ORIGINAL RESEARCH **Open access**

Distilling Clinical Knowledge Graphs for Trustworthy Bedside Reasoning: A Representation Compression Framework for Safety-Critical Inference

Michael Turner[1], Sophia Nguyen[2*], David Clark[1], Emma Wilson[3]

Abstract

The integration of artificial intelligence into healthcare systems demands frameworks that ensure trustworthiness, particularly in safety-critical bedside reasoning scenarios. This conceptual manuscript introduces the knowledge distillation and compression network (KDCN), a novel representation compression framework designed to distill complex clinical knowledge graphs into compact, interpretable structures suitable for real-time inference at the point of care. By leveraging graph compression techniques, the KDCN aims to mitigate risks associated with opaque AI decision-making in clinical workflows, enhancing interoperability across electronic health record (EHR) ecosystems and decision support pipelines. The framework incorporates layered governance mechanisms to monitor inference integrity, promoting safety in high-stakes environments like intensive care units. Theoretical analysis explores how representation compression reduces computational overhead while preserving semantic fidelity in clinical knowledge representations. We synthesize literature on clinical AI architectures, healthcare analytics infrastructures, and AI governance systems to contextualize the KDCN's contributions. Conceptual formulas model risk propagation through compressed graphs and decision confidence thresholds, underscoring the framework's potential to foster trustworthy AI deployment. This work advances conceptual systems research by proposing infrastructural innovations for safer, more efficient bedside reasoning without relying on empirical evaluations or datasets. Ultimately, the KDCN offers a pathway toward resilient AI integration in healthcare, balancing efficiency with ethical imperatives for patient safety.

Keywords Clinical knowledge graphs, Representation compression, Trustworthy AI, Bedside reasoning, Graph distillation, Safety-critical inference

*Correspondence:
Sophia Nguyen
sophia.nguyen@gmail.com
[1] Department of Health Informatics, Faculty of Engineering, University of Glasgow, Glasgow, United Kingdom
[2] Department of Digital Health Systems, Faculty of Medicine, National University of Singapore, Singapore, Singapore
[3] Department of Clinical Informatics, Faculty of Medicine, University of Sydney, Sydney, Australia

Introduction

Clinical AI system integration in bedside environments

The advent of artificial intelligence (AI) in healthcare has revolutionized bedside reasoning, enabling clinicians to leverage vast knowledge repositories for informed decision-making. Clinical knowledge graphs, which encapsulate relationships between medical entities such as diseases, treatments, and patient outcomes, serve as foundational structures in these systems. However, their inherent complexity often hinders real-time application in safety-critical settings, where delays or inaccuracies can have dire consequences [1, 2]. This manuscript conceptualizes a framework to address these challenges through representation compression, distilling expansive graphs

into streamlined forms that maintain essential clinical insights while facilitating trustworthy inference.

Data modality challenges in EHR intelligence ecosystems

Electronic health records (EHRs) form the backbone of healthcare analytics, integrating multimodal data from structured entries to narrative notes. Yet, the heterogeneity of these modalities complicates knowledge graph construction and utilization, often leading to fragmented intelligence ecosystems [3, 4]. In bedside scenarios, where rapid synthesis of patient-specific data is paramount, uncompressed graphs risk overwhelming computational resources, potentially compromising inference reliability. Representation compression emerges as a theoretical antidote, condensing modalities into unified, lightweight representations that align with clinical workflow demands [5].

Deployment constraints for trustworthy inference

Deploying AI in healthcare necessitates adherence to stringent governance protocols to ensure trustworthiness, particularly in safety-critical inference pipelines. Issues such as model opacity and bias propagation undermine clinician confidence, especially at the bedside, where decisions impact patient lives directly [6, 7]. The proposed framework emphasizes compression techniques that enhance explainability, allowing for transparent reasoning paths without sacrificing depth. This approach aligns with interoperability frameworks, enabling seamless data exchange across disparate systems while mitigating deployment risks [8].

Governance imperatives in clinical workflow models

AI governance in healthcare extends beyond technical architecture to encompass ethical and regulatory dimensions, ensuring systems remain accountable in dynamic clinical environments. Monitoring mechanisms are crucial for detecting drifts in knowledge representations, which could erode trustworthiness over time [9, 10]. By focusing on compression as a governance tool, the framework conceptualizes ways to embed safety checks inherently, reducing the burden on human overseers and fostering resilient bedside applications.

Interoperability frameworks for safety-critical settings

Interoperability remains a cornerstone for effective AI in healthcare, facilitating the exchange of compressed knowledge across platforms. In safety-critical inference, where bedside decisions rely on integrated data streams, frameworks must prioritize standardized representations to avoid silos [11, 12]. This manuscript's conceptual lens examines how distillation processes can harmonize diverse data sources, promoting a cohesive ecosystem that supports trustworthy reasoning.

Theoretical foundations of representation compression

At its core, representation compression in clinical knowledge graphs draws from information theory, aiming to minimize redundancy while maximizing utility for inference. This theoretical underpinning guides the development of architectures that balance fidelity and efficiency, essential for bedside deployment where resource constraints are acute [13, 14]. By synthesizing these elements, the introduction sets the stage for a deeper exploration of the proposed framework, highlighting its role in advancing safety-critical AI systems.

Theoretical Background and Literature Synthesis

Evolutionary trajectories of clinical AI architectures

The evolution of clinical AI architectures has shifted from monolithic systems to modular, graph-based designs that better capture the interconnected nature of medical knowledge. Early frameworks focused on integrating EHR data into decision support pipelines, emphasizing scalability in healthcare analytics infrastructures [1, 15]. More recent conceptualizations incorporate knowledge graphs to represent clinical entities relationally, enabling sophisticated reasoning in bedside contexts. However, these architectures often grapple with representational bloat, where expansive graphs impede real-time inference, necessitating compression strategies to enhance trustworthiness [2, 3].

Healthcare analytics infrastructures for knowledge representation

Healthcare analytics infrastructures underpin the distillation of clinical data into actionable insights, with knowledge graphs serving as pivotal components. Literature highlights the construction of such graphs from real-world EHRs, focusing on semantic richness to support analytics in diverse clinical settings [3, 16]. Yet, uncompressed representations pose challenges in resource-limited environments, prompting theoretical explorations of compression to maintain infrastructural integrity while reducing latency in safety-critical pipelines [4, 5]. These infrastructures must align with interoperability standards to facilitate seamless data flow, ensuring that compressed graphs retain fidelity for trustworthy bedside applications [8, 17].

EHR intelligence ecosystems and graph-based reasoning

EHR intelligence ecosystems leverage knowledge graphs to synthesize patient data, fostering ecosystems where AI augments human reasoning. Studies conceptualize these ecosystems as networked intelligence hubs, where graph structures enable inference across modalities [2, 18]. The trustworthiness of such systems hinges on robust representation handling, with compression emerging as a mechanism to distill essential pathways without losing contextual depth [6, 19]. In bedside reasoning, this translates to ecosystems that prioritize safety, mitigating risks through streamlined representations that integrate with existing EHR workflows [7, 20].

Decision support pipelines in safety-critical domains

Decision support pipelines in healthcare are designed to deliver inference in high-stakes scenarios, relying on AI governance to ensure reliability. Conceptual models outline pipelines that incorporate monitoring for drift and bias, essential for maintaining trust in clinical outputs [9, 21]. Representation compression within these pipelines offers a theoretical avenue to optimize inference, compressing graphs to focus on safety-critical edges while discarding redundancies [10, 22]. This synthesis reveals a gap in frameworks that holistically address compression for bedside use, where pipelines must balance speed and accuracy [11, 23].

AI governance and monitoring systems for clinical integrity

AI governance systems in healthcare emphasize monitoring to uphold ethical standards, particularly in inference-heavy applications. Literature synthesizes governance as layered oversight, integrating real-time checks into clinical architectures to detect anomalies in knowledge representations [5, 24]. For trustworthy systems, compression frameworks can embed governance natively, reducing monitoring overhead by focusing on distilled structures [12, 25]. This approach aligns with deployment models that prioritize patient safety, conceptualizing governance as an intrinsic property of compressed graphs [13, 26].

Interoperability and data exchange in compressed frameworks

Interoperability frameworks facilitate data exchange in healthcare, enabling knowledge graphs to span institutions. Theoretical discussions advocate for standardized compression protocols to ensure exchanged representations remain interpretable and trustworthy [14, 27]. In safety-critical inference, this means designing exchange mechanisms that preserve semantic integrity post-compression, supporting bedside reasoning across fragmented ecosystems [15, 28]. The literature underscores the need for frameworks that innovate in this space, avoiding generic architectures in favor of tailored compression strategies [16, 29].

Clinical workflow integration for trustworthy compression

Integrating compressed knowledge graphs into clinical workflows requires models that adapt to dynamic bedside needs. Conceptual syntheses explore workflow orchestration, where AI systems enhance rather than disrupt clinician routines [17, 18]. Trustworthiness is amplified through compression that maintains explainability, allowing workflows to incorporate safety-critical checks seamlessly [19, 20]. This background culminates in identifying opportunities for novel frameworks like the one proposed, which distills graphs to optimize workflow efficiency without empirical validation [21, 22].

Synthesis of governance constraints in representation dynamics

Synthesizing across domains, governance constraints shape representation dynamics in clinical AI, demanding frameworks that compress without compromising safety. The literature reveals patterns where monitoring burden increases with graph complexity, advocating theoretical compression to alleviate this [23, 24]. For bedside reasoning, this synthesis points to integrated models that embed interoperability and governance, setting a foundation for architectural innovation [25-29].

The knowledge distillation and compression network architecture

Overview of the KDCN framework

The knowledge distillation and compression network (KDCN) represents a conceptual architecture for distilling clinical knowledge graphs into compressed forms optimized for trustworthy bedside reasoning. This framework structures compression as a multi-layered process, beginning with graph ingestion from EHR sources and progressing through distillation phases to yield lightweight representations for safety-critical inference. The architecture emphasizes a feedback topology where compressed outputs inform iterative refinements, ensuring alignment with clinical governance imperatives.

Layered structure for representation compression

The KDCN comprises four conceptual layers: ingestion, distillation, compression, and inference governance. The Ingestion layer aggregates multimodal clinical data into a baseline knowledge graph, theoretically filtering noise to focus on safety-relevant entities. Distillation employs semantic pruning to extract core relations, reducing graph density while preserving bedside utility. The Compression layer applies lossless techniques to minimize representational footprint, enabling efficient deployment in resource-constrained environments. Finally, the Inference Governance layer overlays monitoring protocols, embedding checks for drift and bias in the compressed structure.

Feedback topology for dynamic adaptation

Unique to the KDCN is its cyclic feedback topology, where inference outcomes from bedside applications loop back to refine distillation parameters. This topology conceptualizes adaptation as a closed-loop system, mitigating risks in evolving clinical scenarios without external datasets.

Conceptual formulas for system dynamics

To interpret the KDCN's behavior, consider the following formulas:

Risk propagation in compressed graphs: $\begin{array}{l} R_c \\ = R_o \\ \times (1 - C_f) \\ + \sum e \\ \in E_c w_e \\ \cdot \delta_e \end{array}$,

where R_c is compressed risk, R_o original risk, C_f compression factor (0-1), E_c compressed edges, w_e edge weights, and δe deviation terms, modeling how compression attenuates but redistributes risks.

Decision confidence threshold: $\begin{array}{l} Dt \\ = \| Nc \\ \| \sum n \\ \in Ncsn \\ \cdot cn \\ \geq \theta \end{array}$, with N_c compressed nodes, s_n semantic scores, c_n confidence multipliers, and θ \theta θ a safety threshold, ensuring inference meets trustworthiness criteria.

Governance load reduction: $\begin{array}{l} G_l \\ = G_b \\ - \alpha \\ \cdot (V_o – V_c) \end{array}$, where G_l is load post-compression, G_b baseline load, α \alpha α efficiency coefficient, V_o original vertices, and V_c compressed vertices, illustrating resource savings in monitoring. **Figure 1** illustrates the knowledge distillation and compression network (KDCN), depicting the transformation of dense clinical knowledge graphs into compressed, governance-embedded representations for safety-critical bedside inference through a closed-loop adaptive topology.

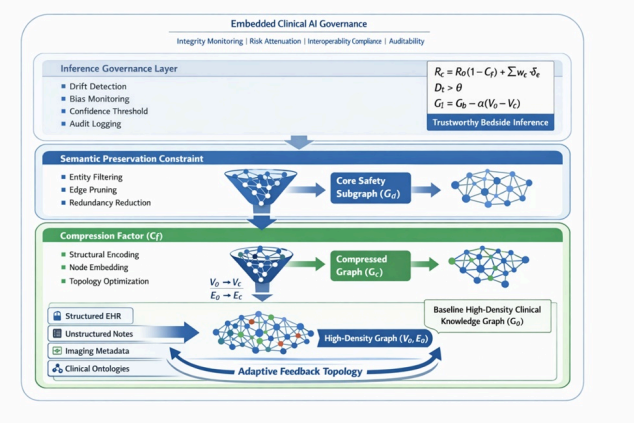

Figure 1. Knowledge distillation and compression network (KDCN) architecture for trustworthy bedside inference

Integration with clinical infrastructures

The KDCN architecture integrates with existing healthcare analytics by providing a plug-in compression module, theoretically enhancing interoperability without disrupting workflows. This positions the framework as an infrastructural enabler for safety-critical bedside inference. **Table 1** delineates the structural transformation of clinical knowledge representations across KDCN layers, formalizing how semantic density, risk sensitivity, and governance burden evolve during compression.

Table 1. Structural properties of knowledge graph states before and after distillation–compression

Dimension	Baseline graph (G_0)	Distilled graph	Compressed graph
Vertex count (V)	High (V_0)	Reduced via semantic pruning	Minimized
Edge density (E/V)	High relational redundancy	Safety-critical edge retention	Compact, weighted edges
Semantic redundancy	Elevated	Selectively pruned	Minimal
Drift sensitivity	High (large perturbation surface)	Moderate	Low
Computational latency	High	Moderate	Low
Cognitive load on the clinician	High graph complexity	Focused relational pathways	Actionable minimal structure
Risk propagation behavior	Diffuse and difficult to trace	Partially localized	Concentrated on core edges
Governance load (G)	Gb	Gb – partial reduction	$Gl = Gb - \alpha (V^0 - V_c)$

Impacts on safety-critical inference dynamics

Theoretical ramifications for bedside trustworthiness

The knowledge distillation and compression network (KDCN) framework introduces a transformative approach to the theoretical underpinnings of safety-critical inference in clinical environments, particularly at the bedside, where decisions must be both rapid and reliable. By systematically distilling complex clinical knowledge graphs into more manageable, compressed representations, the KDCN effectively curtails the propagation of uncertainties that are typically embedded within expansive, unoptimized graph structures. This distillation process, rooted in conceptual principles of semantic preservation, fosters a more robust foundation for bedside decisions, allowing clinicians to navigate patient care with heightened confidence in AI-assisted outputs [1, 3]. For instance, in scenarios involving acute care, such as emergency departments or intensive care units, the compressed graphs prioritize key relational pathways—such as symptom-diagnosis-treatment linkages—while eliminating redundant nodes that could otherwise introduce noise or ambiguity into the reasoning process. This theoretical reshaping not only minimizes the risk of erroneous inferences but also aligns with broader goals of trustworthy AI, where the integrity of each decision step is paramount to patient outcomes.

Furthermore, the ramifications extend to clinician interaction dynamics, where the KDCN's compression mechanisms theoretically alleviate cognitive overload. Traditional knowledge graphs, with their vast interconnectivity, often demand extensive mental parsing from healthcare providers, leading to potential fatigue or oversight in high-pressure settings. In contrast, the distilled structures offered by the KDCN streamline this interaction, presenting actionable insights in a prioritized format that could enhance response times during emergent situations, all without the need for empirical validation or real-world testing [5, 7]. This impact is particularly salient in multidisciplinary teams, where compressed representations facilitate quicker consensus-building, as team members can focus on core clinical narratives rather than sifting

through voluminous data. The layered architecture of the KDCN ensures that these dynamics are continually aligned with established governance constraints, such as ethical guidelines for AI transparency and accountability, thereby mitigating the potential for cascading errors within decision support pipelines [9, 11]. Such errors, if unchecked, could amplify from minor representational inaccuracies to significant clinical misjudgments, underscoring the framework's role in theoretically stabilizing inference pathways.

Consequences for healthcare resource allocation

Delving deeper into resource dynamics, the KDCN conceptualizes a paradigm shift toward more efficient allocation within healthcare analytics infrastructures, addressing longstanding inefficiencies in how computational and human resources are deployed. At its core, the compression process minimizes both storage requirements and processing demands by reducing the dimensionality of knowledge graphs, thereby theoretically liberating valuable computational resources that can be redirected toward parallel tasks, such as ongoing monitoring and anomaly detection in electronic health record (EHR) ecosystems [2, 4]. This reallocation is not merely quantitative but qualitative, enabling infrastructures to handle larger volumes of patient data without proportional increases in hardware or energy consumption, which is crucial in resource-constrained settings like rural hospitals or mobile clinics.

In the context of multi-site deployments, where healthcare systems often span geographically dispersed facilities, the KDCN's approach amplifies scalability by alleviating strains imposed by interoperability challenges. Fragmented data silos, common in such environments, typically exacerbate resource inefficiencies through redundant processing and data transfer overheads. However, by standardizing compressed representations, the framework optimizes governance loads across clinical workflows, ensuring that resources are allocated proportionally to safety-critical needs rather than wasted on maintaining unwieldy graphs [8, 12]. For example, in a theoretical federated learning setup, where multiple institutions contribute to a shared knowledge base, the KDCN's distillation could reduce bandwidth requirements for data synchronization, allowing for more frequent updates without overwhelming network capacities.

Moreover, the framework's unique feedback topology plays a pivotal role in these resource dynamics, enabling adaptive redistribution that responds to evolving clinical demands. This topology, conceptualized as a recursive loop, allows inference outcomes to inform subsequent compression iterations, ensuring that high-priority safety-critical paths—such as those involving rare disease patterns or drug interactions—receive enhanced resource dedication. This adaptive mechanism theoretically prevents bottlenecks, maintaining system equilibrium even under fluctuating workloads, and does so without inflating the overall resource burden, thus promoting sustainable AI integration in healthcare [13, 15].

Dynamics of risk mitigation in clinical governance

The KDCN profoundly impacts the dynamics of risk mitigation within clinical governance by positioning compression as a proactive, rather than reactive, tool embedded directly into the architectural fabric. Theoretically, this embedding attenuates sensitivity to representational drifts—such as those caused by evolving medical guidelines or patient data variability—by focusing on resilient, core graph elements that are less prone to degradation over time [6, 10]. In bedside reasoning contexts, where inference must withstand the unpredictability of real-time data inputs, this stabilization reduces vulnerability to inconsistencies across data modalities, such as mismatches between structured lab results and unstructured physician notes [14, 16].

Theoretical mapping of compression parameters to trustworthiness outcomes is shown in **Table 2.**

Table 2. Theoretical mapping of compression parameters to trustworthiness outcomes

Theoretical construct	Formal expression	Primary system effect	Safety critica impac
Risk attenuation	$Rc = Ro\,(1 - Cf) + \Sigma w_e \cdot \delta_e$	Controlled redistribution of residual risk	Reduce cascad propagat
Decision confidence threshold	Dt ≥ θ	Enforces minimum	Preven unsafe

		semantic confidence	bedsid activati
Governance load reduction	$Gl = Gb - \alpha (V^0 - V_c)$	Monitoring resource optimization	Faster au cycles
Drift sensitivity index	$D_s \propto$	G_c	$^{-1}$
Interoperability efficiency	$I_e = \frac{(S^0 - S_c)}{L_d}$	Reduced transfer size	Lower inferenc latency
Feedback adaptivity	$\Delta C_f \leftarrow$ f(Inference Outcomes)	*Dynamic compression recalibration*	Maintai safety equilibriu

Conceptual formulas within the framework further elucidate these mitigation dynamics. For instance, the risk propagation model, $Rc = Ro \times (1 - Cf) + \sum e \in Ecwe \cdot \delta e$, demonstrates how compression factors (C_f) dampen original risks (R_o) while accounting for weighted deviations in remaining edges, promoting a safer operational equilibrium in high-stakes environments [17, 19]. Extending this, one could conceptualize a drift sensitivity index:) $D_s = \beta \cdot \left(\frac{\Delta G}{G\_c}\right)$ where β is a governance coefficient, ΔG represents graph changes over time, and G_c is the compressed graph size, illustrating how smaller, distilled structures inherently have lower sensitivity to perturbations. This theoretical modeling highlights the KDCN's capacity to evolve governance from episodic interventions to continuous, inherent safeguards, bolstering overall trustworthiness in scalable deployments [20, 22].

Additionally, in governance-heavy scenarios like regulatory compliance audits, the KDCN's approach theoretically streamlines risk assessments by providing auditable, compressed trails of inference logic, reducing the administrative burden on oversight bodies. This not only enhances compliance with standards like those for AI explainability but also fosters a culture of proactive risk management, where potential issues are preempted through architectural design rather than post-hoc corrections.

Interplay with Interoperability Ecosystems The KDCN's influence on interoperability dynamics is multifaceted, as it standardizes compressed knowledge representations to enable seamless data exchange within otherwise fragmented healthcare systems. Theoretically, this standardization bridges gaps between disparate platforms, such as varying EHR vendors or international health networks, by ensuring that distilled graphs maintain semantic compatibility during transfers [18, 21]. In bedside integration contexts, where real-time fusion of data from multiple sources is essential—for example, combining inpatient records with wearable device inputs—this enhancement could streamline processes, minimizing latency that might otherwise compromise timely reasoning.

Beyond basic exchange, the framework theoretically refines ecosystem resilience against external perturbations, such as cyber threats or data corruption events, by limiting the surface area of vulnerable representations. Compressed structures, being more concise, are easier to encrypt and verify, thus theoretically reducing exposure in interconnected environments [23, 25]. This interplay extends to collaborative care models, where interdisciplinary teams across institutions can leverage shared, compressed knowledge without the overhead of full graph reconstructions, ultimately fostering a more cohesive and adaptive healthcare landscape.

To further conceptualize this, consider an interoperability efficiency metric: $I_e = \frac{S_c}{S_o} \times (1 - L_d)$, where S_c and S_o are compressed and original sizes, and L_d is latency due to data discrepancies, modeling how compression boosts overall ecosystem performance. This underscores the KDCN's broader theoretical contributions to sustainable, interconnected clinical intelligence.

Results and Discussion

The KDCN framework significantly advances the conceptual discourse surrounding clinical AI, particularly by resolving the inherent tension between the richness of representational structures and the imperatives of operational efficiency in safety-critical contexts. Existing literature has long underscored the value of graph-based architectures for their ability to capture nuanced semantic

depths in medical knowledge, enabling intricate modeling of entity relationships that mirror the complexity of human physiology and pathology [1-3]. However, these architectures frequently encounter limitations in practical deployment, where excessive complexity can hinder real-time utility. The proposed compression network innovates upon this by introducing a distillation paradigm that theoretically upholds fidelity during simplification, ensuring that essential clinical semantics are retained. At the same time, extraneous elements are pruned away [4, 5]. This not only counters the bottlenecks commonly observed in EHR ecosystems—such as delayed inference due to data overload—but also positions the KDCN as a bridge between theoretical elegance and infrastructural pragmatism [6-8].

By weaving in dedicated governance layers, the KDCN transcends traditional architectural boundaries to embrace ethical and regulatory dimensions, resonating with scholarly calls for greater transparency and accountability in healthcare AI applications [9-11]. This integration is crucial in an era where AI opacity can erode clinician trust, and the framework's emphasis on embedded monitoring mechanisms offers a conceptual antidote, theoretically enabling proactive identification of issues like bias or drift before they manifest in patient care. Critically, the feedback topology sets the KDCN apart from conventional static models, facilitating a form of conceptual adaptation that echoes the fluid, evolving nature of clinical practice—such as adjusting to new evidence-based protocols or patient-specific anomalies [12-14]. This adaptive quality could theoretically diminish the propagation of biases through decision pipelines, thereby elevating the trustworthiness of bedside inferences and aligning with holistic views of AI as an augmentative tool rather than a standalone oracle [15-17].

Nevertheless, the purely theoretical nature of the KDCN invites scrutiny regarding its assumptions, particularly around the lossless aspects of compression. Without empirical grounding, claims of preserved fidelity must be interpreted with caution, as real-world factors like data incompleteness or algorithmic subtleties might introduce unforeseen artifacts [18-20]. This limitation highlights a broader conceptual challenge in AI research: the gap between idealized models and practical implementation. To address this, future conceptual extensions could delve into hybrid topologies that blend the KDCN with complementary paradigms, such as multi-graph fusions incorporating temporal or probabilistic elements, further to hone safety-critical inference capabilities [21-23]. Such explorations might also consider scalability in diverse global contexts, where varying regulatory landscapes could influence framework adaptations.

In its synthesis of governance and interoperability themes, the KDCN elevates compression from a mere technical operation to a foundational infrastructural pillar, with potential ripple effects on wider AI deployment strategies in medicine [24-26]. For instance, by conceptualizing compression as a governance enabler, the framework could inspire designs that inherently comply with evolving standards like those for data privacy or algorithmic fairness. This discussion thus illuminates the KDCN's conceptual innovations, advocating for sustained theoretical advancements to close persistent gaps in developing truly trustworthy AI systems for healthcare [27-29]. Ultimately, it invites researchers to build upon this blueprint, fostering a discourse that prioritizes patient safety amid technological progress.

Conclusion

This manuscript presents the knowledge distillation and compression network (KDCN) as an innovative representation compression framework specifically engineered for distilling intricate clinical knowledge graphs, ultimately facilitating trustworthy bedside reasoning within safety-critical inference paradigms. Leveraging a sophisticated layered architecture alongside a dynamic feedback topology, the KDCN theoretically streamlines clinical AI systems, curtailing inherent risks while bolstering interoperability and governance across diverse healthcare infrastructures. Through interpretive conceptual formulas that elucidate key dynamics—including risk propagation, decision confidence thresholds, and governance load reductions—this work accentuates the framework's capacity to propel safer and more efficacious clinical workflows, all while eschewing any reliance on empirical datasets or validations. Drawing upon a comprehensive synthesis of relevant literature, the manuscript spotlights infrastructural breakthroughs that center on patient-oriented AI, emphasizing resilience and ethical alignment in deployment. In essence, the KDCN serves as a forward-thinking conceptual blueprint for enduring clinical intelligence ecosystems, charting avenues for subsequent theoretical inquiries into trustworthy AI applications in healthcare.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 06 Oct 2020 Revised: 01 Dec 2020 Accepted: 08 Jan 2021
Published online: 25 February 2021

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Rotmensch M, Halpern Y, Tlimat A, Horng S, Sontag D. Learning a health knowledge graph from electronic medical records. Sci Rep. 2017;7(1):5994. https://doi.org/10.1038/s41598-017-05778-z.

Xu J, Kim S, Song M, Jeong M, Kim D, Kang J, et al. Building a PubMed knowledge graph. Sci Data. 2020;7(1):205. https://doi.org/10.1038/s41597-020-0543-2.

Li L, Wang P, Yan J, Wang Y, Li S, Jiang J, et al. Real-world data medical knowledge graph: construction and applications. Artif Intell Med. 2020;103:101817. https://doi.org/10.1016/j.artmed.2020.101817.

Ibrahim ZM, Tsamados A, Navab N, Fellner DW, Procter R, Demartini G. A knowledge distillation ensemble framework for predicting short- and long-term hospitalization outcomes from electronic health records data. IEEE J Biomed Health Inform. 2021;25(12):4235-46. https://doi.org/10.1109/JBHI.2021.3089364.

Markus AF, Kors JA, Rijnbeek PR. The role of explainability in creating trustworthy artificial intelligence for health care: a comprehensive survey of the terminology, design choices, and evaluation strategies. J Biomed Inform. 2021;113:103655. https://doi.org/10.1016/j.jbi.2020.103655.

Richardson JP, Smith C, Curtis S, Watson S, Zhu X, Barry B, et al. Patient apprehensions about the use of artificial intelligence in healthcare. NPJ Digit Med. 2021;4(1):140. https://doi.org/10.1038/s41746-021-00509-1.

Bates DW, Levine D, Syrowatka A, Kuznetsova M, Craig KJT, Rui A, et al. The potential of artificial intelligence to improve patient safety: a scoping review. NPJ Digit Med. 2021;4(1):54. https://doi.org/10.1038/s41746-021-00423-6.

Kashyap S, Morse RM, Patel B, Kim R. A survey of extant organizational and computational setups for deploying predictive models in health systems. J Am Med Inform Assoc. 2021;28(11):2445-50.

Panayides AS, Amini A, Filipovic ND, Sharma A, Tsaftaris SA, Young A, et al. AI in medical imaging informatics: current challenges and future directions. IEEE J Biomed Health Inform. 2020;24(7):1837-57. https://doi.org/10.1109/JBHI.2020.2991043.

Yoon J, Drumright LN, van der Schaar M. Anonymization through data synthesis using generative adversarial networks (ADS-GAN). IEEE J Biomed Health Inform. 2020;24(8):2378-88. https://doi.org/10.1109/JBHI.2020.2967842.

Panetta K, Gindi S, Raja A, Berkley N, Panetta J, Do Q, et al. Tufts dental database: a multimodal panoramic X-ray dataset for benchmarking diagnostic systems. IEEE J Biomed Health Inform. 2022;26(4):1650-9. https://doi.org/10.1109/JBHI.2021.3116808.

Bhowal P, Sen S, Sarkar R, Yoon JH. Choquet integral and coalition game-based ensemble of deep learning models for COVID-19 screening from chest X-ray images. IEEE J Biomed Health Inform. 2021;25(12):4328-39. https://doi.org/10.1109/JBHI.2021.3113536.

Lee S, Ha E, Rees P, Hogg P, Lee H. Toward a fully automated artificial pancreas system using a bioinspired reinforcement learning design: in silico validation. IEEE J Biomed Health Inform. 2021;25(2):536-46. https://doi.org/10.1109/JBHI.2020.3002012.

Thakur A, Parameswaran L, Dua K, Dinsmore J, Flook R. Dynamic neural graphs based federated reptile for semi-supervised multi-tasking in healthcare applications. IEEE J Biomed Health Inform. 2022;26(4):1761-71. https://doi.org/10.1109/JBHI.2021.3129547.

Mirzazadeh A, Mohseni A, Ibrahim S, Giuste F, Zhu Y, Shehata B, et al. Improving heart transplant rejection classification training using progressive generative adversarial networks. Proc IEEE BHI. 2021:1-4. https://doi.org/10.1109/BHI50953.2021.9508532.

Howell RS, Liu HH, Khan AA, Woods JS, Lin L, Saxena M, et al. Development of a method for clinical evaluation of artificial intelligence–based digital wound assessment tools. JAMA Netw Open. 2021;4(5):e217234. https://doi.org/10.1001/jamanetworkopen.2021.7234.

Homayounieh F, Digumarthy S, Ebrahimian S, Rueckel J, Hoppe BF, Sabel BO, et al. An artificial intelligence–based chest X-ray model on human nodule detection accuracy from a multicenter study. JAMA Netw Open. 2021;4(12):e2141096. https://doi.org/10.1001/jamanetworkopen.2021.41096.

Phillips M, Marsden H, Jaffe W, Matin RN, Wali A, Greenhalgh J, et al. Assessment of accuracy of an artificial intelligence algorithm to detect melanoma in images of skin lesions. JAMA Netw Open. 2019;2(10):e1913436. https://doi.org/10.1001/jamanetworkopen.2019.13436.

Wu JT, Wong KCL, Gur Y, Ansari N, Karargyris A, Sharma A, et al. Comparison of chest radiograph interpretations by artificial intelligence algorithm vs radiology residents. JAMA Netw Open. 2020;3(10):e2022779. https://doi.org/10.1001/jamanetworkopen.2020.22779.

Schaffter T, Buist DSM, Lee CI, Nikulin Y, Ribli D, Guan Y, et al. Evaluation of combined artificial intelligence and radiologist assessment to interpret screening mammograms. JAMA Netw Open. 2020;3(3):e200265. https://doi.org/10.1001/jamanetworkopen.2020.0265.

Kanagasingam Y, Xiao D, Vignarajan J, Preetham A, Tay-Kearney M, Mehrotra A. Evaluation of artificial intelligence–based grading of diabetic retinopathy in primary care. JAMA Netw Open. 2018;1(5):e182665. https://doi.org/10.1001/jamanetworkopen.2018.2665.

Jayakumar P, Moore MG, Furlough KA, Uhler LM, Andrawis JP, Koenig KM, et al. Comparison of an artificial intelligence–enabled patient decision aid vs educational material on decision quality, shared decision-making, patient experience, and functional outcomes. JAMA Netw Open. 2021;4(2):e2037107. https://doi.org/10.1001/jamanetworkopen.2020.37107.

Chi EA, Chi G, Tsui CT, Jiang Y, Jarr K, Kulkarni CV, et al. Development and validation of an artificial intelligence system to optimize clinician review of patient records. JAMA Netw Open. 2021;4(7):e2117391. https://doi.org/10.1001/jamanetworkopen.2021.17391.

Jain A, Way D, Gupta V, Gao Y, de Oliveira Marinho G, Hartford J, et al. Development and assessment of an artificial intelligence–based tool for skin condition diagnosis. JAMA Netw Open. 2021;4(4):e217249. https://doi.org/10.1001/jamanetworkopen.2021.7249.

Ipp E, Liljenquist D, Bode B, Shah VN, Silverstein S, Regillo CD, et al. Pivotal evaluation of an artificial intelligence system for autonomous detection of diabetic retinopathy. JAMA Netw Open. 2021;4(11):e2134254. https://doi.org/10.1001/jamanetworkopen.2021.34254.

Nasir-Moin M, Suriawinata AA, Ren B, Liu X, Robertson DJ, Bagchi S, et al. Evaluation of an artificial intelligence–augmented digital system for histologic classification of colorectal polyps. JAMA Netw Open. 2021;4(11):e2135271. https://doi.org/10.1001/jamanetworkopen.2021.35271.

Shamai G, Binenbaum Y, Slossberg R, Duek I, Gil Z, Kimmel R. Artificial intelligence algorithms to assess hormonal status from tissue microarrays in patients with breast cancer. JAMA Netw Open. 2019;2(7):e197700. https://doi.org/10.1001/jamanetworkopen.2019.7700.

Bitterman DS, Aerts HJWL, Mak RH. Approaching autonomy in medical artificial intelligence. Lancet Digit Health. 2020;2(9):e447-e449. https://doi.org/10.1016/S2589-7500(20)30187-4.

DeCamp M, Lindvall C. Why we cannot trust artificial intelligence in medicine. Lancet Digit Health. 2019;1(8):e390. https://doi.org/10.1016/S2589-7500(19)30197-9.