



Connected Evidence: Scalable Graph Risk Prediction

Ganesh Pambala

Independent Researcher

ganeshpambal@gmail.com

Abstract

Scalable graph intelligence offers a way to integrate clinical knowledge in real time to support immediate clinical decision-making — particularly knowledge expressed as feature-extraction rules for clinical risk stratification tasks such as early detection of acute disease. By modeling entities as graphs, structured data streaming from multiple hospitals can be represented as a temporal graph, capturing either the temporal or causal structure of the data depending on whether it is used for rapid, real-time risk stratification or stored for later analysis. Both representations serve a dual purpose: meeting the immediate need to extract features for risk stratification, and subsequently supporting unified, real-time integration of clinical knowledge.

Existing research on applying graphs to clinical data typically adopts one of two narrow perspectives — knowledge graphs for integrating clinical rules, or temporal graphs for risk stratification — treating them as largely separate problems. In practice, however, the two are closely linked: clinical knowledge is readily translated into features, and risk-stratification rules inevitably carry an implicit temporal structure. As a result, clinical knowledge is often captured concisely in rule form, and rule-based risk stratification approaches remain widely used. Yet these rules remain constrained by the limitations of manual knowledge engineering, underscoring the need for pathways that enable real-time, automated knowledge integration. Moreover, the feature sets produced by current approaches often prove insufficient for the early detection of acute, life-threatening conditions.

Keywords: Real-time clinical decision support, scalable graph intelligence, entity and relationship modeling,

temporal graphs, risk stratification, causal inference, event sequencing, ensemble learning.

1. Introduction

Scalable graph intelligence aims to facilitate real-time integration of clinical knowledge across heterogeneous distributed sources for patient risk stratification. The hypothesis is that risk features may be efficiently extracted from an evolving temporal graph encompassing core entities, interrelations, and events known to influence risk. Current approaches to clinical knowledge integration and patient risk prediction are largely based on SQL, limiting the ability to leverage knowledge and data across institutional/organizational boundaries. This underutilization of established clinical terminologies and ontologies may contribute to unsatisfactory predictive performance for many common diseases. A scalable architecture for integrating clinical knowledge across schema-mapped heterogeneous distributed data sources is proposed.

Table 1. Core Concepts of the Proposed System

Concept	Short Description
Scalable Graph Intelligence	Integrates clinical knowledge from multiple sources in real time
Temporal Graph	Represents patient events, diagnoses, procedures, and outcomes over time



Concept	Short Description
Risk Stratification	Predicts patient risk levels within a fixed clinical time horizon
Clinical Knowledge Integration	Combines rules, features, and ontologies into a unified graph model
Real-Time Decision Support	Supports fast clinical inference for early disease detection

Graphed schemas support knowledge integration and enable encoding of rich temporal patterns and causal structures for risk stratification. Recent graph-based knowledge representation methods offer important advantages: a single distributed knowledge repository may support knowledge representation, inference, and reasoning while remaining open to future enhancements and extensions. Knowledge graphs merge data across sources within a unified graph schema, enabling symbol manipulation capability anticipated by the creators of artificial intelligence and easily extending to other domains of human knowledge.

1. Clinical Graph Representation

$$G_t = (V_t, E_t, X_t)$$

A. Entity and Relationship Modeling for Healthcare

Entity-relationship modeling offers a tractable approach to graph schema structuring. Core entities typically include patients, healthcare providers, diagnoses, procedures, medications, laboratory test results, and outcome markers. Relationships of immediate concern

include the mapping from a patient to a diagnosis, the mapping from a healthcare provider to an administered procedure, and the mapping from a test result to a diagnosis indicating elevated risk. Refinements of these relationships based on data provenance can enable expanded use of time-stamped data, for example by incorporating lab result time stamps into risk stratification. By systematically expanding the core entity set to include such relationships or capturing additional entities of interest, such as comorbidities and risk factors, the risk stratification model can potentially be made more accurate.

Table 2. Major Clinical Entities in the Graph Model

Entity	Role in the System
Patient	Central node for risk prediction
Diagnosis	Indicates disease conditions and clinical history
Procedure	Captures clinical interventions
Medication	Represents prescribed or administered drugs
Laboratory Result	Provides measurable clinical evidence
Outcome Marker	Supports risk classification and prediction

Use of a medically aligned graph schema can facilitate integration of domain knowledge with patient temporal data on demand. A PROMIS SF score corresponding to at least a moderate level of pain in the past seven days could be indicated as clear data, for example, by a proven, actively monitored diagnosis for chronic pain. In contrast to post hoc, machine-learning strategies, the graph-based approach offers natural opportunities for rule-based reasoning. For example, if a patient has had a principal



diagnosis of heart disease for ten years and another of diabetes for five years, one might assign those conditions an indication of poor status on the basis of known relationships rather than on a less interpretable inductive model.

2. Patient Risk Score

$$R_i = f(X_i, G_t)$$

2. Background and Motivation

Existing efforts for real-time clinical decision support rely on machine learning models that consume rich data streams to infer clinical knowledge. The present approach considers knowledge integration, risk stratification, and a full representation of reasoning capabilities as first-class citizens and exploits graph-based representations. Moreover, temporal information and the causes of clinical events are captured, enabling either investigative queries or refinements to static temporal knowledge for risk stratification. Graph-based approaches frequently regard their temporal dimension as superfluous or implement it as a straightforward Boolean predicate. Temporal graphs, however, enforce strict constraints on edges and can ease reasoning by explicitly representing the ordering of events. Temporal graphs can be used to align heterogeneous data sources, while machine learning models trained on such sources can infer possible future clinical events. The underlying causal structures of these events can then be identified and used to reason about their potential occurrence.

Table 3. Data Sources Used for Clinical Integration

Data Source	Purpose
Electronic Health Records	Patient history and clinical events

Data Source	Purpose
Laboratory Information Systems	Test results and lab observations
Health Information Exchanges	Cross-institutional patient data sharing
Clinical Decision Support Systems	Rule-based and model-based recommendations
Monitoring Systems	Streaming patient condition updates

Two types of clinical decision support are considered: the integration of clinical knowledge and the stratification of risk for a specific clinical outcome within a fixed time horizon. Both share the objective of capturing the interaction between patients and their healthcare environment. Clinical knowledge integration identifies relationships among entities through observation, while stratification seeks to predict future entities and relationships. These capabilities encompass a wide range of temporal reasoning in graphs; some specialized categories such as workflow automation are not investigated here.

3. Temporal Edge Relation

$$e_{ij}^t = (v_i, v_j, t)$$

A. Temporal and Causal Reasoning in Graphs

Temporal graphs are commonly used when temporal information is important for the semantics of the model. In such cases, ML algorithms can benefit from having the temporal dimension present in the graph itself. A temporal graph is a graph whose edges are active at certain points in time. In a temporal graph, it is a special case when

edges are active in continuous intervals. The temporal graph can be regarded as a graph stream where a constant number of vertices is kept in memory. Graphs with such temporal knowledge can also be used for enumerating time-constrained sub-graph patterns in temporal graphs. For risk stratification, events are often sequenced into a temporal-ordering treatment pathway to include longitudinal time-ordering information for patients.

Table 4. Real-Time Inference Architecture Components

Component	Function
Data Ingestion	Collects clinical data from multiple endpoints
Normalization	Cleans and harmonizes heterogeneous data
Graph Store	Maintains temporal clinical relationships
Feature Extraction	Converts graph patterns into risk features
Classification	Predicts risk category or clinical outcome
Inference Engine	Generates real-time clinical decision support

When performing causal inference operations on temporal data, the main goal is to examine whether it is possible to infer causal information from the non-experimental data when expert knowledge about the problem domain is unavailable or incomplete. Counterfactual reasoning is a particular form of causal inference that allows the system to derive plausible explanations for uncertain nodes in a causal graph by making successive, forced-explanation-requests. Another issue with temporal

data such as those in health care is the presence of various forms of uncertainty.

4. Graph Feature Extraction

$$Z_i = \phi(G_t, V_i)$$

3. Graph-Based Knowledge Representation in Clinical Domains

The knowledge underlying risk stratification has a temporal aspect and is often accumulated gradually over long time spans. It is therefore advantageous to maintain an easy-to-query graph representation of clinical events—diagnoses, procedures, medication administration and prescription—recorded in electronic health records. Such a representation can also support the integration of knowledge from predictive and other clinical models whose risk factors map to events in the electronic health record.

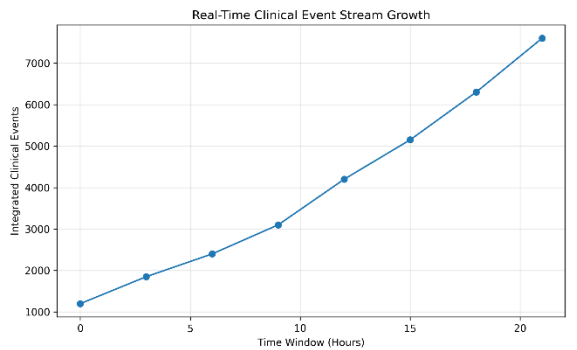


Fig 1: **System Architecture** — the six-component pipeline (ingestion → normalization → graph store → feature extraction → classification → inference engine) with the 5-second real-time target.

Temporal reasoning supports the designation of different realistic temporal relations among events (e.g., generally preceding a future event, preceding a future event in only a few instances, or determining the occurrence of a future event). Causal reasoning enables causal discovery from observational clinical data. It addresses questions involving “what-if” scenarios (counterfactual reasoning): for example, how a given event would have affected the outcome if it had taken place or been absent from the patient's clinical history.

5. Risk Probability

$$P(y_i = 1) = \sigma(WZ_i + b)$$

A. Ontologies and Shared Vocabularies

Domain-appropriate ontologies and shared vocabularies are key to effective and reliable AI methods in the healthcare domain, and mapping these standards for alignment should be a priority. A semantic monitoring and reasoning engine for risk stratification could operate a temporal graph containing the patients, procedures, medications and/or devices administered or engaged along with timestamps. Potential sources for risk could include the International Statistical Classification of Diseases and Related Health Problems (ICD), Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT), Logical Observation Identifiers Names and Codes (LOINC), and the Fast Healthcare Interoperability Resources (FHIR) vocabulary standard. The fact that different healthcare providers tend to favour different vocabularies leads to coordination challenges in heterogeneous health information deployments. Methods for governance and versioning would establish a semantically coherent ontology-process in the sense of aligning health risk with semantic severity levels in the language of health conditions.

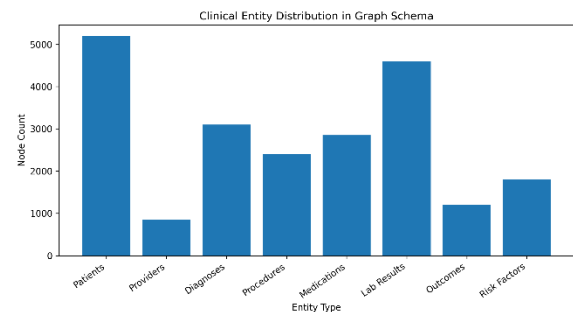


Fig 2: **Entity-Relationship Schema** — core patient-centered graph (provider, diagnosis, procedure, medication, lab result, outcome, comorbidity).

Mappings supplied by the Health Level Seven (HL7) standards organization ensure Semantic Interoperability between FHIR and the various classifications, categorizations and terminology standards of HL7, including the LOINC, SNOMED CT, and International Classification of Nursing Practice (ICNP). A mapping project executed by the National Library of Medicine was enabled to further align SNOMED CT content with most aspects of ICD-10-CM, producing the designated subsets of SNOMED CT-ICD-10-CM map files. Additionally, the Eclipse Foundation provided a FHIR-to-ICD-10-CM mapping for US health data.

6. Event Sequence Modeling

$$S_i = \{e_1, e_2, \dots, e_n\}$$

4. System Architecture for Real-Time Inference

The architecture for real-time inference is composed of six primary components: data ingestion, normalization, the graph store, feature extraction,

classification, and the inference engine (the last two together represent a real-time inference pipeline). Data are streamed into the real-time pipeline, processed, and aggregated into the graph store. Toward real-time knowledge integration, ready-to-use knowledge is extracted from the graph: knowledge in the form of features or rules is fed into a classification model or rules engine, respectively.

Data from clinical environments arrive at specified endpoints and undergo data ingestion and normalisation, all described systematically in the following subsections. Security is paramount; all personal health information is kept within a secured infrastructure, without touching outside systems, and is cryptographically encrypted to mitigate the risk of leaks. The objective is to clean, normalise, align, and update the graph-anonymized data ready for fast and high-availability access.

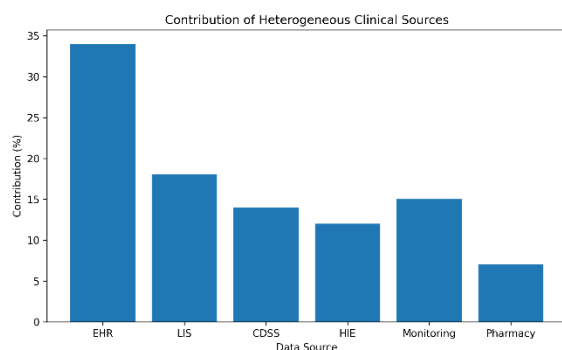


Fig 3: Temporal Graph Pathway — a time-ordered patient treatment sequence illustrating the paper's "treatment pathway" concept.

A. Data Ingestion and Normalization

Real-time clinical knowledge integration, risk stratification, and decision support require up-to-date clinical data from diverse heterogeneous sources within the

clinical environment. To this end, relevant data must be continuously ingested, normalized, and integrated into a common representational model, such as a temporal graph. The data ingestion component includes all mechanisms and processes for continuously extracting, cleaning, and harmonizing clinical data, thus preparing it for further processing by temporal graph engines.

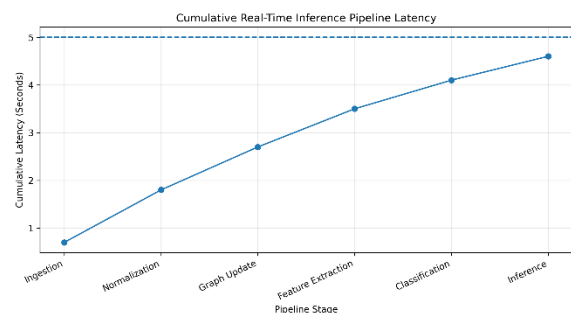


Fig 4: Causal/Counterfactual Graph — DAG showing causal structure plus the counterfactual "what-if" reasoning branch discussed in Section 2.A.

Clinical data are notoriously challenging to integrate due to inconsistencies, differences in schemas, heterogeneous sources, and temporal misalignment. Provenance disclosures, audit trails, and difficulty in handling schema evolution also exacerbate the problem. Common sources of clinical data are Electronic Health Records (EHRs), health information exchanges, Laboratory Information Systems (LIS), and Clinical Decision Support Systems (CDSS), among others. In addition to identification of common entities (patients, providers, locations, etc.), data integration involves resolving naming and representation variations among common entities and a temporally coherent merging of data from multiple sources.

7. Ontology Mapping Function

$$M: O_s \rightarrow O_t$$



5. Conclusion

The proposed approach contributes a model for clinic-level knowledge integration and risk stratification, based on scalable graph intelligence. It unites the strengths of existing temporal graph-based systems: simultaneously representing a clinical knowledge base, a feature set for real-time classifiers, and a logic for reasoning over the knowledge base, feature set, and patient temporal histories. The approach satisfies functional, operational, and performance requirements for integration and stratification within real-time clinical decision support. Assembling patient features and inferring patient outcomes within seconds of change at the underlying data sources positions the solution at the center of a typical real-time decision-support workflow.

8. Real-Time Update Rule

$$G_{t+1} = G_t + \Delta E_t + \Delta V_t$$

A methodology for aligning different views of the same core concept in a graph context is also proposed. In contrast to standard ontology-based alignment, which forces equivalence between notions in different source graphs, the presented procedure coordinates naming patterns, model scales, and ontological commitments, and also governs versioning of a shared list of entity categories. Seven ontologies from clinical decision-support research form one coordinate graph, while a support vector machine model for heart-attack risk prediction based on temporal medical histories becomes the second. Future work will apply the expanding body of temporal-graph-based causality-inference techniques to the clinic-level knowledge-graph integration case.

References

1. Fang, C., Xu, D., Su, J., Dry, J. R., & Linghu, B. (2021). DeePaN: Deep patient graph convolutional network integrating clinico-genomic evidence to stratify lung cancers for immunotherapy. *npj Digital Medicine*, 4, 14.
2. Lu, H., & Uddin, S. (2021). A weighted patient network-based framework for predicting chronic diseases using graph neural networks. *Scientific Reports*, 11, 22607.
3. Li, Y., Jin, Y., Song, G., Zhu, Z., Shi, C., & Wang, Y. (2021). GraphMSE: Efficient meta-path selection in semantically aligned feature space for graph neural networks. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(5), 4206–4214.
4. Yandamuri, U. S. (2022). Cloud-Based Data Integration Architectures for Scalable Enterprise Analytics. *International Journal of Intelligent Systems and Applications in Engineering*, 10, 472–483.
5. Li, Y., Tam, D. S. H., Xie, S., Liu, X., Ying, Q. F., Lau, W. C., Chiu, D. M., & Chen, S. Z. (2021). Temporal graph representation learning for detecting



- anomalies in E-payment systems. Proceedings of the 21st IEEE International Conference on Data Mining Workshops, 983–990.
6. Wang, C., Dou, Y., Chen, M., Chen, J., Liu, Z., & Yu, P. S. (2021). Deep fraud detection on non-attributed graph. Proceedings of the 2021 IEEE International Conference on Big Data, 5470–5477.
 7. Mangalampalli, B. M. Generative AI Applications In Healthcare Data Mart Design And Optimization.
 8. Xu, B., Shen, H., Sun, B., An, R., Cao, Q., & Cheng, X. (2021). Towards consumer loan fraud detection: Graph neural networks with role-constrained conditional random field. Proceedings of the AAAI Conference on Artificial Intelligence, 35(5), 4537–4545.
 9. Ma, X., Wu, J., Xue, S., Yang, J., Zhou, C., Sheng, Q. Z., Xiong, H., & Akoglu, L. (2023). A comprehensive survey on graph anomaly detection with deep learning. IEEE Transactions on Knowledge and Data Engineering, 35(12), 12012–12038.
 10. Fu, T., Huang, K., Xiao, C., Glass, L. M., & Sun, J. (2022). HINT: Hierarchical interaction network for clinical-trial-outcome predictions. Patterns, 3(4), 100445.
 11. An, Y., Li, R., & Chen, X. (2022). MERGE: A multi-graph attentive representation learning framework integrating group information from similar patients. Computers in Biology and Medicine, 151, 106245.
 12. Mangala, N. (2022). Implementing Databricks Unity Catalog For Centralized Data Governance In Multi-Business-Unitenterprises. Journal of International Crisis and Risk Communication Research, 101-122.
 13. Wang, Y., Wang, Y. G., Hu, C., Li, M., Fan, Y., Otter, N., Sam, I., Gou, H., Hu, Y., Kwok, T., Zalcborg, J., Boussioutas, A., Daly, R. J., Montúfar, G., Liò, P., Xu, D., Webb, G. I., & Song, J. (2022). Cell graph neural networks enable the precise prediction of patient survival in gastric cancer. npj Precision Oncology, 6, 45.
 14. Dong, X., Wong, R., Lyu, W., Abell-Hart, K., Deng, J., Liu, Y., Hajagos, J. G., Rosenthal, R. N., Chen, C., & Wang, F. (2023). An integrated LSTM-HeteroRGNN model for interpretable opioid overdose risk prediction. Artificial Intelligence in Medicine, 135, 102439.



15. Diaz Ochoa, J. G., & Mustafa, F. E. (2022). Graph neural network modelling as a potentially effective method for predicting and analyzing procedures based on patients' diagnoses. *Artificial Intelligence in Medicine*, 131, 102359.
16. Xu, Y., Ying, H., Qian, S., Zhuang, F., Zhang, X., Wang, D., Wu, J., & Xiong, H. (2023). Time-aware context-gated graph attention network for clinical risk prediction. *IEEE Transactions on Knowledge and Data Engineering*, 35(7), 7557–7568.
17. Mangala, N. (2022). Real-Time Data Quality Monitoring and Gating Frameworks in Cloud-Based Data Pipelines. *International Journal of Research and Applied Innovations*, 5(6), 8197-8219.
18. Fouladvand, S., Reyes Gomez, F., Nilforoshan, H., Schwede, M., Noshad, M., Jee, O., You, J., Sasic, R., Leskovec, J., & Chen, J. (2023). Graph-based clinical recommender: Predicting specialists procedure orders using graph representation learning. *Journal of Biomedical Informatics*, 143, 104407.
19. Fridgeirsson, E. A., Sontag, D., & Rijnbeek, P. (2023). Attention-based neural networks for clinical prediction modelling on electronic health records. *BMC Medical Research Methodology*, 23, 285.
20. Deng, A., & Hooi, B. (2021). Graph neural network-based anomaly detection in multivariate time series. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(5), 4027–4035.
21. Fu, B., Yu, X., & Feng, T. (2022). CT-GCN: A phishing identification model for blockchain cryptocurrency transactions. *International Journal of Information Security*, 21(6), 1223–1232.
22. Duan, X., Yan, B., Dong, A., Zhang, L., & Yu, J. (2022). Phishing frauds detection based on graph neural network on Ethereum. In *Wireless Artificial Intelligence and Security* (pp. 351–363). Springer.
23. Huang, M., Liu, Y., Ao, X., Li, K., Chi, J., Feng, J., Yang, H., & He, Q. (2022). AUC-oriented graph neural network for fraud detection. *Proceedings of the ACM Web Conference 2022*, 1311–1321.
24. Li, Q., He, Y., Xu, C., Wu, F., Gao, J., & Li, Z. (2022). Dual-Augment graph neural network for fraud detection. *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, 4188–4192.



25. Deng, Z., Xin, G., Liu, Y., Wang, W., & Wang, B. (2022). Contrastive graph neural network-based camouflaged fraud detector. *Information Sciences*, 618, 39–52. 110984.
26. Mangalampalli, B. M. (2023). AI-Driven Anomaly Detection in Healthcare Claims Data: A Business Intelligence Perspective. *Journal of Rare Cardiovascular Diseases*.
27. Zhang, G., Li, Z., Huang, J., Wu, J., Zhou, C., Yang, J., & Gao, J. (2022). eFraudCom: An e-commerce fraud detection system via competitive graph neural networks. *ACM Transactions on Information Systems*, 40(3), Article 47.
28. Wu, B., Chao, K.-M., & Li, Y. (2023). Heterogeneous graph neural networks for fraud detection and explanation in supply chain finance. *Information Systems*, 121, 102335.
29. Wang, X., Liu, Z., Liu, J., & Liu, J. (2023). Fraud detection on multi-relation graphs via imbalanced and interactive learning. *Information Sciences*, 642, 119153.
30. Tong, G., & Shen, J. (2023). Financial transaction fraud detector based on imbalance learning and graph neural network. *Applied Soft Computing*, 149,