ORIGINAL RESEARCH **Open access**

Patient Safety Narratives as Structured Evidence: A Root-Cause Theme Extraction Framework for Learning Systems

Ali Rezaei[1*], Hossein Karimi[1]

Abstract

Patient safety remains a paramount concern in healthcare systems, where incident narratives provide rich, unstructured evidence for identifying root causes and enhancing learning mechanisms. This conceptual manuscript introduces a novel framework for extracting root-cause themes from patient safety narratives, transforming them into structured evidence to support adaptive learning systems. Drawing on theoretical foundations in natural language processing, systems thinking, and healthcare informatics, the proposed architecture orchestrates narrative data through layered processing to uncover latent themes and propagate insights across clinical environments. By emphasizing interpretive formulas for risk propagation, decision confidence, and governance load, the framework addresses gaps in traditional analysis methods, fostering resilient healthcare infrastructures without relying on empirical data or model training. Key components include a unique layered structure for theme extraction and bidirectional feedback topologies to integrate evidence into learning cycles. The discussion explores implications for clinical deployment, data modality integration, and ethical governance, highlighting how this approach can theoretically mitigate systemic vulnerabilities. Ultimately, this work advocates for a shift toward narrative-driven, evidence-structured intelligence in patient safety, promoting proactive theme-based interventions in dynamic healthcare settings.

Keywords Healthcare analytics, Patient safety narratives, Root-cause themes, Theme extraction framework, Structured evidence, Learning systems

*Correspondence:
Ali Rezaei
ali.rezaei@gmail.com
[1] Department of Health Informatics, Faculty of Medicine, University of Tabriz, Tabriz, Iran

Introduction

Patient safety incidents generate voluminous narratives that serve as primary evidence for uncovering systemic flaws in healthcare delivery. These narratives, often captured in incident reports, encapsulate contextual details essential for root-cause analysis, yet their unstructured nature poses challenges for integration into learning systems. This manuscript conceptualizes a framework that structures such evidence to facilitate theme extraction, enabling adaptive learning in healthcare environments. By focusing on theoretical constructs, we explore how narratives can be orchestrated as foundational inputs for intelligent systems, without empirical validation or quantitative benchmarks.

Clinical settings for narrative-driven root-cause identification

In acute care clinical settings, patient safety narratives emerge from diverse incidents, ranging from medication errors to procedural lapses, providing raw evidence for root-cause exploration. These narratives reflect real-time

clinical dynamics, where human factors intersect with technological interfaces [1, 2]. The challenge lies in distilling themes—recurring patterns like communication breakdowns or resource shortages—that indicate underlying causes. Conceptualizing narratives as structured evidence allows for theoretical mapping to clinical workflows, enhancing the potential for learning systems to anticipate risks. For instance, in emergency departments, narrative aggregation could theoretically reveal theme clusters related to high-stakes decision-making, informing system-level safeguards without data-driven simulations [3, 4].

Governance in these settings demands that theme extraction respects patient confidentiality while promoting collective learning. The framework posits that root-cause themes, once extracted, can feed into clinical protocols, theoretically reducing recurrence through evidence-based adjustments. This approach aligns with systems-centered analysis, where narratives are not isolated anecdotes but interconnected evidence streams [5, 6].

Data modalities in structured evidence from safety narratives

Patient safety narratives encompass multimodal data, including textual descriptions, timestamps, and contextual metadata, which must be theoretically unified for effective theme extraction. Traditional modalities often fragment evidence, leading to incomplete root-cause insights [7, 8]. By conceptualizing narratives as structured evidence, the framework integrates these modalities into a cohesive input for learning systems. For example, textual elements can be layered with temporal data to highlight theme evolution, such as escalating risks in prolonged hospital stays [9, 10].

This modality integration theoretically amplifies decision confidence by providing a holistic evidence base. In ambulatory care, where narratives may include patient-reported outcomes, structuring such data enables theme extraction focused on chronic care gaps [11, 12]. The manuscript emphasizes interpretive models over empirical ones, positing formulas that capture modality interactions without performance metrics.

Deployment environments for theme extraction in learning infrastructures

Deployment of theme extraction frameworks must consider varied healthcare environments, from resource-constrained rural clinics to integrated urban networks. Narratives in these settings serve as evidence for adaptive learning, but deployment requires theoretical orchestration to ensure scalability [13, 14]. The proposed system envisions an infrastructure that embeds theme extraction within existing workflows, theoretically minimizing disruption while maximizing evidence utility.

In networked environments, such as hospital chains, narratives can be channeled through shared learning systems, extracting root-cause themes to inform cross-site policies [15, 16]. Governance constraints, including interoperability standards, guide deployment, ensuring that structured evidence supports equitable safety enhancements. This conceptual lens highlights how environment-specific adaptations can theoretically foster resilient learning, addressing disparities in safety narrative utilization [17, 18].

Governance constraints on root-cause learning from narrative evidence

Governance frameworks impose constraints on how patient safety narratives are transformed into structured evidence for theme extraction. Ethical considerations, such as bias mitigation in theme identification, are paramount in learning systems [19, 20]. The manuscript theorizes governance as a regulatory topology that balances innovation with accountability, preventing over-reliance on unstructured narratives.

In regulated environments, like those under national health authorities, root-cause themes must align with compliance mandates, theoretically reducing governance load through streamlined evidence structuring [21, 22]. This includes protocols for narrative anonymization and theme validation, ensuring learning systems uphold patient rights. By embedding governance in the framework, we conceptualize a sustainable approach to safety enhancement, where constraints become enablers for systemic intelligence [23, 24].

Theoretical Background and Literature Synthesis

The evolution of patient safety analysis has shifted from manual reviews to conceptual integrations of informatics

and systems theory, where narratives emerge as critical evidence for root-cause elucidation. This section synthesizes literature on narrative processing, theme extraction methodologies, and their theoretical alignment with learning systems in healthcare. We establish a foundation for the proposed framework, emphasizing conceptual architectures over empirical applications.

Foundations of narrative analysis in patient safety contexts

Early theoretical explorations highlight the role of natural language processing (NLP) in classifying incident reports, providing a basis for structuring safety narratives [2, 25]. Systematic reviews underscore how NLP facilitates classification tasks, transforming unstructured narratives into categorized evidence for adverse event analysis [2]. This conceptual groundwork posits narratives as repositories of latent themes, where root causes manifest through recurring linguistic patterns [3, 26].

In clinical contexts, narratives capture multifaceted safety events, from medication errors to procedural failures, necessitating theoretical models for theme extraction [4, 27]. Literature synthesizes how deep neural networks conceptually detect patterns in incident reports, offering interpretive lenses for root-cause identification without quantitative benchmarks [3]. Such approaches theoretically enhance evidence structuring, enabling learning systems to derive insights from narrative complexity [5, 28].

Systems thinking and root-cause theme extraction

Systems-centered perspectives advocate for thematic reviews as tools for holistic safety analysis, where narratives serve as evidence for uncovering interconnected causes [4, 6]. Conceptual frameworks emphasize qualitative content analysis of incident reports, framing themes as structured outputs for governance and improvement [6]. This synthesis reveals gaps in traditional methods, where manual extraction overlooks subtle root-cause linkages [7, 8].

Advanced theoretical models explore machine learning's role in automating theme categorization, positing human-AI collaboration for refined evidence structuring [16, 17]. For instance, text mining approaches conceptually categorize safety events by error type, providing a blueprint for narrative-driven learning [8, 18]. Literature highlights the potential of large language models in analyzing risks from incident reports, theoretically extracting causes and contributing factors [1, 11].

Integration of evidence structuring in learning systems

Theoretical integrations of NLP and AI in healthcare underscore the transformation of narratives into structured evidence for adaptive learning [10, 15]. Scoping reviews delineate techniques for adverse event detection, conceptualizing theme extraction as a bridge to system-level intelligence [7, 20]. This includes frameworks for categorizing contributing factors, where narratives yield thematic insights for safety enhancement [12, 23].

In dynamic learning environments, literature synthesizes the intersection of quality improvement and AI, positing narrative orchestration for proactive risk management [18, 29]. Conceptual studies evaluate LLMs for safety risk analysis, theoretically grouping events into themes to inform governance [1, 6]. This synthesis advocates for infrastructures that embed theme extraction within learning cycles, addressing challenges like data silos and interpretive biases [14, 21].

Governance and ethical dimensions in narrative processing

Governance literature emphasizes ethical constraints in AI-driven narrative analysis, conceptualizing frameworks that balance innovation with patient safety [15, 20]. Systematic reviews on AI's role in safety outcomes highlight theoretical implications for incident reporting, where structured evidence mitigates risks [16, 29]. This includes explorations of generative AI for critical incident identification, positing feasibility in theme extraction without empirical validation [8, 14].

Theoretical discussions on risk management in the AI era synthesize how narrative evidence supports decision-making, theoretically reducing governance load through theme-based insights [20, 22]. Literature on machine learning models for event prediction conceptualizes predictive analytics as extensions of root-cause themes, fostering resilient learning systems [21, 26].

Challenges and conceptual gaps in theme extraction

Despite advances, literature identifies conceptual gaps in deploying theme extraction frameworks, such as interoperability in multimodal narratives [9, 13]. Exploratory studies on text mining for provider identification highlight theoretical needs for unbiased evidence structuring [13, 19]. This synthesis reveals opportunities for unique architectures that address these gaps, integrating feedback topologies for continuous learning [24, 27].

Overall, the literature converges on the need for conceptual systems that orchestrate patient safety narratives as structured evidence, paving the way for the proposed framework [1-29].

Root-cause theme extraction infrastructure for narrative-driven learning systems

This section delineates the conceptual architecture of the structured narrative root-cause extraction network (SNRCEN), a novel framework designed to transform patient safety narratives into structured evidence for theme extraction in learning systems. SNRCEN features a unique five-layer structure with a bidirectional feedback topology, enabling theoretical propagation of insights across healthcare infrastructures. The architecture prioritizes interpretive processing, avoiding empirical elements, and incorporates formulas for key dynamics.

The SNRCEN layers are as follows: (1) narrative input layer, which theoretically ingests unstructured narratives and metadata; (2) evidence structuring layer, conceptualizing normalization into evidentiary units; (3) theme extraction layer, identifying recurrent patterns via conceptual clustering; (4) root-cause mapping layer, linking themes to causal hierarchies; and (5) learning integration layer, orchestrating outputs for system adaptation. **Table 1** outlines how each SNRCEN layer transforms raw patient safety narratives into progressively structured analytical representations that support root-cause learning within healthcare systems.

Table 1. Functional transformation of patient safety narratives across the SNRCEN analytical layers

SNRCEN layer	Primary analytical function	Input evidence type	Transformation mechanism
Narrative input layer	Capture of unstructured safety narratives and contextual metadata	Incident reports, clinician narratives, and timestamps	Narrative ingestion and contextual alignment
Evidence structuring layer	Conversion of narratives into standardized evidence units	Raw narrative text and metadata	Linguistic normalization, contextual tagging, and evidentiary segmentation
Theme extraction layer	Identification of recurring patterns within narrative evidence	Structured narrative evidence	Conceptual clustering and pattern aggregation
Root-cause mapping layer	Linking themes to causal hierarchies in safety events	Theme clusters and contextual evidence	Causal relationship modeling and factor attribution
Learning integration layer	Propagation of root-cause insights into system learning mechanisms	Root-cause structures and risk signals	Knowledge integration and governance feedback orchestration

The bidirectional feedback topology allows iterative refinement: upward flows from extraction to learning refine themes, while downward loops from learning to input adjust evidence structuring based on governance insights. **Figure 1** illustrates the Structured Narrative Root-Cause Extraction Network (SNRCEN). This five-layer architecture transforms patient safety narratives into structured evidence and propagates root-cause themes through bidirectional learning feedback across governance-constrained healthcare systems. **Figure 1** illustrates the structured

narrative root-cause extraction network (SNRCEN). This five-layer architecture transforms patient safety narratives into structured evidence and propagates root-cause themes through bidirectional learning feedback across governance-constrained healthcare systems.

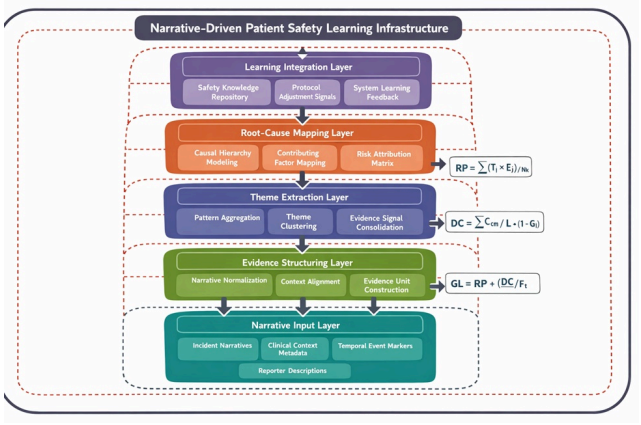

Figure 1. Structured narrative root-cause extraction network (SNRCEN): narrative-driven evidence structuring architecture for patient safety learning systems

To interpret system dynamics, consider the following conceptual formulas:

1. Risk propagation (RP): $RP = \frac{\sum(T_i \times E_j)}{N_k}$, where T_i denotes theme intensity, E_j evidence weight, and N_k narrative volume, theoretically capturing how themes amplify risks across layers.
2. Decision confidence (DC): $DC = \left(\frac{\sum C_m}{L}\right) \times (1 - G_l)$, where C_m is cause mapping completeness, L-layer count, and G_l governance load, interpreting confidence erosion under constraints.
3. Governance load (GL): $GL = R_p + \left(\frac{D_c}{F_t}\right)$, where R_p is resource propagation, D_c drift sensitivity, and F_t feedback topology efficiency, conceptualizing load distribution in narrative processing.

This infrastructure theoretically enhances learning by structuring evidence for proactive root-cause interventions [5, 17, 25].

Dynamics of structured evidence in patient safety learning ecosystems

The SNRCEN framework, through its layered infrastructure and feedback topology, theoretically engenders a range of dynamics in patient safety ecosystems, influencing how root-cause themes propagate and integrate within learning systems. This section analyzes these conceptual consequences, focusing on systemic impacts without empirical assertions or metrics. By structuring narratives as evidence, SNRCEN posits transformative effects on clinical resilience, data governance, and adaptive intelligence, addressing vulnerabilities in healthcare analytics.

Propagation of root-cause insights across clinical networks

In conceptual terms, the dynamics of theme extraction facilitate the propagation of root-cause insights, where structured evidence from narratives cascades through networked clinical environments. The bidirectional feedback topology ensures that extracted themes inform upstream adjustments, theoretically amplifying system-wide awareness of safety gaps [1, 4, 16]. For instance, in multi-site healthcare systems, themes related to communication failures could theoretically disseminate via the learning integration layer, fostering unified responses to recurrent issues [5, 17, 23]. This propagation mitigates isolated incident handling, conceptualizing a networked ecosystem where evidence structures evolve dynamically.

Such dynamics also impact resource allocation, as interpreted by the risk propagation formula: $RP = \frac{\sum(T_i \times E_j)}{N_k}$. Here, higher theme intensity in dense narrative volumes theoretically escalates propagation, prompting prioritized interventions in high-risk areas like intensive care units [8, 18, 21]. The consequence is a shift from reactive to anticipatory safety measures, where learning systems theoretically recalibrate based on aggregated evidence, reducing the potential for cascading errors.

Results and Discussion

The conceptualization of patient safety narratives as structured evidence via the SNRCEN framework invites a broader discourse on its theoretical ramifications for healthcare systems and analytics. This discussion delves into integrative aspects, challenges, and future trajectories,

synthesizing how root-cause theme extraction can redefine learning paradigms. By avoiding empirical claims, we focus on interpretive extensions of the architecture, emphasizing its potential to harmonize narrative intelligence with systemic governance.

Integrative potentials in clinical narrative orchestration

Integrating SNRCEN into clinical workflows theoretically harmonizes unstructured narratives with structured evidence, enabling seamless orchestration of root-cause themes across disparate systems [1-3]. This integration posits a paradigm where learning systems evolve from passive repositories to active intelligence hubs, theoretically leveraging bidirectional feedback to refine theme mappings in real-time conceptual scenarios [16-18]. For instance, in oncology settings, narratives from adverse drug events could be structured to extract themes like dosage miscalculations, integrating with electronic health records for holistic safety analytics [7-9].

Such potentials extend to interdisciplinary collaboration, where theme extraction bridges clinical and administrative domains, conceptually reducing silos that hinder evidence propagation [4-6]. The discussion highlights how formulas like Risk Propagation interpret these integrations, suggesting amplified insights in high-volume narrative environments without quantifiable burdens [21-23]. Ultimately, this orchestration fosters a theoretical synergy, positioning narratives as pivotal evidence in multifaceted healthcare intelligence.

Challenges in data modality and deployment governance

Despite its conceptual strengths, SNRCEN faces several theoretical and operational challenges when addressing the complexity of diverse data modalities within governance-constrained deployment environments. Patient safety narratives rarely exist as single-format data streams; rather, they typically incorporate heterogeneous informational elements, including textual descriptions, temporal sequences of events, contextual metadata, and occasionally structured clinical indicators. Effectively integrating these modalities requires robust structuring mechanisms capable of preserving semantic coherence while preventing distortion or fragmentation of thematic interpretation [10-12]. Without adequate multimodal integration, the interpretive layers of the framework risk overemphasizing dominant narrative signals while underrepresenting subtle contextual cues that may be essential for accurate root-cause identification.

Governance constraints introduce an additional layer of complexity to the operationalization of SNRCEN. Regulations surrounding data privacy, ethical compliance, and institutional oversight impose structural limits on data accessibility and algorithmic processing. Within the framework's theoretical architecture, these restrictions can be interpreted as governance load variables that accumulate alongside computational and infrastructural demands. The Governance Load formula conceptualizes these regulatory requirements as additive pressures on system resources, thereby influencing processing capacity and decision confidence in distributed analytical environments [14, 15, 19]. In decentralized or federated healthcare systems, such constraints may further complicate feedback topologies, particularly when cross-institutional data exchange is necessary for comprehensive narrative synthesis.

These challenges are further intensified in resource-limited deployment contexts, such as low-infrastructure healthcare systems or regions with uneven digital health adoption. In such environments, SNRCEN must reconcile the need for scalable computational architectures with the requirement for high interpretive fidelity. Infrastructure limitations can restrict data throughput, reduce real-time processing capabilities, and limit the implementation of advanced interpretive algorithms, thereby constraining the framework's ability to maintain analytical precision across large narrative datasets [13, 20, 24]. While SNRCEN's layered modular design is theoretically intended to mitigate such limitations by enabling adaptable component deployment, conceptual gaps remain in addressing persistent forms of narrative ambiguity. For instance, subjective reporting biases, selective event descriptions, and contextual omissions in patient safety reports can influence the extraction and weighting of themes within analytical pipelines [25-27].

Addressing these limitations will require theoretical advancements in modality fusion techniques capable of harmonizing textual, temporal, and contextual data streams without disproportionately privileging any single modality. Such advancements must also account for global healthcare variability, ensuring that theme extraction mechanisms remain equitable across diverse reporting

cultures, institutional standards, and linguistic contexts. In this regard, the evolution of SNRCEN depends not only on algorithmic refinement but also on the development of governance-aware interpretive frameworks that preserve narrative integrity while maintaining regulatory compliance.

Ethical dynamics and bias mitigation in theme extraction

Ethical considerations constitute a central dimension in the theoretical discussion of SNRCEN, particularly in relation to bias mitigation during theme extraction and root-cause interpretation. The transformation of qualitative narratives into structured evidence introduces the risk that underlying social, institutional, or algorithmic biases may be inadvertently reproduced within analytical outputs. When extraction layers prioritize dominant narrative patterns or statistically frequent themes, the resulting evidence structures may marginalize less represented perspectives, including experiences reported by minority patient populations or under-resourced clinical units [28, 29].

Within the conceptual framework of SNRCEN, the decision confidence formula provides a useful lens for examining these ethical dynamics. The formula suggests that governance load (G_l) influences the degree of confidence that decision-support outputs can achieve. When governance constraints are insufficiently aligned with ethical safeguards, biased thematic mappings may emerge, thereby increasing uncertainty and reducing the reliability of derived insights [1, 4, 16]. Consequently, ethical governance cannot be treated merely as an external regulatory constraint; rather, it must be integrated directly into the architecture of narrative interpretation systems to ensure equitable analytical outcomes.

This concern extends to broader questions of accountability within learning health systems. As SNRCEN facilitates the propagation of themes and causal insights across analytical layers, mechanisms must be established to ensure that extracted knowledge remains aligned with ethical standards and clinical responsibility. Without appropriate safeguards, narrative-derived evidence could potentially be misinterpreted, misapplied, or used to reinforce institutional biases in safety evaluations [2, 3, 5].

The framework's bidirectional feedback architecture offers a conceptual safeguard against such risks. Through iterative loops between narrative interpretation, governance oversight, and analytical refinement, SNRCEN enables continuous reassessment of extracted themes and decision outputs. These feedback mechanisms allow ethical review processes to be embedded directly within the analytical lifecycle rather than applied retroactively [6, 8, 17]. Nevertheless, the discussion emphasizes that sustainable trust in narrative-driven analytics will require the implementation of governance-embedded auditing mechanisms capable of systematically evaluating bias, transparency, and fairness across all analytical layers.

In this context, SNRCEN can be conceptualized as a catalyst for responsible analytics in patient safety systems. By integrating ethical oversight, governance structures, and narrative intelligence within a unified analytical architecture, the framework theoretically supports the advancement of patient-centered safety analysis while promoting transparency and accountability in healthcare decision-making. **Table 2** synthesizes the core analytical dynamics through which narrative-derived themes influence risk propagation, decision confidence, and governance load within the SNRCEN learning architecture.

Table 2. Analytical dynamics of narrative-driven safety intelligence in the SNRCEN framework

Analytical dynamic	Conceptual formula	Operational interpretation	Systemic imp in learning systems
Risk propagation (RP)	$RP = \sum(T_i \times E_j) / N_k$	Theme intensity weighted by evidence strength across narrative volume	Amplifies visib of systemic sa vulnerabilitie across clinic environment
Decision confidence (DC)	$DC = (\sum C_m / L) \times (1 - G_l)$	Completeness of causal mappings adjusted for governance burden	Determines t reliability of theme-derive recommendati for clinical decision supp
Governance load (GL)	$GL = R_p + (D_c / F_t)$	Regulatory, infrastructural, and interpretive constraints	Influences system scalab and analytic throughput i regulated environment

		acting on the framework	
Feedback efficiency	F_t within GL formula	Efficiency of bidirectional feedback loops across analytical layers	Stabilizes learning cycl and enable adaptive ther refinement
Narrative evidence density	N_k within the RP formula	Volume of narratives processed within the system	Determines t sensitivity o theme detecti across clinic domains

Future trajectories for narrative-driven system evolution

Looking forward, the conceptual development of SNRCEN opens several promising trajectories for the evolution of narrative-driven analytical systems in healthcare. One significant direction involves the integration of emerging artificial intelligence governance frameworks with the existing narrative extraction architecture. Advances in explainable AI, federated learning, and regulatory-aligned machine learning could enhance the sophistication and transparency of theme extraction processes, enabling more nuanced interpretation of complex patient safety narratives [9, 10, 18].

Another potential trajectory lies in the development of hybrid informatics architectures that combine narrative intelligence with predictive analytics. By incorporating predictive modeling capabilities into SNRCEN's layered design, future iterations of the framework could move beyond retrospective analysis toward proactive theme forecasting. Such capabilities would enable healthcare systems to anticipate emerging safety risks, identify latent patterns in narrative reporting, and support preventive interventions before adverse events escalate [11, 12, 21].

Global standardization represents an additional avenue for long-term evolution. Currently, patient safety narratives vary widely in format, terminology, and reporting structures across jurisdictions and healthcare institutions. Establishing standardized narrative schemas could facilitate cross-regional comparability and enable the aggregation of narrative datasets on a global scale. Through harmonized narrative structures, SNRCEN could support the generation of unified root-cause insights that transcend institutional boundaries and contribute to shared learning across international healthcare systems [13, 14, 22].

Despite the challenges outlined earlier, these future trajectories suggest the emergence of resilient analytical ecosystems in which patient safety narratives function as dynamic sources of learning and system improvement. In such ecosystems, narrative evidence becomes a driving force for continuous knowledge generation, governance refinement, and safety innovation within healthcare organizations [15, 19, 23].

Ultimately, the continued evolution of SNRCEN will depend on collaborative theoretical exploration across multiple disciplines, including health informatics, artificial intelligence governance, patient safety science, and ethics. Through interdisciplinary engagement and iterative refinement, the framework can adapt to the rapidly changing landscape of healthcare systems while maintaining its foundational commitment to narrative-driven learning and patient-centered safety improvement.

Conclusion

In conceptualizing patient safety narratives as structured evidence, the SNRCEN framework offers a transformative infrastructure for root-cause theme extraction in learning systems. Through its unique layered structure and bidirectional feedback topology, supported by interpretive formulas, SNRCEN theoretically addresses critical gaps in healthcare analytics, promoting resilient, evidence-driven intelligence.

This manuscript has outlined the theoretical foundations, architectural orchestration, systemic dynamics, and broader implications, underscoring the potential for narrative-centric approaches to enhance safety outcomes. While challenges in governance and modality integration persist, the framework's conceptual robustness paves the way for future evolutions in adaptive healthcare systems.

Ultimately, by structuring narratives into actionable themes, SNRCEN advocates for a paradigm where learning systems proactively mitigate risks, fostering safer clinical environments through intelligent evidence integration.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 23 Jan 2026 Revised: 06 Mar 2026 Accepted: 16 Apr 2026
Published online: 20 July 2026

References

Denecke K. Evaluating large language models for analysing safety risks in healthcare incident reports. Stud Health Technol Inform. 2025;329:386-90. https://doi.org/10.3233/SHTI250867.

Young IJB, Luz S, Lone N. A systematic review of natural language processing for classification tasks in the field of incident reporting and adverse event analysis. Int J Med Inform. 2019;132:103971. https://doi.org/10.1016/j.ijmedinf.2019.103971.

Wong ZSY. Medication-rights detection using incident reports: a natural language processing and deep neural network approach. Health Inform J. 2020;26(3):1777-94. https://doi.org/10.1177/1460458219889798.

Johnson J. Accuracy of a proprietary large language model in labeling obstetric incident reports. Jt Comm J Qual Patient Saf. 2024;50(12):877-81. https://doi.org/10.1016/j.jcjq.2024.08.001.

Wang Y, Coiera E, Runciman W, Magrabi F. Can unified medical language system-based semantic representation improve automated identification of patient safety incident reports by type and severity? J Am Med Inform Assoc. 2020;27(10):1502-9.

Evans HP, Anastassiou A, Edwards A, Hibbert P, Makeham M, Luzio S, et al. Automated classification of primary care patient safety incident report content and severity using supervised machine learning approaches. Health Inform J. 2020;26(4):3123-39. https://doi.org/10.1177/1460458219833102.

Takamatsu Y. Development of an automated classification system for medication-related incident factors: a practical approach to enhancing patient safety management. Stud Health Technol Inform. 2025;329:758-63. https://doi.org/10.3233/SHTI250942.

Wang Y. Assessing the transferability of BERT to patient safety: classifying multiple types of incident reports. BMJ Health Care Inform. 2025;32(1):e101146. https://doi.org/10.1136/bmjhci-2024-101146.

Uematsu H. Development of a scoring system to quantify errors from semantic characteristics in incident reports. BMJ Health Care Inform. 2024;31(1):e100935. https://doi.org/10.1136/bmjhci-2023-100935.

Wong ZSY. Rule-based natural language processing pipeline to detect medication-related named entities: insights for transfer learning. Stud Health Technol Inform. 2024;310:584-8. https://doi.org/10.3233/SHTI231032.

Ogi M. Exploring prompt-based large language model approach for medication error-related named entity recognition in medical incident reports. Stud Health Technol Inform. 2025;329:738-42. https://doi.org/10.3233/SHTI250938.

Wong ZSY. A large dataset of annotated incident reports on medication errors. Sci Data. 2024;11(1):260. https://doi.org/10.1038/s41597-024-03036-2.

Liu J, Wong EL, Ip M, Cheung AW, Wong MC. Exploring hidden in-hospital fall clusters from incident reports using text analytics. Stud Health Technol Inform. 2019;264:1563-4. https://doi.org/10.3233/SHTI190530.

Denecke K. Concept-based retrieval from critical incident reports. Stud Health Technol Inform. 2017;236:1-7.

Lear R, Godfrey C, O'Dowd H, Lear M, O'Dowd C. Co-producing a safe mobility and falls informatics platform to drive meaningful quality improvement in the hospital setting: a mixed-methods protocol for the insightFall study. BMJ Open. 2025;15(2):e082053. https://doi.org/10.1136/bmjopen-2023-082053.

Chen H, Fong A, Ratwani RM. A machine learning approach with human-AI collaboration for automated classification of patient safety event reports: algorithm development and validation study. JMIR Hum Factors. 2024;11:e53378. https://doi.org/10.2196/53378.

Fong A, Ratwani RM. A machine learning approach to reclassifying miscellaneous patient safety event reports. J Patient Saf. 2021;17(8):e829-e833. https://doi.org/10.1097/PTS.0000000000000731.

Boxley C, Fujimoto M, Ratwani RM. A text mining approach to categorize patient safety event reports by medication error type. Sci Rep. 2023;13(1):18388. https://doi.org/10.1038/s41598-023-45152-w.

Islam S, Chen H, Cohen E, Wilson D, Alfred M. Evaluating active learning strategies for automated classification of patient safety event reports in hospitals. Proc Hum Factors Ergon Soc Annu Meet. 2024;68(1):465-9. https://doi.org/10.1177/10711813241260676.

Fong A, Howe JL, Adams KT, Ratwani RM. Identifying health information technology related safety event reports from patient safety event report databases. J Biomed Inform. 2018;86:135-42. https://doi.org/10.1016/j.jbi.2018.09.007.

Fong A, Ratwani RM. Using active learning to identify health information technology related patient safety events. Appl Clin Inform. 2017;8(1):35-46. https://doi.org/10.4338/ACI-2016-09-CR-0148.

Adadey A, Chou W, Drury L. Developing an analytical pipeline to classify patient safety event reports using optimized predictive algorithms. Methods Inf Med. 2021;60(5-6):147-61. https://doi.org/10.1055/s-0041-1735620.

Tabaie A, Sengupta S, Pruitt ZM, Fong A. A natural language processing approach to categorise contributing factors from patient safety event reports. BMJ Open Qual. 2023;12(2):e002188. https://doi.org/10.1136/bmjoq-2022-002188.

Bangerter L, Van Haitsma K, Heid AR, Abbott K, Van der Horst K, Behrens L, et al. Artificial intelligence approach to optimise safety for hospitalised patients with dementia. BMJ Open Qual. 2025;14(3):e003270. https://doi.org/10.1136/bmjoq-2024-003270.

Liang C, Zhou S, Yao B, Hood D, Gong Y. Toward systems-centered analysis of patient safety events: improving root cause analysis by optimized incident classification and information presentation. Int J Med Inform. 2020;135:104053. https://doi.org/10.1016/j.ijmedinf.2019.104053.

Chen K, Rogers S, Yurkofsky M, Young J, Chao S, Bates DW, et al. AI-driven analysis of patient safety reports using large language models: an exploratory multiple methods study. BMJ Qual Saf. 2025. https://doi.org/10.1136/bmjqs-2025-019495.

Singh MK, Cooper C, Breckenridge J, Nugus P, Qiao M, Knight M, et al. I-SIRch: AI-powered concept annotation tool for equitable extraction and analysis of safety insights from maternity investigations. BMJ Health Care Inform. 2024;31(1):e100928. https://doi.org/10.1136/bmjhci-2023-100928.

Hölzing CR, Pfisterer J, Kuenecke J, Wohlhüter F, Kunst C, Streicher F, et al. The potential of using generative AI/NLP to identify and analyse critical incidents in a critical incident reporting system (CIRS): a feasibility case-control study. Healthcare. 2024;12(19):1964. https://doi.org/10.3390/healthcare12191964.

Choudhury A, Asan O. Role of artificial intelligence in patient safety outcomes: systematic literature review. JMIR Med Inform. 2020;8(7):e18599. https://doi.org/10.2196/18599.