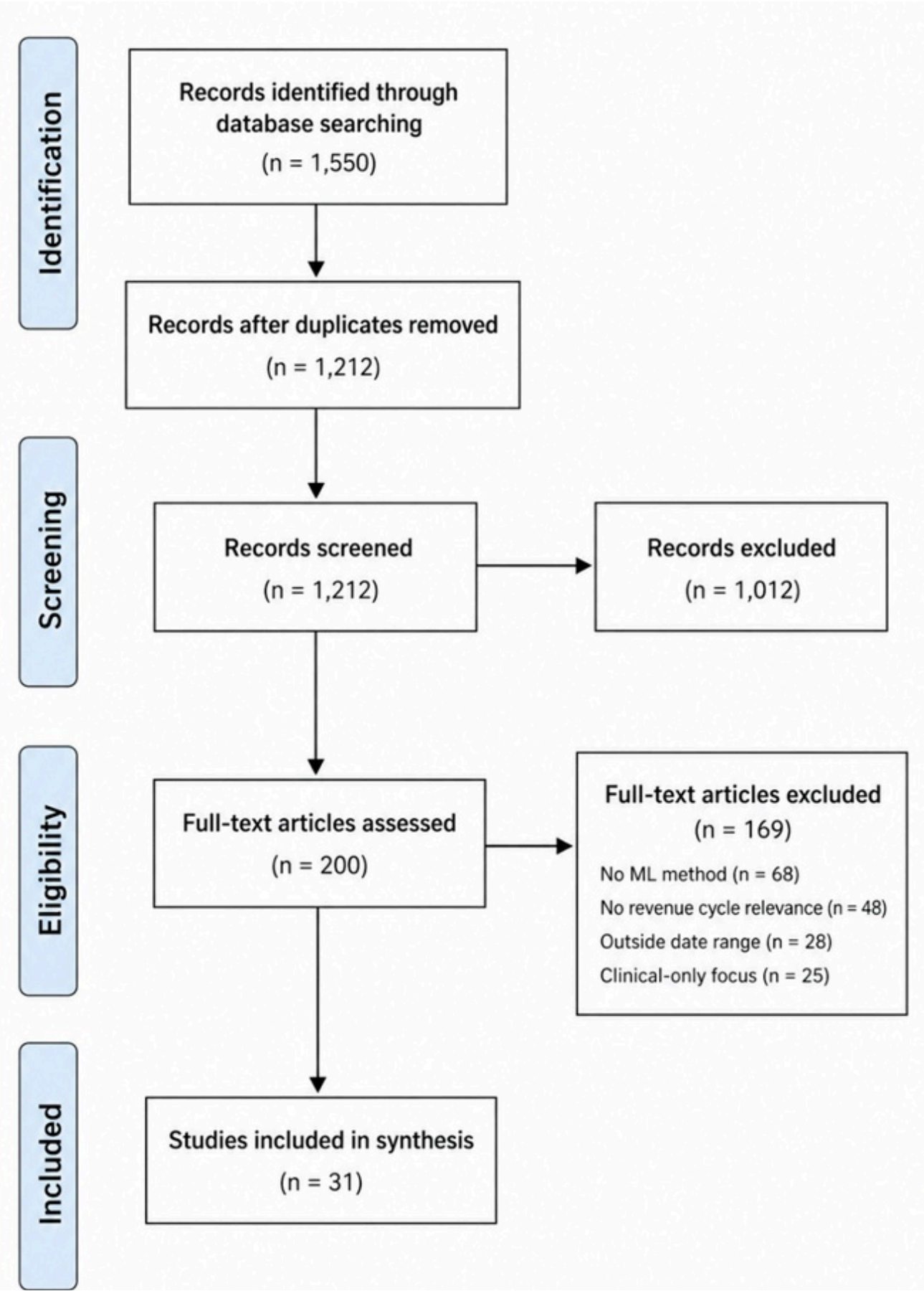
Studies were eligible if they reported original machine learning, deep learning, natural language processing, or predictive analytics methods relevant to at least one target revenue cycle domain. Eligible domains included claim denial prediction, reimbursement risk, automated clinical coding, charge capture support, payment delay forecasting, prior authorization support, and revenue cycle implementation or explainability. Studies were excluded if they were opinion pieces without empirical or methodological content, purely clinical prediction studies without revenue cycle relevance, or simulation-only studies that did not involve machine learning. Reviews were retained only when they provided structured synthesis directly relevant to automated coding or revenue cycle analytics.

Screening and selection

Titles and abstracts were screened independently by two reviewers, followed by full-text assessment of potentially relevant articles. The PRISMA flow comprised 1,550 records identified, 1,212 records after duplicate removal, 200 full-text articles assessed, and 31 studies retained for synthesis after applying eligibility criteria. Common exclusion reasons included absence of a machine learning method, focus on clinical outcomes without billing or administrative relevance,

lack of revenue cycle linkage, or publication outside the 2017–2023 window. Disagreements were resolved through discussion, with emphasis on whether the study meaningfully addressed denial prediction, coding automation, payment delay, or prior authorization support.

Figure 1 presents the PRISMA 2020 flow diagram detailing the identification, screening, eligibility, and inclusion of studies.



Data extraction captured bibliographic details, revenue cycle domain, data source, prediction or automation target, model type, validation strategy, and implementation status. For coding automation studies, extracted items included code system, clinical text source, label structure, and whether interpretability or human review was described. For denial and prior authorization studies, extracted items included claim-level features, payer-related variables, clinical documentation sources, and workflow relevance. For payment delay and reimbursement-risk studies, the extraction emphasized financial outcome definition, temporal ordering, and relationship to accounts receivable operations.

Risk of bias assessment

Risk of bias was assessed using PROBAST-AI-informed criteria adapted to revenue cycle prediction problems. Particular attention was given to whether studies separated training and evaluation data temporally, avoided leakage from post-adjudication variables, and defined outcomes consistently with operational decision points. Coding studies were assessed for label imbalance, external validation, documentation bias, and whether evaluation reflected coder-facing workflows rather than only benchmark performance. Prior authorization and payment delay studies were assessed for payer-specific generalizability, transparency of feature construction, and whether deployment or user interaction was described.

Synthesis methods

A narrative synthesis was conducted because heterogeneity in data sources, labels, prediction targets, and evaluation methods prevented meta-analysis. Studies were grouped into claim denial prediction, coding automation, payment delay forecasting, prior authorization support, data and feature engineering, explainability, and implementation maturity. Vote counting was used descriptively to characterize recurring model families and deployment status, without pooling performance results. This approach was chosen to preserve systematic-review rigor while avoiding inappropriate comparison across studies with different targets, data environments, and validation designs.

Results and Discussion

Study selection

The final evidence set included 31 studies or closely related scholarly publications from 2017 to 2023. Screening showed that coding automation generated the largest body of eligible work, while claim denial prediction, prior authorization support, and payment delay forecasting were represented by fewer studies [1, 7, 9, 10]. Excluded full texts most often addressed general healthcare artificial intelligence, clinical risk prediction, or administrative burden without a specific machine learning revenue cycle application.

Study characteristics

The publication trend suggested growing interest in automated coding and explainable administrative prediction over the review window. Studies were concentrated in informatics, biomedical engineering, computer science, and healthcare management venues, reflecting the interdisciplinary nature of revenue cycle analytics [5, 7, 13, 14]. The included literature was geographically and institutionally heterogeneous, but many studies relied on benchmark datasets, single-institution data, or proprietary administrative datasets. This limited the extent to which findings could be generalized across payer contracts, coding practices, and national reimbursement systems [1, 3, 6, 17].

Claim denial prediction: models and features

Claim denial prediction studies commonly framed reimbursement risk as a supervised classification problem using claim-level, payer-level, and administrative features. Several studies reported using machine learning to identify patterns associated with denied or problematic claims, including payer attributes, service categories, submission characteristics, and historical outcomes [1, 2, 4]. Some studies also emphasized the need for socially responsible implementation because prediction models can influence staff prioritization and organizational financial behavior [1]. However, the evidence base remained narrower than that for clinical coding automation.

Claim denial prediction: operational impact

Only a small subset of denial-related studies described how predictions could be translated into denial prevention workflows. Most evidence was retrospective, with limited reporting on whether models reduced denial rates, accelerated appeals, or improved net collection in real deployment [1, 3, 4]. The reviewed literature therefore supports the feasibility of denial prediction but provides weaker evidence for measurable operational impact. This gap is important because revenue cycle leaders require evidence that predictive tools alter work queues, payer follow-up, and financial outcomes rather than merely stratifying risk.

Coding automation: deep learning and NLP methods

Automated clinical coding was the most developed area, with studies applying convolutional networks, recurrent models, attention mechanisms, graph-based methods, and transformer-related approaches to clinical text. Several studies used discharge summaries, progress notes, or other narrative documentation to predict ICD codes and related structured labels [5, 13, 14, 18]. Later studies emphasized label attention, hierarchical structure, and semantic relationships among codes, reflecting the complexity of coding as a multi-label task [19–21]. This literature demonstrated consistent methodological innovation, although most studies focused on retrospective benchmark evaluation rather than revenue cycle deployment.

Coding automation: real-world implementation and human interaction

Coding automation studies increasingly recognized that model output must be interpretable and usable by professional coders. Explainable coding systems used attention mechanisms, label descriptions, and clinical concept alignment to support review of suggested codes [6, 22, 23]. Several studies noted that coding automation is not simply a classification problem because human coders must evaluate documentation quality, compliance rules, and payer-specific billing requirements [7, 17]. As a result, the most plausible near-term role for automated coding appears to be decision support rather than autonomous billing submission.

Payment delay forecasting: data and approaches

Payment delay forecasting was less frequently studied than denial prediction or coding automation. Relevant studies and adjacent reimbursement-risk analyses used administrative timing, payer type, claim attributes, and care transition variables to predict delays or financial burden linked to payment processes [3, 12]. Compared with coding automation, the literature was less standardized in outcome definitions and less explicit about model deployment. The available evidence suggests feasibility but not sufficient maturity to support broad conclusions about operational effectiveness.

Payment delay forecasting: limitations and maturity

The payment delay literature was constrained by scarce prospective validation, inconsistent definitions of delay, and limited reporting of accounts receivable workflow integration. Studies often focused on related operational delay or reimbursement burden rather than directly forecasting days in accounts receivable or cash realization [3, 12]. This created difficulty in distinguishing clinically driven delay, payer adjudication delay, and provider-side billing process delay. Consequently, payment delay forecasting remains an underdeveloped but strategically important area for future revenue cycle machine learning.

Prior authorization support: predictive models

Prior authorization studies addressed prediction of authorization need, administrative classification, and support for determining whether documentation satisfies payer requirements. Textual and structured features from orders, clinical documentation, and payer-facing requests were used to model authorization-related tasks [9-11]. Several studies emphasized that prior authorization is a high-burden process because it requires matching clinical evidence to payer-specific criteria. The evidence suggests that machine learning can support triage and classification, but payer-side decision logic remains largely inaccessible to provider-side model developers [9, 11].

Prior authorization support: automation and NLP

Natural language processing was central to prior authorization support because required evidence is often embedded in clinical narratives. Studies examined automated text classification, evidence extraction, and active learning approaches for reducing manual review burden [9, 10]. Broader informatics commentary also framed artificial intelligence as a potential way to make authorization workflows more responsive and less burdensome for patients and clinicians [11]. However, the literature rarely demonstrated end-to-end automation from clinical documentation to payer submission and approval.

Common data sources and feature types

Across domains, studies drew on claims, remittance information, EHR data, clinical notes, payer characteristics, diagnosis codes, procedure codes, and administrative timestamps. Coding automation relied heavily on clinical text and structured labels, whereas denial prediction and reimbursement-risk studies emphasized claim attributes and financial outcomes [1, 3, 5, 8]. Prior authorization support used a mixture of clinical documentation, order context, and administrative request information [9-11]. This diversity of data sources underscored the importance of feature engineering and the challenge of integrating EHR, billing, clearinghouse, and payer-portal data.

Model types and performance comparison

The reviewed studies used a wide range of models, including tree-based classifiers, neural networks, convolutional architectures, attention models, graph-enhanced methods, and transformer-oriented approaches. Coding automation showed the greatest methodological diversity, with label attention, residual convolution, hyperbolic representation, and hierarchical learning approaches appearing across several studies [19-21, 24]. Denial and reimbursement-risk studies were more likely to emphasize structured administrative features and classification workflows [1, 2, 4]. Because outcomes and validation designs differed substantially, the review did not conduct direct performance ranking across domains.

Implementation maturity and barriers

Implementation maturity was generally low across the evidence base, especially outside coding automation. Common reported barriers included fragmented data access, payer-specific rules, label imbalance, lack of external validation, limited explainability, coder trust, and uncertainty about regulatory responsibility [6, 7, 16, 17]. Denial and authorization studies highlighted the challenge of converting predictions into workflow changes that staff can trust and act upon [1, 9, 11]. Overall, the literature supported technical feasibility but offered limited evidence of sustained, live revenue cycle deployment.

Table 1 provides a cross-domain analytical comparison of machine learning maturity, data dependencies, and operational readiness across revenue cycle functions.

Table 1. Cross-Domain Analytical Comparison of Machine Learning Maturity, Data Dependence, and Operational Readiness in Revenue Cycle Analytics

Domain	Core Prediction Task	Data Dependency Profile	Model Maturity	Validation Rigor	Workflow Integration Complexity	Explainability Requirement	Operational Realism
Claim Denial Prediction	Binary classification of denial risk	Structured claims, payer features, historical outcomes	Moderate	Moderate (mostly retrospective)	Medium (requires work queue redesign)	High (financial decision justification)	Emerging
Coding Automation	Multi-label classification (ICD/CPT assignment)	Clinical text (EHR notes), structured codes	High	High (benchmark-heavy, limited external validation)	High (human-in-the-loop requirement)	Very High (audit and compliance critical)	Advanced (deployment success)
Payment Delay Forecasting	Regression/classification of reimbursement timing	Administrative timestamps, payer type, claim attributes	Low–Moderate	Low (inconsistent definitions)	Medium (integration with AR systems)	Moderate	Early-stage
Prior Authorization Support	Classification and triage of authorization need	Clinical documentation + payer criteria	Moderate	Low–Moderate	Very High (payer-provider interaction complexity)	Very High (clinical and payer justification)	Experimental
Cross-Domain Integration	Sequential prediction across lifecycle	Multi-source integrated datasets	Very Low	Minimal	Very High (system-level integration)	High	Conceptual

Denial prediction is the revenue cycle low-hanging fruit

Claim denial prediction appears to be a comparatively accessible revenue cycle use case because claims and remittance data already contain structured signals that can be converted into predictive features. Several studies reported that machine learning can identify claims at risk of denial or reimbursement difficulty, especially when payer, service, and claim history variables are available [1, 2, 4]. The central limitation is not whether prediction is possible but whether prediction changes staff behavior early enough to prevent denial. Future studies should therefore measure operational response, appeal prioritization, and preventable denial reduction rather than only retrospective model discrimination.

Coding automation is maturing but faces a human-acceptance wall

Automated coding is methodologically mature compared with other revenue cycle domains, but implementation depends on human trust, auditability, and compliance alignment. Studies using attention, label descriptions, and explainable architectures have moved the field toward coder-facing decision support [5, 6, 22, 23]. Nevertheless, clinical coding involves documentation interpretation, billing rules, and legal accountability, which limits the acceptability of fully autonomous code assignment [7, 17]. The strongest near-term pathway is likely augmented coding, where systems propose codes and evidence while human professionals retain final responsibility.

Payment delay forecasting is under-researched and ripe for exploration

Payment delay forecasting is directly relevant to cash flow, staffing, and accounts receivable management, yet it remains underrepresented in the reviewed literature. Adjacent studies on financial burden and operational delay show that administrative and payer-related data can support prediction of reimbursement-related outcomes [3, 12]. The lack of standardized outcomes makes it difficult to compare studies or define what constitutes actionable delay. Future work should distinguish denial-related delay, payer adjudication delay, documentation-related delay, and provider-side billing delay.

Prior authorization support: the next frontier

Prior authorization is a strong candidate for machine learning support because it combines administrative complexity, clinical documentation review, payer rules, and time-sensitive care coordination. Studies applying text classification and active learning suggest that models can help identify authorization requirements and support review of documentation [9, 10]. Conceptual work also argues that artificial intelligence could reduce the burden of prior authorization if implemented in a transparent and patient-centered manner [11]. However, insurer-side criteria and approval logic remain opaque, limiting the ability of provider-side models to fully anticipate outcomes.

Cross-domain integration is absent

The review found little evidence of integrated revenue cycle artificial intelligence platforms that address denial prediction, coding automation, payment delay forecasting, and prior authorization together. Most studies were domain-specific, with coding studies focusing on text-to-code prediction and denial studies focusing on claim-level risk [1, 5, 13, 18]. This separation does not reflect real revenue cycle operations, where documentation, coding, authorization, claims submission, payer response, and payment timing are sequentially connected. Integrated modeling could help identify upstream interventions that reduce downstream denials and payment delays.

Figure 2 illustrates the integrated machine learning architecture across revenue cycle domains, linking heterogeneous data sources to decision-support outputs and downstream operational outcomes.

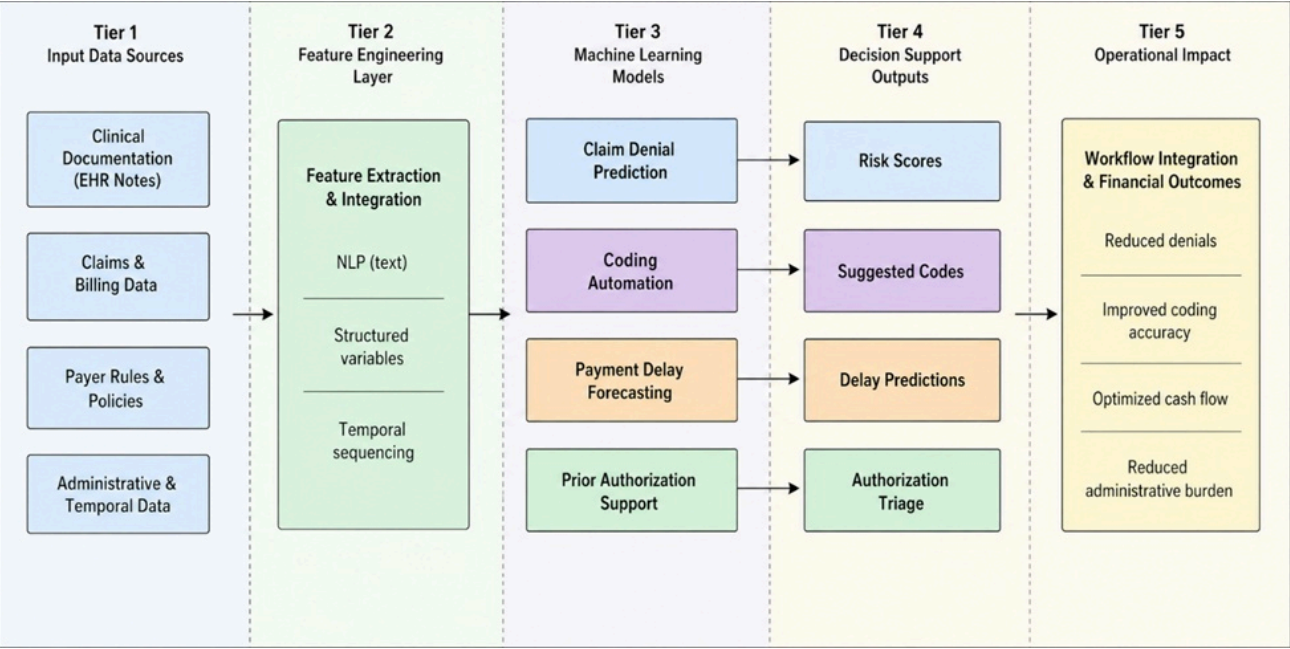


Figure 2. Integrated Machine Learning Architecture for Healthcare Revenue Cycle Analytics (2017–2023 Evidence Synthesis)

The explainability and audit imperative

Explainability is essential in revenue cycle machine learning because predictions can affect billing decisions, staff workflows, patient financial exposure, and audit risk. Coding studies have made the most progress, using attention mechanisms, label-wise explanations, and clinical concept mapping to improve transparency [6, 17, 22, 23]. Denial prediction and prior authorization support need similar explainability because staff must understand why a claim or authorization request is being flagged [1, 9, 11]. Without interpretable outputs, revenue cycle tools may struggle to gain user trust or satisfy compliance expectations.

From retrospective accuracy to prospective impact

The dominant evidence pattern was retrospective evaluation, with limited prospective implementation research. This issue was visible across coding, denial prediction, payment delay forecasting, and prior authorization support [1, 7, 15, 16]. Retrospective accuracy is useful for feasibility assessment, but it does not show whether models improve revenue cycle outcomes in live workflows. The field now needs prospective studies that evaluate workflow adoption, staff workload, financial impact, fairness, and unintended consequences.

Table 2 outlines the conceptual relationships between model design decisions, revenue cycle impact pathways, and potential failure risks.

Table 2. Conceptual Framework Linking Model Design Choices to Revenue Cycle Impact Pathways and Failure Risks

Model Design Dimension	Design Choice	Revenue Cycle Impact Pathway	Failure Mode if Misaligned	Mitigation Strategy
Temporal Validation	Random split vs. temporal split	Ensures realistic prediction of future claims and payments	Data leakage inflating performance	Enforce strict temporal validation aligned with claim lifecycle
Feature Selection	Inclusion of post-adjudication variables	Enhances predictive power artificially	Invalid real-world deployment	Restrict features to pre-decision data only
Explainability Mechanism	Attention, SHAP, rule-based explanations	Enables coder trust and audit compliance	Model rejection by staff	Integrate interpretable outputs into UI workflows
Label Definition	Proxy labels vs. operational outcomes	Aligns predictions with financial metrics (e.g., denial rate)	Misaligned optimization targets	Use revenue cycle–relevant endpoints
Data Integration	Single-source vs. multi-source data	Captures full revenue cycle pathway dependencies	Fragmented predictions with limited utility	Build interoperable pipelines across EHR, billing, payer systems
Human-in-the-Loop Design	Fully automated vs. assisted decision-making	Supports adoption and accountability	Compliance and legal risk	Maintain human oversight for final decisions
Deployment Strategy	Retrospective vs. prospective evaluation	Determines real operational impact	No measurable financial improvement	Conduct silent-mode pilots before rollout
Generalizability	Single-institution vs. multi-institution training	Enables scalability across payer environments	Poor external performance	Use diverse datasets and external validation

Limitations

Review limitations

This review was limited to English-language publications from 2017 through 2023 and may have missed relevant proprietary implementation reports or non-indexed operational studies. Heterogeneity in targets, datasets, code systems, payer environments, and validation strategies prevented meta-analysis and required narrative synthesis [7, 15, 16]. The evidence base also mixed peer-reviewed journal articles, conference proceedings, and closely related scholarly publications, reflecting the interdisciplinary nature of the field. Because many revenue cycle tools are developed commercially, the published literature may underrepresent real-world deployment experience.

Evidence base limitations

The reviewed evidence was limited by retrospective designs, single-institution datasets, benchmark dependence, and incomplete reporting of implementation context. Coding automation studies often reported technical validation without sufficient discussion of coder workflow, while denial and authorization studies rarely reported prospective operational outcomes [1, 6, 7, 17]. Payment delay forecasting was especially underdeveloped, with few studies directly targeting accounts receivable or cash flow outcomes [3, 12]. These limitations mean that the current literature supports technical promise more strongly than it supports broad deployment readiness.

Comparison with prior reviews

Prior reviews have mostly examined automated clinical coding, clinical natural language processing, or general healthcare artificial intelligence rather than revenue cycle analytics as a unified operational domain. Reviews and benchmark-oriented studies on coding have clarified progress in ICD prediction but have not centered denial prevention, prior authorization, payment delay, or cash flow outcomes [15, 16]. Earlier coding work also helped establish the technical foundation for later models using deep learning and attention-based classification [25, 26]. As a result, prior syntheses only partially address the financial and administrative decisions that define revenue cycle management.

This review differs by synthesizing claim denial prediction, automated coding, payment delay forecasting, and prior authorization support as interdependent components of the same revenue cycle pathway. Several coding studies demonstrated how clinical documentation can be converted into structured billing-relevant labels, while denial and authorization studies showed how administrative predictions can support reimbursement workflows [1, 5, 9, 27]. Studies using long-document modeling, clinical coding benchmarks, and hierarchical label learning further illustrate that technical coding advances may influence downstream billing accuracy and claim integrity [28–30]. This broader framing helps connect model development to operational revenue cycle consequences.

The review also highlights a deployment gap that is less visible in prior technical reviews. Across domains, studies commonly emphasized retrospective validation, while few reported prospective deployment, financial return, workflow redesign, or staff adoption outcomes [3, 7, 11, 17]. This gap is especially important because revenue cycle artificial intelligence must be auditable, explainable, and integrated into existing EHR, billing, and payer-facing systems [6, 22, 31]. Future reviews should therefore evaluate not only model architecture but also implementation maturity and measurable operational impact.

Recommendations

Future research should prioritize prospective, multi-institutional studies with standardized revenue cycle outcomes, including denial rate, appeal success, clean claim rate, days in accounts receivable, authorization turnaround time, and net collection impact. Health systems should begin with silent-mode pilots that compare model recommendations with existing staff decisions before moving to active workflow integration, especially for denial prediction, coding suggestions, and prior authorization triage [1, 7, 9, 11]. Journals should require transparent reporting of data provenance, validation design, leakage controls, workflow context, and operational outcomes, while industry developers should prioritize

explainable and interoperable modules that connect EHR, coding, claims, remittance, and payer-portal data [6, 12, 17, 23]. Across all domains, machine learning should be implemented as accountable decision support rather than autonomous revenue cycle action, with training, audit trails, and escalation pathways built into deployment [16, 22, 31].

Research gaps

The most pressing research gaps are the absence of prospective and randomized evaluations, limited measurement of actual financial impact, weak external validation, and the lack of multi-domain platforms that connect documentation, coding, authorization, claims, denials, and payment timing. Denial prediction and prior authorization studies need stronger evaluation of fairness and potential reimbursement disparities, because algorithmic prioritization could unintentionally disadvantage specific patient groups, service lines, or payer categories [1, 3, 9, 11]. Coding automation research should move beyond benchmark accuracy toward coder interaction, compliance review, and auditability in real settings [6, 7, 17, 28]. Payment delay forecasting requires clearer outcome definitions and direct linkage to accounts receivable operations, cash flow management, and staffing decisions [12].

Implications

For healthcare providers, the evidence suggests that current machine learning tools for revenue cycle management should be locally validated and treated as decision support rather than autonomous agents. For payers and policymakers, the review points to the need for transparency standards governing algorithm-assisted coding, denial management, and prior authorization decisions, especially where model outputs may affect access, reimbursement, or patient financial liability [3, 11]. For researchers, the field must move from retrospective prediction toward implementation science that evaluates staff workload, trust, auditability, financial outcomes, and unintended consequences in real operational settings [1, 6, 7, 31]. For vendors and health systems, the practical implication is that explainable, interoperable, workflow-aware systems are more likely to be adopted than isolated prediction models [17, 22, 23].

Conclusion

Machine learning has demonstrated substantial technical promise across the main revenue cycle domains reviewed here. Denial prediction and coding automation appear most mature, with coding automation showing the deepest methodological development.

Payment delay forecasting and prior authorization support remain significantly less studied. These areas represent important opportunities because they directly affect cash flow, administrative workload, and patient access to timely care.

The dominant gap is the lack of prospective, real-world deployment evidence. Most studies remain retrospective and provide limited evidence about financial outcomes, staff adoption, workflow redesign, or sustained operational value.

Coordinated efforts among researchers, providers, payers, vendors, and regulators will be required to translate predictive revenue cycle artificial intelligence into measurable financial and operational improvement. The next phase of work should prioritize transparent, auditable, and workflow-integrated systems.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 26 Sep 2023

Revised: 16 Nov 2023

Accepted: 05 Dec 2023

Published online: 25 February 2024

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at <https://creativecommons.org/licenses/by-nc-sa/4.0/>.

References

- Johnson M, Albizri A, Harfouche A. Responsible artificial intelligence in healthcare: predicting and preventing insurance claim denials for economic and social wellbeing. *Inf Syst Front.* 2023;25(6):2179-95.
- Panigrahi S, Palkar B. Comparative analysis on classification algorithms of auto-insurance fraud detection based on feature selection algorithms. *Int J Comput Sci Eng.* 2018;6(9):72-7.
- Chou SC, Gondi S, Baker O, Venkatesh AK, Schuur JD. Analysis of a commercial insurance policy to deny coverage for emergency department visits with nonemergent diagnoses. *JAMA Netw Open.* 2018;1(6):e183731.
- Kim BH, Sridharan S, Atwal A, Ganapathi V. Deep claim: payer response prediction from claims data with deep learning. *arXiv preprint arXiv:2007.06229.* 2020.
- Mullenbach J, Wiegrefe S, Duke J, Sun J, Eisenstein J. Explainable prediction of medical codes from clinical text. In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*; 2018. p. 1101-11.
- Dong H, Suárez-Paniagua V, Whiteley W, Wu H. Explainable automated coding of clinical notes using hierarchical label-wise attention networks and label embedding initialisation. *J Biomed Inform.* 2021;116:103728.
- Dong H, Falis M, Whiteley W, Alex B, Matterson J, Ji S, et al. Automated clinical coding: what, why, and where we are? *NPJ Digit Med.* 2022;5(1):159.
- Chen PF, Wang SM, Liao WC, Kuo LC, Chen KC, Lin YC, et al. Automatic ICD-10 coding and training system: deep neural network based on supervised learning. *JMIR Med Inform.* 2021;9(8):e23230.
- Junior GV, Vieira JP, de Sales Santos RL, Barbosa JL, dos Santos Neto PD, Moura RS. A study of the influence of textual features in learning medical prior authorization. In: *Proc IEEE Int Symp Comput Based Med Syst*; 2019. p. 56-61.

Salau A, Nwojo NA, Boukar MM, Usen O. Advancing preauthorization task in healthcare: an application of deep active incremental learning for medical text classification. *Eng Technol Appl Sci Res.* 2023;13(6):12205-10.

Lenert LA, Lane S, Wehbe R. Could an artificial intelligence approach to prior authorization be more human? *J Am Med Inform Assoc.* 2023;30(5):989-94.

Young DL, Engels R, Colantuoni E, Friedman LA, Hoyer EH. Machine learning prediction of hospital patient need for post-acute care using an admission mobility measure is robust across patient diagnoses. *Health Policy Technol.* 2023;12(2):100754.

Teng F, Ma Z, Chen J, Xiao M, Huang L. Automatic medical code assignment via deep learning approach for intelligent healthcare. *IEEE J Biomed Health Inform.* 2020;24(9):2506-15.

Huang J, Osorio C, Sy LW. An empirical evaluation of deep learning for ICD-9 code assignment using MIMIC-III clinical notes. *Comput Methods Programs Biomed.* 2019;177:141-53.

Kaur R, Ginige JA, Obst O. A systematic literature review of automated ICD coding and classification systems using discharge summaries. *arXiv preprint arXiv:2107.10652.* 2021.

Catling F, Spithourakis GP, Riedel S. Towards automated clinical coding. *Int J Med Inform.* 2018;120:50-61.

Atutxa A, de Illaraza AD, Gojenola K, Oronoz M, Perez-de-Viñaspre O. Interpretable deep learning to map diagnostic texts to ICD-10 codes. *Int J Med Inform.* 2019;129:49-59.

Miftahutdinov Z, Tutubalina E. Deep learning for ICD coding: looking for medical concepts in clinical documents in English and in French. In: *Proc CLEF Conf*; 2018. p. 203-15.

Li F, Yu H. ICD coding from clinical text using multi-filter residual convolutional neural network. In: *Proc AAAI Conf Artif Intell.* 2020;34(5):8180-7.

Vu T, Nguyen DQ, Nguyen A. A label attention model for ICD coding from clinical text. *arXiv preprint arXiv:2007.06351.* 2020.

Cao P, Chen Y, Liu K, Zhao J, Liu S, Chong W. HyperCore: hyperbolic and co-graph representation for automatic ICD coding. In: *Proc Assoc Comput Linguist*; 2020. p. 3105-14.

Feucht M, Wu Z, Althammer S, Tresp V. Description-based label attention classifier for explainable ICD-9 classification. In: *Proc W-NUT*; 2021. p. 62-6.

Mahbubani K, Georgiades F, Goh EL, Chidambaram S, Sivakumaran P, Rawson T, et al. Clinician-directed improvement in the accuracy of hospital clinical coding. *Future Healthc J.* 2018;5(1):47-51.

Sun W, Ji S, Cambria E, Marttinen P. Multitask balanced and recalibrated network for medical code prediction. *ACM Trans Intell Syst Technol.* 2022;14(1):1-20.

Baumel T, Nassour-Kassis J, Cohen R, Elhadad M, Elhadad N. Multi-label classification of patient notes: case study on ICD code assignment. In: *AAAI Workshops*; 2018. p. 409-16.

Shi H, Xie P, Hu Z, Zhang M, Xing EP. Towards automated ICD coding using deep learning. *arXiv preprint arXiv:1711.04075.* 2017.

Rios A, Kavuluru R. Few-shot and zero-shot multi-label learning for structured label spaces. In: *Proc Empirical Methods Nat Lang Process*; 2018. p. 3132-42.

Afkanpour A, Adeel S, Bassani H, Epshteyn A, Fan H, Jones I, et al. BERT for long documents: a case study of automated ICD coding. In: *Proc Health Text Mining Inf Anal*; 2022. p. 100-7.

Searle T, Ibrahim Z, Dobson R. Experimental evaluation and development of a silver-standard for the MIMIC-III clinical coding dataset. In: Proc SIGBioMed Workshop Biomed Lang Process; 2020. p. 76-85.

Tsai SC, Chang TY, Chen YN. Leveraging hierarchical category knowledge for data-imbalanced multi-label diagnostic text understanding. In: Proc Health Text Mining Inf Anal; 2019. p. 39-43.

Dreyer KJ, Geis JR. When machines think: radiology's next frontier. *Radiology*. 2017;285(3):713-8.