Explainable AI-Assisted Clinical Decision Support Using Electronic Health Record Time-Series Data

Vinay Venkataraman
School of Information Technology, University of Cincinnati, Cincinnati, OH, USA.
vinayv@uc.edu

Brandon Lindgren
Department of Computer Science, University of Houston, Houston, TX, USA.
brandonmail@uh.edu

Claudio J. Jacobs
Department of Computer Science, University of Alabama at Birmingham, Birmingham, AL, USA.
claudiojacobs94@uab.edu

Abstract

The integration of artificial intelligence into clinical decision support systems holds transformative potential for modern healthcare, particularly when leveraging the rich, longitudinal data contained within electronic health records. However, the widespread adoption of such systems is critically dependent on their interpretability, as opaque model predictions undermine clinician trust and pose significant regulatory and ethical challenges. This paper presents a comprehensive analysis of explainable AI-assisted clinical decision support systems that operate on electronic health record time-series data. It moves beyond the technical specifics of individual algorithms to address the broader systemic, architectural, and governance challenges inherent in deploying these systems at scale. The discussion begins by characterizing the unique structural properties of electronic health record time-series, including irregular sampling, high dimensionality, and missing data mechanisms, and how these properties constrain model design. The paper then examines the architectural trade-offs between predictive fidelity and interpretability, contrasting deep learning frameworks with intrinsically interpretable models and post-hoc explanation methods. A substantial portion of the analysis is dedicated to the deployment infrastructure required to operationalize such systems within hospital information technology environments, focusing on data pipelines, real-time inference latency, and the sustainability of continuous model monitoring. The paper further explores the critical dimensions of fairness and robustness, demonstrating how biases embedded in historical clinical data can be amplified by opaque models, and how explainability can serve as a diagnostic tool for auditing these biases. Finally, the discussion extends to policy implications and the evolving regulatory landscape, particularly concerning model validation, accountability, and patient safety. The paper concludes that the future of AI in clinical decision support lies not in pursuing ever-greater complexity, but in architecting systems that are inherently transparent, resilient, and aligned with the socio-technical realities of clinical practice.

Keywords

explainable artificial intelligence, clinical decision support, electronic health records, time-series analysis, healthcare infrastructure, algorithmic fairness, model governance, predictive analytics.

1. Introduction

The digitization of healthcare systems has generated massive repositories of electronic health record data, offering an unprecedented opportunity to apply machine learning for improving patient outcomes. Clinical decision support systems powered by artificial intelligence promise to augment clinician judgment by predicting patient deterioration, recommending treatments, and identifying high-risk cohorts from historical data patterns. However, the transition from algorithmic development in research laboratories to reliable deployment in hospital settings is fraught with challenges that extend far beyond model accuracy [1]. The opaque nature of many high-performing models, particularly deep neural networks, creates a fundamental tension between predictive power and the need for interpretability in high-stakes medical environments. Clinicians require not just accurate predictions, but explanations that align with their causal reasoning and domain expertise to trust and act upon system recommendations [2]. This paper argues that the successful integration of artificial intelligence into clinical workflows depends on a paradigm shift from black-box optimization to transparent, explainable systems that are designed from the ground up with clinical governance, system robustness, and fairness as primary requirements, rather than afterthoughts. The analysis adopts a systems engineering perspective to examine how explainability functions as a critical infrastructure component spanning data management, model architecture, human-computer interaction, and regulatory compliance.

2. Background and Systemic Context of EHR Time-Series Data

Electronic health record data fundamentally differ from the clean, regularly sampled time-series commonly found in other engineering domains. Patient data are generated through clinical necessity rather than experimental design, resulting in irregular temporal spacing, heterogeneous measurement modalities, and significant portions of missing observations [3]. Vital signs, laboratory results, medication administrations, and clinical notes each have distinct temporal cadences and are subject to different documentation practices across institutions. This structural irregularity challenges conventional time-series models that assume uniformly spaced intervals, requiring sophisticated data processing pipelines that can manage alignment, imputation, and variable-rate sampling without introducing systematic bias. Furthermore, the longitudinal nature of patient records means that individual trajectories can span years, creating massive storage and retrieval demands for operational systems [4]. From a systems perspective, the preprocessing infrastructure often constitutes the most complex and error-prone component of the entire decision support pipeline, yet it receives significantly less attention in academic literature than the model architecture itself. Any explainability mechanism that operates on the final model output must also account for the transformations applied upstream, as explanations can be rendered misleading if they ignore the impact of data filtering, imputation strategies, or cohort selection criteria.

3. Architectural Trade-offs in Model Design for Interpretability

The design of clinical decision support systems involves navigating a fundamental trade-off between model complexity and interpretability. Deep learning approaches such as long short-term memory networks and transformer architectures have demonstrated strong predictive performance on tasks like early sepsis detection and mortality prediction [5]. Yet these models operate through layers of non-linear transformations that are notoriously difficult to decompose into human-understandable reasoning. Conversely, inherently interpretable models like decision trees, logistic regression, and generalized additive models offer complete transparency but often underperform on capturing complex temporal dependencies [6]. A growing body of research has attempted to bridge this gap through post-hoc explainability techniques, including feature attribution methods, attention mechanisms, and surrogate models. However, these methods carry their own caveats, as they approximate but do not replicate the model's internal decision process, and their reliability can vary depending on the stability of the underlying features and the distribution of the training data [7]. The architectural decision thus becomes a governance choice: deploying a highly complex model with post-hoc explanations that may be inconsistent, versus deploying a simpler model whose lower accuracy is offset by complete transparency. This trade-off is not purely technical; it involves considerations about who bears the liability for incorrect predictions, how physicians will integrate explanations into their workflow, and what level of performance degradation is acceptable for improved interpretability.

4. Deployment Infrastructure and Operational Sustainability

Translating explainable AI models into bedside decision support requires a robust and sustainable deployment infrastructure that integrates with existing hospital information systems. Real-time clinical decision support must operate within strict latency constraints, as delays in processing incoming laboratory results or vital sign measurements can render recommendations clinically irrelevant [8]. The computational pipeline must stream data from the electronic health record, validate and clean it, compute features, feed the input to the model, generate predictions, produce explanations, and deliver both to a clinical interface within seconds. This imposes stringent requirements on data storage, network bandwidth, and compute resource allocation. Moreover, the system must be designed for continuous monitoring and retraining, as clinical data distributions shift over time due to changes in practice patterns, population demographics, or treatment protocols [9]. Explainable models offer a particular advantage in this context, as stability monitoring tools can track changes in feature importance and decision boundaries over time, providing early warning of model drift. The sustainability of such a system also depends on its maintainability by hospital information technology staff who may not have deep machine learning expertise. This argues for modular architectures that separate data ingestion, model inference, explanation generation, and user interface into independently maintainable components, each with clear performance metrics and error handling protocols.

5. Interpretability as a Mechanism for Clinical Trust and User Adoption

The ultimate success of any clinical decision support system hinges on whether clinicians choose to engage with it. Numerous studies have demonstrated that physicians are reluctant to act on recommendations from black-box models, even when those models are objectively accurate [10]. This resistance stems from a legitimate professional responsibility to understand the rationale behind a clinical decision. Explainability functions not merely as a technical feature but as a communication channel between the algorithmic system and the human expert. For an explanation to be effective, it must be tailored to the cognitive context of the clinician, highlighting the specific patient factors that most influenced the prediction and presenting them in a clinically meaningful vocabulary [11]. For example, a prediction of impending decompensation should be accompanied by an explanation that identifies which vital sign trends or laboratory derangements drove the alert, expressed in terms a clinician can rapidly evaluate and potentially confirm. This requires close collaboration between model developers, user interface designers, and clinical domain experts during the deployment process. Furthermore, explanations must be robust and consistent; if a model provides contradictory explanations for similar patient profiles, or if explanations are driven by spurious correlations, trust will rapidly erode. Building this trust is a long-term process that involves iterative refinement, transparent reporting of model limitations, and mechanisms for clinician feedback that inform model updates.

6. Fairness, Bias, and the Diagnostic Role of Explainability

Historical electronic health record data are a product of existing healthcare disparities, documenting patterns of unequal access, differential treatment, and systematic underdiagnosis in certain populations [12]. Machine learning models trained on such data risk perpetuating and even amplifying these biases, leading to decision support systems that are less accurate or less informative for minority groups. Addressing algorithmic fairness in clinical AI requires more than demographic parity metrics; it demands a structural understanding of how biases enter the data generation process and propagate through model training. Explainability techniques provide a powerful diagnostic tool for this purpose [13]. By examining which features drive predictions for different subpopulations, researchers can identify whether the model has learned proxy variables for race, socioeconomic status, or insurance type. For example, an explainability analysis might reveal that a model for readmission risk is heavily weighting features related to healthcare utilization, which are themselves correlated with access barriers rather than clinical acuity. Such findings can guide both model remediation and broader policy interventions to address upstream social determinants of health. However, it is critical to recognize that explainability alone cannot guarantee fairness; it can only surface potential issues that then require deliberate action by system designers, healthcare administrators, and policymakers. A comprehensive fairness framework must include rigorous audit processes, differential validation across demographic subgroups, and mechanisms for accountability when disparate impacts are discovered.

7. Regulatory Compliance and Governance Frameworks

The deployment of AI-assisted clinical decision support systems is increasingly subject to regulatory oversight, particularly in jurisdictions with stringent medical device approval processes. Regulatory bodies are grappling with how to evaluate not just model accuracy but also the transparency and safety of algorithmic recommendations [14]. Explainability is emerging as a key requirement in this evolving landscape, as regulators seek assurance that models do not rely on spurious correlations or hidden confounders that could lead to harmful recommendations. The governance of these systems extends beyond initial approval to encompass post-market surveillance, requiring ongoing monitoring of model performance and explanation consistency [15]. This creates a need for audit trails that log every prediction and its associated explanation, enabling retrospective analysis in the event of an adverse outcome. Governance frameworks must also define who is responsible when a system recommends an incorrect course of action: the model developer, the deploying institution, or the clinician who accepted or overrode the recommendation. These questions of accountability are profoundly influenced by the transparency of the system, as opaque models make it impossible to determine whether a fault lies in the data, the algorithm, or the decision context. Establishing clear governance policies is essential for building the legal and ethical foundation upon which widespread clinical adoption can be built.

8. Cross-Domain Comparisons and Lessons Learned

The challenges of deploying explainable AI in healthcare are not unique, and valuable lessons can be drawn from other high-stakes domains such as autonomous driving, financial credit scoring, and predictive policing. In each of these domains, the tension between complex model performance and the need for human oversight has led to different regulatory and design responses [16]. The financial sector, for example, has long required that credit denial decisions be accompanied by specific, auditable explanations, leading to the widespread adoption of inherently interpretable models. The autonomous vehicle industry, by contrast, has pursued complex perception systems with post-hoc explanation mechanisms for black-box event analysis. Healthcare faces a hybrid challenge: the decision-making context is as ethically charged as credit scoring, with consequences as severe as autonomous driving, yet the complexity of physiological time-series data makes simple models insufficient. The emerging consensus is that hybrid architectures, combining deep learning components for feature extraction with intrinsically interpretable decision layers, may offer the best balance of accuracy and transparency [17]. Furthermore, cross-domain experience emphasizes the importance of human-centered design, where the explanation format and timing are optimized for the specific decision-making context rather than providing generic feature importance plots.

9. Sustainability and Environmental Considerations

As computational demands of deep learning models continue to grow, the environmental footprint of AI systems has become an important sustainability consideration. Deploying large transformers or ensemble models for clinical decision support at thousands of hospital beds requires substantial energy for both inference and model retraining [18]. Sustainability is not merely an environmental concern; it directly impacts the economic feasibility of deploying AI in resource-constrained healthcare settings, including community hospitals and clinics in lower-income regions. Explainable models that achieve acceptable performance with simpler architectures can significantly reduce energy consumption and hardware requirements, making advanced decision support more accessible. Moreover, the long-term sustainability of any AI system depends on its maintainability, which is enhanced by transparency. Models that can be understood and debugged by a broader range of technical staff are less likely to be abandoned when the original developers move on. From a systems perspective, sustainability should be evaluated not just as energy per inference but as total cost of ownership over the system lifecycle, including development, validation, deployment, monitoring, and eventual decommissioning.

10. Conclusion

This paper has presented a systems-oriented analysis of explainable AI-assisted clinical decision support using electronic health record time-series data, emphasizing the structural, architectural, and governance dimensions that determine real-world impact. The transition from promising research algorithms to trusted clinical tools requires careful consideration of data irregularity, the trade-off between accuracy and interpretability, the demands of deployment infrastructure, and the necessity of building trust through transparent explanations. Explainability is not a single technical feature but a systemic property that must be designed into every layer of the system, from data preprocessing to user interface design. The diagnostic power of explainability for detecting bias and model drift makes it an indispensable tool for ensuring fairness and robustness in clinical AI. Regulatory frameworks are evolving to require transparency, and the field can benefit from cross-domain experiences in high-stakes decision making. Ultimately, the goal is not to replace clinical judgment but to augment it with algorithmic insights that are presented in a form that respects the complexity and responsibility of medical decision-making. Future research should focus on developing standardized evaluation protocols for explanation quality, establishing best practices for governance and audit, and designing hybrid architectures that combine the strengths of deep learning with the clarity of interpretable models. The most successful clinical decision support systems of the next decade will be those that are not only intelligent but also trustworthy, transparent, and deeply integrated into the human ecosystem of care.

References

1. Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 25(1), 44–56.

2. Rajpurkar, P., Chen, E., Banerjee, O., & Topol, E. J. (2022). AI in health and medicine. Nature Medicine, 28(1), 31–38.

3. Jensen, A. B., Moseley, P. L., Oprea, T. I., Elledge, S. J., & Brunak, S. (2017). Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients. Nature Communications, 8(1), 650.

4. Shickel, B., Tighe, P. J., Bihorac, A., & Rashidi, P. (2017). Deep EHR: A survey of recent advances in deep learning techniques for electronic health record analysis. IEEE Journal of Biomedical and Health Informatics, 22(5), 1589–1604.

5. Futoma, J., Simons, M., Panch, T., Doshi-Velez, F., & Celi, L. A. (2020). The myth of generalisability in clinical machine learning. NPJ Digital Medicine, 3(1), 1–5.

6. Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82–115.

7. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206–215.

8. Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A., & Escobar, G. (2014). Big data in health care: Using analytics to identify and manage high-risk and high-cost patients. Health Affairs, 33(7), 1123–1131.

9. Nestor, B., McDermott, M. B. A., Chauhan, G., Naumann, T., Hughes, M. C., Wang, S., ... & Ghassemi, M. (2019). Advances in machine learning for clinical decision support. Journal of the American Medical Informatics Association, 26(8-9), 801–806.

10. Cabitza, F., Rasoini, R., & Gensini, G. F. (2017). Unintended consequences of machine learning in medicine. JAMA, 318(6), 517–518.

11. Tonekaboni, S., Joshi, S., McCradden, M. D., & Goldenberg, A. (2019). What clinicians want: Contextualizing explainable machine learning for clinical end use. In Machine Learning for Healthcare Conference (pp. 359–380). PMLR.

12. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447–453.

13. Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.

14. Gerke, S., Minssen, T., & Cohen, G. (2020). Ethical and legal challenges of artificial intelligence-driven healthcare. In Artificial Intelligence in Healthcare (pp. 295–336). Academic Press.

15. Challen, R., Denny, J., Pitt, M., Gompels, L., Edwards, T., & Tsaneva-Atanasova, K. (2019). Artificial intelligence, bias and clinical safety. BMJ Quality & Safety, 28(3), 231–237.

16. Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. In Proceedings of the Conference on Fairness, Accountability, and Transparency (pp. 59–68).

17. Lipton, Z. C. (2018). The mythos of model interpretability. Queue, 16(3), 31–57.

18. Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2019). Green AI. Communications of the ACM, 63(12), 54–63.