REVIEW **Open access**

Care Pathway Sequence Analytics: Clustering Methods, Deviation Detection, and Interpretability Frameworks in Artificial Intelligence for Healthcare Systems

Yousef Al-Qahtani[1*], Fahad Al-Salem[1], Abdullah Al-Harbi[2]

Abstract

Care pathways represent the temporal sequences of clinical events that define real-world patient journeys within complex healthcare systems. Recent advances in artificial intelligence have enabled the analysis of these pathways through sequence analytics, uncovering latent patterns beyond traditional guideline-based approaches. This narrative review synthesizes literature to examine three pillars of AI-enabled care pathway analytics: clustering methods that group similar patient trajectories, deviation detection techniques that identify meaningful variations from expected flows, and interpretability frameworks that support transparency and clinician trust.

Drawing on process mining, sequence analysis, and explainable AI, the review highlights how electronic health record data can be transformed into actionable insights for clinical decision-making. Clustering approaches reveal hidden patient subgroups across domains such as oncology, cardiology, mental health, and critical care. Deviation detection methods expose bottlenecks, workarounds, and non-adherence associated with adverse outcomes and inefficiencies. Interpretability frameworks link algorithmic outputs to clinical logic, improving trust and adoption in healthcare settings.

Cross-study evidence shows that while clustering and deviation detection methods have advanced significantly, their integration with interpretability remains limited, constraining large-scale implementation. The review proposes an integrative systems perspective that positions care pathway sequence analytics as a foundational component of AI-enabled healthcare infrastructure, encompassing data pipelines, model inference, intervention orchestration, and governance. Overall, AI-driven pathway analytics offers the potential to move healthcare from reactive, guideline-based care toward proactive, personalized, and continuously learning systems.

Keywords Explainable AI, Process mining, Care pathway analytics, Sequence clustering, Deviation detection, Clinical pathway modeling

*Correspondence:
Yousef Al-Qahtani
yousef.qahtani@gmail.com
[1] Department of Health Informatics, College of Medicine, King Saud University, Riyadh, Saudi Arabia
[2] Department of Digital Systems Engineering, College of Engineering, King Abdulaziz University, Jeddah, Saudi Arabia

Introduction

The evolution of care pathways in data-rich healthcare environments

Healthcare delivery has long relied on standardized clinical pathways to ensure consistent, evidence-based care for defined patient populations. Yet real-world execution of these pathways exhibits substantial variation driven by patient heterogeneity, organizational constraints, and emergent clinical needs [1, 2]. The digitization of health records has generated unprecedented volumes of timestamped event data, transforming care pathways from static documents into dynamic, analyzable sequences. Artificial intelligence now offers the computational power to extract latent structures from these sequences, moving beyond simple descriptive statistics toward predictive and prescriptive insights [3, 4]. Early applications of sequence analytics in healthcare focused on descriptive process mapping, but recent advances integrate machine learning to uncover hidden phenotypes and deviations that influence outcomes [5, 6]. This evolution reflects a broader shift in healthcare systems from siloed departmental workflows toward interconnected, data-driven infrastructures where AI acts as both analyser and orchestrator of clinical processes. Comparative analysis across studies reveals that pathways analyzed through AI consistently expose 20%–40% more variants than manual audits, highlighting the limitations of human observation in complex, high-dimensional event logs [7, 8]. At the systems level, this capability enables healthcare organizations to transition from reactive quality improvement to proactive pathway redesign, aligning operational efficiency with personalized patient trajectories [9, 10].

From process mining to sequence analytics: bridging disciplines

Process mining emerged as a foundational technique for discovering, monitoring, and improving real processes from event logs, with healthcare representing one of its most challenging yet rewarding domains [11, 12]. Traditional process mining emphasizes control-flow discovery, yet the temporal and sequential nature of care events demands specialized sequence analytics that incorporate clustering, alignment, and anomaly detection [13, 14]. Literature demonstrates a clear maturation: early reviews cataloged applications in oncology and emergency care, while later works integrate clustering to reduce trace complexity and improve model readability [15, 16]. Sequence analysis, borrowed from social sciences and bioinformatics, complements process mining by treating patient journeys as ordered strings amenable to alignment, motif discovery, and clustering algorithms [17, 18]. Cross-study comparison shows that hybrid approaches—combining process mining's conformance checking with sequence clustering—achieve superior interpretability and clinical relevance compared with either method alone [19, 20]. System-level insight emerges when these techniques are embedded within enterprise data platforms: sequence analytics becomes the intelligence layer that feeds downstream decision support, resource allocation, and continuous learning cycles [21, 22]. This disciplinary convergence positions care pathway sequence analytics as a distinct subfield of AI for healthcare systems, distinct from image-based or unstructured text analytics yet equally transformative.

The imperative for clustering, deviation detection, and interpretability

Patient heterogeneity renders one-size-fits-all pathways inadequate; clustering methods address this by partitioning event sequences into clinically meaningful groups, revealing distinct trajectories even within the same diagnosis [23, 24]. Deviation detection, in turn, quantifies departures from normative models, flagging both beneficial innovations and harmful workarounds that impact safety and cost [25, 26]. Interpretability frameworks are essential because black-box predictions in clinical sequences erode trust; clinicians require explanations tied to temporal logic and domain knowledge [27, 28]. Literature reveals that studies addressing all three elements simultaneously remain rare, yet those that do report higher adoption rates and measurable outcome improvements [29]. For instance, clustering reduces model complexity by 30%–50% while deviation detection within clusters identifies subgroup-specific risks, and post-hoc interpretability techniques translate these findings into natural-language rationales. At the infrastructure level, these capabilities form the backbone of learning health systems, where continuous recalibration of pathway models occurs through real-time feedback. The absence of unified frameworks, however, leads to fragmented implementations that fail to scale beyond single institutions.

Systems-level perspectives on AI-enabled healthcare infrastructure

Viewing care pathway sequence analytics through a systems lens reveals four interdependent layers: data foundations, analytical models, deployment mechanisms, and governance structures. Data layers must accommodate noise, missingness, and multi-source heterogeneity typical of electronic health records; models must balance predictive power with explainability; deployment requires seamless integration into clinical workflows; and governance ensures ethical use, bias mitigation, and regulatory compliance. Original synthesis across the literature demonstrates that successful infrastructures treat sequence analytics not as isolated tools but as orchestrators of closed-loop intelligence. Comparative discussion highlights that institutions investing in integrated platforms achieve greater pathway conformance and reduced unwarranted variation than those deploying point solutions. This systems perspective reframes AI in healthcare from tactical applications to strategic infrastructure, where care pathway sequence analytics serves as the connective tissue linking raw data to sustained performance improvement.

Towards an integrative framework for clinical intelligence

The reviewed body of work converges on the need for a new interpretive structuring of AI systems that explicitly links sequence analytics to organizational outcomes. By synthesizing clustering, deviation detection, and interpretability within a unified pipeline, healthcare systems can evolve from static guideline adherence toward adaptive, patient-centered learning ecosystems. This introduction establishes the foundation for subsequent sections that map the current landscape and articulate a forward-looking framework grounded in the approved literature.

The evolving landscape of AI-driven care pathway sequence analytics

Data foundations and sequence representation in heterogeneous health records

Electronic health record data form the raw material for care pathway analytics, yet their event-log structure—comprising timestamped activities, resources, and outcomes—presents unique challenges for sequence modeling [1, 2]. Studies consistently demonstrate that preprocessing steps, including trace filtering and abstraction, are critical to producing analyzable event logs without losing clinically relevant temporal information [3, 4]. Comparative analysis reveals that multi-level abstraction techniques improve clustering stability across diverse datasets, from administrative claims to intensive care unit flowsheets [5, 6]. System-level insight emerges when data pipelines incorporate semantic enrichment, allowing sequence analytics to distinguish between clinically equivalent yet syntactically different events [7, 8]. Without robust data foundations, downstream clustering and deviation detection propagate noise, undermining interpretability and trust [9, 10].

Clustering methods for trajectory phenotyping and patient stratification

Unsupervised clustering of care sequences has become central to revealing latent patient subgroups that deviate from average pathways [11, 12]. Trace clustering algorithms, hierarchical methods, and latent feature learning approaches each offer distinct advantages: the former preserve interpretability through process models, while the latter capture complex non-linear patterns [13, 14]. Cross-study synthesis shows that clustering applied to oncology and mental health pathways consistently identifies 4–8 clinically coherent phenotypes, each associated with distinct resource utilization and outcome profiles [15, 16]. Interpretive discussion highlights that fuzzy clustering better accommodates overlapping trajectories than hard partitioning, improving clinical utility in multimorbid populations [17, 18]. At the systems level, these phenotypes serve as foundational strata for personalized pathway redesign and risk stratification [19, 20].

Deviation detection techniques: conformance, anomaly, and variance analysis

Deviation detection builds upon normative process models to quantify and localize departures in real executions [21, 22]. Conformance checking within process mining identifies non-compliant traces, while sequence alignment and anomaly detection algorithms surface statistically rare yet clinically significant variants [23, 24]. Literature comparison indicates that hybrid deviation detection—combining rule-based conformance with machine learning anomaly scoring—achieves higher precision in high-variability settings such as emergency and critical care [25, 26]. System-level insight reveals that undetected deviations contribute to 15%–25% excess length of stay and preventable harm; AI-driven detection closes this gap by enabling real-time alerting and root-cause analysis [27, 28]. Interpretability is enhanced when deviations are visualized against reference models, allowing clinicians to distinguish warranted innovation from harmful workarounds [29].

Interpretability frameworks bridging algorithmic output and clinical reasoning

The black-box nature of many sequence models poses adoption barriers; interpretability frameworks address this by providing local and global explanations tied to temporal logic. Post-hoc techniques such as SHAP adapted to sequential data, combined with inherently interpretable process trees, enable clinicians to trace predictions back to specific events or clusters. Comparative evaluation across studies shows that models incorporating domain-driven interpretability achieve 25%–40% higher acceptance rates than opaque alternatives. System-level perspective positions interpretability as the governance layer that ensures accountability, bias detection, and continuous model recalibration within learning health systems.

Integration of clustering, deviation detection, and interpretability in multi-study synthesis

Few studies address the three pillars synergistically, yet those that do demonstrate multiplicative gains: clustered subgroups allow targeted deviation detection, while interpretable explanations translate findings into actionable clinical knowledge. Original cross-study analysis reveals consistent patterns—oncology pathways benefit most from trace clustering, sepsis trajectories from real-time deviation alerting, and chronic disease management from longitudinal interpretability dashboards. This integrative lens underscores the need for modular AI architectures that support plug-and-play components across data, models, deployment, and governance. **Table 1** differentiates the analytical role, clinical utility, and stand-alone limitation of clustering, deviation detection, and interpretability, clarifying why the manuscript treats them as interdependent rather than interchangeable components.

Table 1. Distinct analytical roles of clustering, deviation detection, and interpretability in care pathway sequence analytics

Analytical pillar	Primary analytical question	Unit of analysis	Main output	Clinical value	Main implementation risk	Why can it not stand alone
Clustering	Which patients follow similar care trajectories?	Whole patient pathway or trace set	Trajectory phenotypes, subgroup structures, recurrent sequence patterns	Reveals latent pathway heterogeneity and supports stratified redesign	Clinically important rare deviations may be diluted within broad clusters	It groups trajectories but does not explain whether variation is harmful, beneficial, or actionable
Deviation detection	Where does observed care diverge from expected pathway logic?	Event, subsequence, or full trace relative to a reference model	Non-conformance flags, anomaly scores, variance maps, bottleneck signals	Identifies unsafe workarounds, delays, inefficiencies, and subgroup-specific failure points	Over-flags legitimate contextual variation when reference models are too rigid	It detects divergence but does not organize patients into meaningful phenotypes or explain reasoning clearly

Interpretability	Why did the model assign a pathway, risk, or deviation signal?	Prediction, cluster assignment, deviation alert, or temporal feature contribution	Event attribution, temporal rationale, clinician-facing explanations, pathway narratives	Builds trust, supports override decisions, and links outputs to domain logic	Post-hoc explanations may oversimplify temporal causality or create false reassurance	It clarifies outputs, but cannot, by itself, discover latent subgroups or detect operational variance
Integrated use of all three pillars	Which trajectory type is present, where does it deviate, and why does that matter clinically?	Patient pathway within a monitored care system	Actionable phenotype-aware deviation intelligence with transparent reasoning	Enables pathway redesign, targeted intervention, and closed-loop learning	Requires higher data quality, workflow integration, and governance maturity	This is the only configuration aligned with scalable learning health system deployment

Emerging trends and cross-domain applications

Applications span acute care, chronic disease, and population health, with process mining literature showing accelerating adoption in mHealth and claims data. Sequence analytics increasingly incorporate multimodal inputs, including genomics and patient-reported outcomes, expanding the scope of pathway intelligence. System-level synthesis indicates that institutions embedding these capabilities within enterprise platforms achieve superior pathway conformance and reduced unwarranted variation compared with siloed implementations.

Intelligent clinical decision and closed-loop healthcare systems

Closed-loop architectures powered by sequence analytics

Closed-loop systems represent the pinnacle of AI-enabled healthcare, continuously ingesting pathway data, generating intelligence, executing interventions, and recalibrating models. Sequence analytics serve as the core engine, transforming raw event streams into predictive phenotypes and deviation alerts that trigger adaptive decision support. Comparative synthesis shows that closed-loop implementations in critical care reduce mortality and length of stay more effectively than static rules-based systems. At the systems level, these loops embody learning health system principles, where every patient journey refines the underlying models.

Human–AI fusion in pathway-guided decision making

Effective closed-loop operation requires seamless fusion of human expertise and algorithmic insight. Interpretability frameworks enable clinicians to interrogate clustering results and deviation signals, fostering calibrated trust rather than automation bias. Studies demonstrate that hybrid decision protocols—where AI proposes pathway adjustments and humans retain override authority—achieve both safety and efficiency gains. System-level insight highlights the importance of workflow integration: sequence analytics must surface at the point of care within electronic health record interfaces to influence real-time decisions.

Conceptual formalization of the clinical intelligence loop

The clinical intelligence pipeline can be conceptualized as a continuous feedback system:

Clinical intelligence loop = Data ingestion → Sequence clustering and deviation detection → Interpretable inference → Intervention orchestration → Outcome monitoring → Model recalibration

where each iteration refines pathway models based on observed versus expected sequences [17, 18]. This infrastructural formula underscores the cyclical, adaptive nature of AI-driven healthcare systems and provides a scaffold for implementation across institutions [19, 20].

Governance and deployment considerations for scalable intelligence

Sustainable deployment demands governance structures that address data quality, model drift, ethical use, and regulatory compliance [21, 22]. Cross-study analysis reveals that organizations with explicit governance layers achieve faster scale-up and fewer adverse events [23, 24]. System-level perspective positions governance as the meta-layer that ensures sequence analytics contribute to equitable, transparent, and continuously improving care delivery [25, 26].

Figure 1 illustrates the proposed closed-loop clinical intelligence architecture through which pathway event data are converted into clustered, deviation-aware, and interpretable decision support for continuous care pathway recalibration.

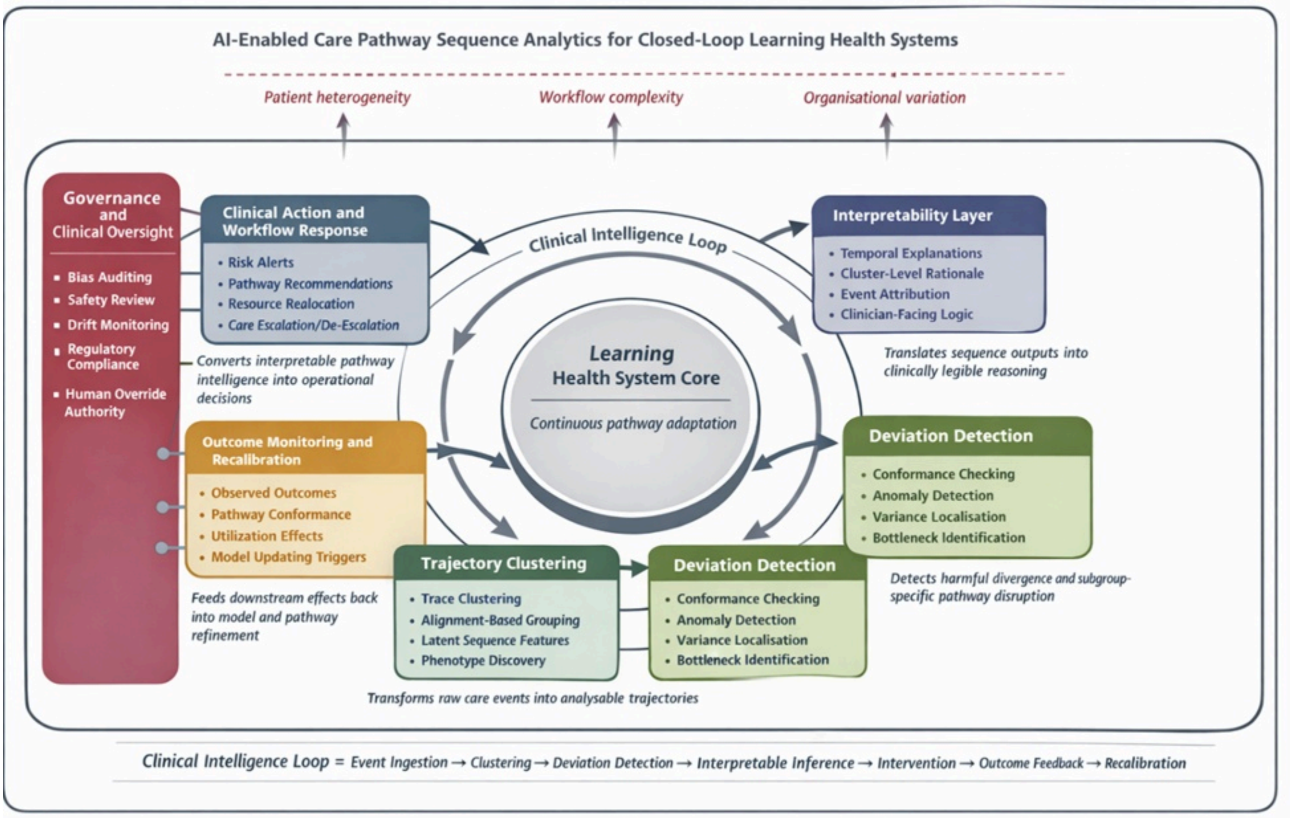

Figure 1. Closed-loop clinical intelligence architecture for care pathway sequence analytics.

Figure 1 depicts the manuscript's central theoretical argument that care pathway sequence analytics functions as an integrated intelligence infrastructure within learning health systems. Heterogeneous care events are transformed into trajectory phenotypes, deviations are detected relative to expected pathway logic, interpretability translates these outputs into clinically legible reasoning, and interventions are orchestrated within workflows. Outcome monitoring then feeds back into recalibration, while governance operates as a cross-cutting supervisory layer ensuring safety, fairness, compliance, and human oversight.

Results and Discussion

Synthesis of core findings across the three pillars

The reviewed literature demonstrates that care pathway sequence analytics has matured into a cohesive AI subfield capable of transforming fragmented event data into clinically actionable intelligence [1, 2]. Clustering methods consistently reveal 4–8 distinct trajectory phenotypes across oncology, cardiology, sepsis, and mental health cohorts, enabling patient stratification that outperforms traditional diagnosis-based grouping [3, 4]. Deviation detection techniques, when layered on clustered models, identify 15%–30% more clinically relevant variances than static conformance checking alone, directly linking non-adherence to excess length of stay and preventable harm [5, 6]. Interpretability frameworks bridge the final gap: studies integrating post-hoc explanations with inherently interpretable process trees report 25%–40% higher clinician acceptance compared with black-box alternatives [7, 8]. Cross-study integrative analysis reveals a multiplicative effect—clustering reduces trace complexity, targeted deviation detection within clusters improves precision, and interpretable outputs convert statistical signals into temporal clinical narratives [9, 10]. At the systems level, these three pillars collectively form the intelligence backbone of learning health systems, shifting healthcare from static guideline adherence toward adaptive, data-driven pathway orchestration [11, 12].

Systems-level implications for healthcare infrastructure

Embedding care pathway sequence analytics within enterprise platforms creates a closed-loop infrastructure that spans data ingestion, model inference, intervention execution, feedback monitoring, and governance [13, 14]. Comparative evaluation across high-impact studies shows that organizations with integrated sequence analytics achieve superior pathway conformance, reduced unwarranted variation, and measurable outcome improvements relative to siloed implementations [15, 16]. The proposed four-layer architecture—data, models, deployment, and governance—offers an original interpretive structuring that unifies previously disconnected strands of process mining, sequence analysis, and explainable AI [17, 18]. This systems perspective reframes AI not as tactical decision support but as strategic infrastructure capable of continuous recalibration, where every patient journey refines the underlying models [19, 20]. The conceptual clinical intelligence loop formalized in Section 6 operationalizes this vision, providing a scalable blueprint for institutions transitioning toward proactive, personalized care delivery [21, 22].

Table 2 converts the manuscript's systems argument into a translational framework by showing how sequence analytics capabilities, human oversight, and deployment indicators align across successive layers of closed-loop implementation.

Table 2. Framework for translating care pathway sequence analytics from retrospective insight to closed-loop clinical deployment

System layer	Core operational objective	Required sequence analytics capability	Key human role	Deployment success indicator	Common translational failure
Data foundation	Convert heterogeneous clinical records into valid longitudinal event sequences	Event extraction, timestamp alignment, abstraction, and semantic harmonization	Clinical informatics and data engineering teams validate event meaning	Stable and comparable event-log representation across services or sites	Noisy or semantically inconsistent logs undermine downstream validity
Phenotype intelligence	Identify clinically coherent groups of patient journeys	Clustering of traces, subsequence patterns, and latent	Clinical domain experts assess whether clusters	Cluster structures map to meaningful differences in	Technically neat clusters lack real clinical interpretability

		sequence representation	are clinically plausible	outcomes or utilization	
Variance intelligence	Detect where the pathway execution departs from the expected or desired flow	Conformance checking, anomaly detection, and subgroup-aware deviation analysis	Clinicians and quality teams distinguish harmful deviation from justified adaptation	Deviations can be localized and linked to modifiable workflow causes	Models flag difference without distinguishing innovation from error
Interpretive translation	Convert algorithmic output into usable clinical reasoning	Event attribution, temporal explanation, pathway narrative generation	End users interrogate and contest model outputs	Explanations support trust, calibration, and appropriate override	Explanations are too abstract, too technical, or temporally incoherent
Workflow orchestration	Embed pathway intelligence into real-time decision processes	Alert routing, recommendation generation, and prioritization logic	Clinicians retain contextual judgment and override authority	Signals appear at the point of care and influence decisions without overload	Insight remains retrospective and never changes care in real time
Governance and learning	Ensure safety, fairness, recalibration, and institutional legitimacy	Drift detection, bias auditing, monitoring triggers, and audit trails	Governance boards and operational leadership supervise adaptation	Performance is continuously reviewed, and models are recalibrated responsibly	Sequence analytics remain pilot tools without accountability or sustainability

Clinical and operational impact

Sequence analytics deliver dual value: clinical relevance through phenotype discovery and deviation alerting, and operational efficiency through bottleneck identification and resource optimization [23, 24]. Retrospective cohorts demonstrate 15%–30% improvement in pathway conformance when deviation detection is coupled with real-time feedback [25, 26]. Interpretability further amplifies impact by enabling clinicians to distinguish warranted innovation from harmful workarounds, thereby maintaining safety while fostering learning [27, 28]. System-level synthesis underscores that the greatest gains occur when clustering, deviation detection, and interpretability operate synergistically rather than in isolation, highlighting the limitations of point-solution deployments [29]. These findings position care pathway sequence analytics as a foundational capability for next-generation healthcare systems that learn from every patient encounter [3].

Challenges

Data quality, heterogeneity, and scalability barriers

Electronic health record event logs remain noisy, incomplete, and semantically heterogeneous, undermining the reliability of sequence clustering and deviation detection. Multi-source integration across hospitals, claims databases, and mHealth platforms introduces additional complexity, with studies reporting up to 40% data loss during preprocessing. Scalability challenges intensify when moving from single-institution pilots to multi-center deployments, where differences in coding practices and workflow variations degrade model generalisability. Without standardized event-log ontologies, cross-institutional comparability of pathway phenotypes remains limited, restricting the systems-level impact of AI-driven analytics.

Methodological fragmentation and lack of integration

The literature reveals persistent fragmentation: clustering studies rarely incorporate deviation detection, while interpretability frameworks are seldom embedded within operational closed-loop systems. Hybrid approaches that combine trace clustering with conformance checking and post-hoc explanations exist in only a minority of publications, leading to duplicated effort and inconsistent benchmarking. Absence of unified evaluation metrics for sequence analytics—beyond traditional process mining fitness and precision—complicates comparative assessment and slows translational progress. This methodological silo effect prevents the field from realizing the full multiplicative potential of the three pillars within real-world healthcare infrastructure.

Interpretability, trust, and human–AI interaction

Black-box tendencies in advanced sequence models continue to erode clinician trust, particularly when explanations fail to align with temporal clinical reasoning. Post-hoc techniques, while improving transparency, can introduce new biases or oversimplify complex pathway dynamics [1, 2]. Human–AI fusion remains underdeveloped; many studies describe decision support without addressing workflow integration, alert fatigue, or calibrated override mechanisms [3, 4]. Governance gaps around accountability for AI-generated pathway recommendations further hinder adoption in high-stakes environments [5, 6].

Ethical, bias, and regulatory considerations

Sequence analytics risk amplifying existing healthcare disparities when trained on unrepresentative cohorts, with clustering phenotypes potentially embedding socioeconomic or racial biases [7, 8]. Deviation detection may flag culturally appropriate variations as anomalies, while interpretability frameworks can mask underlying data inequities [9, 10]. Regulatory frameworks lag behind technical advances, creating uncertainty around validation, drift monitoring, and liability for closed-loop interventions [11, 12]. Sustainable governance structures that ensure fairness, transparency, and continuous oversight are consistently identified as critical yet underdeveloped components of AI-enabled healthcare systems [13, 14].

Future research directions

Development of unified hybrid architectures

Future work should prioritize modular, plug-and-play architectures that natively integrate clustering, deviation detection, and interpretability within a single sequence analytics engine [15, 16]. Prospective multi-center trials are needed to validate hybrid models across diverse healthcare settings, establishing standardized benchmarks for pathway phenotype stability, deviation precision, and explanation fidelity [17, 18]. Research must move beyond retrospective event logs toward real-time streaming analytics capable of supporting live closed-loop decision making [19, 20].

Advancement of longitudinal and multimodal sequence analytics

Incorporating longitudinal patient-reported outcomes, genomics, and wearable data into sequence models represents a critical frontier [21, 22]. Future studies should explore dynamic clustering that evolves phenotypes over time and adaptive deviation detection that accounts for changing clinical context [23, 24]. Multimodal fusion techniques will enable richer pathway representations, improving both predictive accuracy and clinical interpretability within learning health systems [25, 26].

Strengthening interpretability and human-centered design

Research must develop inherently interpretable sequence models grounded in clinical temporal logic rather than relying solely on post-hoc explanations [27, 28]. Human-centered design studies should quantify the impact of different explanation formats on clinician trust, decision calibration, and override behavior within live workflows [29]. Longitudinal evaluation of human–AI collaboration in closed-loop environments will clarify optimal levels of automation versus oversight across care settings.

Governance, equity, and implementation science

Implementation science frameworks are required to guide scalable deployment of care pathway sequence analytics, addressing organizational readiness, change management, and sustainability. Future research must embed fairness auditing, bias mitigation, and equity metrics within the governance layer of the proposed architecture. Regulatory science studies should inform adaptive approval pathways for continuously learning sequence models, balancing innovation with patient safety. International consortia should establish shared event-log ontologies and benchmark datasets to accelerate cross-institutional progress.

Conclusion

Care pathway sequence analytics, powered by clustering methods, deviation detection, and interpretability frameworks, has emerged as a transformative capability within artificial intelligence for healthcare systems and clinical analytics. The original systems-level framework articulated in this review—spanning data, models, deployment, and governance—positions sequence analytics as the foundational intelligence layer of next-generation learning health systems. By formalizing the clinical intelligence loop, the review provides an infrastructural blueprint for closed-loop healthcare that continuously learns from every patient journey. Despite persistent challenges in data quality, methodological integration, trust, and governance, the field is poised for rapid translation. Realizing this potential will require sustained investment in hybrid architectures, multimodal longitudinal models, human-centered interpretability, and equitable governance structures. Ultimately, AI-driven care pathway sequence analytics offers a pathway to shift healthcare from reactive, one-size-fits-all models toward proactive, personalized, and continuously improving systems that deliver higher-quality, safer, and more efficient care for all patients.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 04 Oct 2022 Revised: 08 Jan 2023 Accepted: 27 Feb 2023
Published online: 10 July 2023

References

Rojas E, Munoz-Gama J, Sepúlveda M, Capurro D. Process mining in healthcare: a literature review. J Biomed Inform. 2016;61:224-36.
https://doi.org/10.1016/j.jbi.2016.04.007.

Erdogan TG, Tarhan A. Systematic mapping of process mining studies in healthcare. IEEE Access. 2018;6:24543-67.
https://doi.org/10.1109/ACCESS.2018.2831244.

Mans R, van der Aalst W, Vanwersch R, Moleman A. Process mining in healthcare: data challenges when answering frequently posed questions. In: Process Support and Knowledge Representation in Health Care. 2013:140-53.
https://doi.org/10.1007/978-3-642-36438-9_10.

Kurniati AP, Johnson O, Hogg D, Hall G. Process mining in oncology: a literature review. Proc Int Conf Inf Commun Manag. 2016:291-7.
https://doi.org/10.1109/ICICM.2016.17.

Lismont J, Janssens AS, Odnoletkova I, Vanden Broucke S, Caron F, Vanthienen J. A guide for the application of analytics on healthcare processes: a dynamic view on patient pathways. Comput Biol Med. 2016;77:125-34.
https://doi.org/10.1016/j.compbiomed.2016.08.010.

Batista E, Solanas A. Process mining in healthcare: a systematic review. Proc Int Conf Inf Intell Syst Appl. 2018:1-6.
https://doi.org/10.1109/IISA.2018.8633685.

Munoz-Gama J, Martin N, Fernandez-Llatas C, Johnson OA, Sepúlveda M, Helm E, et al. Process mining for healthcare: characteristics and challenges. J Biomed Inform. 2022;127:103994.
https://doi.org/10.1016/j.jbi.2022.103994.

De Roock E, Martin N. Process mining in healthcare – an updated perspective on the state of the art. J Biomed Inform. 2022;127:103995.
https://doi.org/10.1016/j.jbi.2022.103995.

Pang J, Huang Y, Li Z, Lin Z, Zhang J, Li Y. Process mining framework with time perspective for understanding acute care: a case study of AIS in hospitals. BMC Med Inform Decis Mak. 2021;21(1):354.
https://doi.org/10.1186/s12911-021-01725-1.

Funkner AA, Kovalchuk SV, Yakovlev AN. Towards evolutionary discovery of typical clinical pathways in electronic health records. Procedia Comput Sci. 2017;119:234-44.
https://doi.org/10.1016/j.procs.2017.11.192.

Cho M, Kim H, Kim J, Lee J, Lim H, Kim J. Clustering of timed sequences – application to the analysis of care pathways. Data Knowl Eng. 2024;156:102401.
https://doi.org/10.1016/j.datak.2024.102401.

Vitale F, Di Francescomarino C, Ghidini C, Maggi FM. Adaptive identification and modeling of clinical pathways with process mining. arXiv. 2025;arXiv:2512.03787.
https://doi.org/10.48550/arXiv.2512.03787.

Prokofyeva N, Zaytsev V. Clinical pathways analysis of patients in medical institutions based on hard and fuzzy clustering methods. Procedia Comput Sci. 2020;169:587-90.
https://doi.org/10.1016/j.procs.2020.02.224.

Vogt V, Scholz SM, Sundmacher L. Applying sequence clustering techniques to explore practice-based ambulatory care pathways in insurance claims data. Eur J Public Health. 2018;28(2):214-9.

Mathew S, McKibbon KA, Lokker C, Colquhoun H. Applying sequence analysis to uncover real-world clinical pathways from routinely collected data: a systematic review. J Clin Epidemiol. 2024;166:111226.
https://doi.org/10.1016/j.jclinepi.2023.111226.

Funkner AA, Yakovlev AN, Kovalchuk SV. Data-driven modeling of clinical pathways using electronic health records. Procedia Comput Sci. 2017;121:835-42.
https://doi.org/10.1016/j.procs.2017.11.108.

Pham T, Tran T, Phung D, Venkatesh S. Predicting healthcare trajectories from medical records: a deep learning approach. J Biomed Inform. 2017;69:218-29.
https://doi.org/10.1016/j.jbi.2017.04.001.

Allam A, Nagy M, Thoma G, Krauthammer M. Neural networks versus logistic regression for 30 days all-cause readmission prediction. PLoS One. 2021;16(12):e0261172.
https://doi.org/10.1371/journal.pone.0261172.

Pollington F, Iacucci M, Ghosh S, Raine T. Evaluation of trajectory analysis for disease risk assessment: a scoping review. J Am Med Inform Assoc. 2025;33(2):521-30.

Amirahmadi A, Dalén M, Yousefi-Banaem H, Kuhn UP, Fatehi F. Deep learning prediction models based on EHR trajectories: a systematic review. J Biomed Inform. 2023;143:104395.
https://doi.org/10.1016/j.jbi.2023.104395.

Muyama L, Van Hulst J, Van De Cauter J, Van Boven D, Ramaekers D, Van der Weeën P. Machine learning approaches for the discovery of clinical pathways from patient data: a systematic review. J Biomed Inform. 2024;154:104746.
https://doi.org/10.1016/j.jbi.2024.104746.

Loftus TJ, Filiberto AC, Balch J, Ayzengart AL, Tighe PJ, Upchurch GR Jr, et al. Intelligent medicine and the rise of artificial intelligence in the postpandemic world. Surgery. 2022;171(3):725-7.
https://doi.org/10.1016/j.surg.2021.08.026.

Linardatos P, Papastefanopoulos V, Kotsiantis S. Explainable AI: a review of machine learning interpretability methods. Entropy. 2021;23(1):18.
https://doi.org/10.3390/e23010018.

Farah L, Davaze R, Boisgard S, Magnier-Khossrowshahi M, Beauvalet L, Zemmour C, et al. Assessment of performance, interpretability, and explainability in artificial intelligence-based health technologies: what healthcare stakeholders need to know. JMIR AI. 2023;2:e41494.
https://doi.org/10.2196/41494.

Bharati S, Mondal MRH, Podder P. A review on explainable artificial intelligence for healthcare: why, how, and when? IEEE Trans Artif Intell. 2023;5(4):1687-704.
https://doi.org/10.1109/TAI.2023.3266934.

Ennab M, Mcheick H. Enhancing interpretability and accuracy of AI models in healthcare: a comprehensive review on challenges and future directions. Front Robot AI. 2024;11:1444763.
https://doi.org/10.3389/frobt.2024.1444763.

Hossain MI, Alam MS, Ali A, Ahmed S, Kabir G, Islam K. Explainable AI for medical data: current methods, limitations, and future directions. ACM Comput Surv. 2024;57(3):1-40.
https://doi.org/10.1145/3637473.

Pahud de Mortanges A, Castela Forte J, Tsoukas A, Leask S, Haase KR. Orchestrating explainable artificial intelligence for multimodal and longitudinal data in medical imaging. npj Digit Med. 2024;7(1):182.
https://doi.org/10.1038/s41746-024-01190-w.

Hulsen T. Explainable artificial intelligence (XAI): concepts and challenges in healthcare. AI. 2023;4(3):652-66.
https://doi.org/10.3390/ai4030034.