Graph-Driven Anomaly Detection in EHR Data Using Isolation Forests and Large Language Models

Abe Mathew Mankavil
University of Texas at Austin

ABSTRACT

This paper addresses the challenge of identifying clinically meaningful anomalies in complex electronic health record (EHR) data. It presents a Knowledge Graph-based Retrieval Augmented Generation (KG-RAG) framework for anomaly detection within EHR data using structured knowledge graph representations, statistical modeling, and large language models (LLMs) to identify and interpret atypical patient information. These anomalies may include rare combinations of diagnoses and treatments, inconsistencies between lab results and presenting conditions, or demographic mismatches in care protocols. By modeling EHR data as interconnected graphs and applying unsupervised anomaly detection algorithms, this solution is able to identify anomalous cases without relying on labeled data. The graph-based representation enables the capture of both direct and multi-hop interactions among different clinical elements, enriching the context for anomaly detection. Node embeddings were generated for each admission subgraph using random walk–based methods and added to the graph as vector properties to support structure-aware anomaly detection. This paper focuses on the use of the isolation forest algorithm to detect anomalous admissions that diverge from typical patterns of care. To complement this detection process and further interpretability, the framework incorporates LLMs that generate natural language explanations for identified anomalies, ensuring transparent clinical insights. The proposed framework is applicable to a wide range of healthcare settings and can be adapted to different data infrastructures and modeling tools.

1 INTRODUCTION

The detection of anomalies in clinical data is a critical task for ensuring patient safety, improving quality of care, and identifying latent patterns in healthcare delivery. Electronic health records offer a rich view of patient encounters, but their inherent complexity pose challenges for traditional anomaly detection approaches. Hospital admissions, for example, may involve dozens of interconnected events whose clinical relevance depends not only on individual features, but also on how these features interact. As a result, there is a growing need for methods that can capture both the structural and contextual nuances of healthcare data to expose clinically significant irregularities.

This paper proposes a Knowledge Graph-based Retrieval Augmented Generation (KG-RAG) framework for anomaly detection in EHR data, leveraging the expressive power of graph representations and the unsupervised capabilities of Isolation Forests. Patient records are modeled as a knowledge graph in which admissions serve as central nodes connected to related clinical entities such as diagnoses, medications, lab results, and demographic features. This structure enables the modeling of complex, multi-hop relationships that are often lost in flattened tabular data. To detect anomalies, node embeddings were extracted that capture the local structure of each admission subgraph, and combined with clinical and demographic node attributes to construct a comprehensive feature representation. Isolation Forests are then used to identify admissions that deviate from normative care patterns—highlighting rare medication-lab-diagnosis combinations, treatment-demographic mismatches, or unusual sequences of care.

To interpret these anomalies in a clinically meaningful way, the framework incorporates large language models (LLMs) to generate natural language explanations. Rather than relying on fixed templates or manual review, a retrieval-augmented generation (RAG) system is setup such that the LLM dynamically queries the underlying graph and constructs a narrative that contextualizes why a particular case may be anomalous. This combination of unsupervised detection and language-based explanation allows for a scalable and interpretable approach to clinical auditing.

By integrating graph-based modeling, machine learning, and generative language technologies, this work offers a generalizable method for detecting and explaining irregularities in complex EHR datasets. The approach has the potential to support applications in a wide variety of future healthcare systems.

1.1 Related Work

Recent work at the intersection of graph modeling, machine learning, and LLMs has significantly advanced anomaly detection in EHR data. Khan et al. demonstrated the effectiveness of node embedding–based graph autoencoders in identifying adverse pregnancy outcomes, underscoring the value of learning compact graph representations for outlier detection [1]. Our project builds on this idea by generating node embeddings for admission subgraphs, enabling structurally informed detection of anomalous clinical events using Isolation Forests.

Efforts to improve interpretability in EHR-based prediction tasks have also gained momentum. Bopche's work applied explainable AI techniques to forecast bloodstream infections from time-series EHRs, highlighting the importance of avoiding "models that operate as *black boxes*" [2]. Our use of LLMs to produce natural language explanations for detected anomalies serves a similar purpose, ensuring that the reasoning behind each outlier is accessible to clinicians.

Soman and Gao explored ways to enhance LLM performance using biomedical knowledge graphs. These studies support our use of retrieval-augmented generation (RAG) techniques, where the LLM dynamically queries a knowledge graph to explain deviations in care, and motivate the use of graph structures to represent clinical information [3,4].

Lastly, Niu et al. show the effectiveness of graph-based anomaly detection across hospital networks, in their work combining machine learning with graph algorithms to uncover system-wide irregularities [5]. Our framework adopts a lot of the same graph-centric ideas from Niu et al. but operates at the patient-admission level, enabling fine-grained analysis of individual outliers in a clinical context.

Together, these studies form a foundation that our anomaly detection framework extends through the integration of structural embeddings, unsupervised learning, and generative explanation in one unified pipeline.

2 METHODS

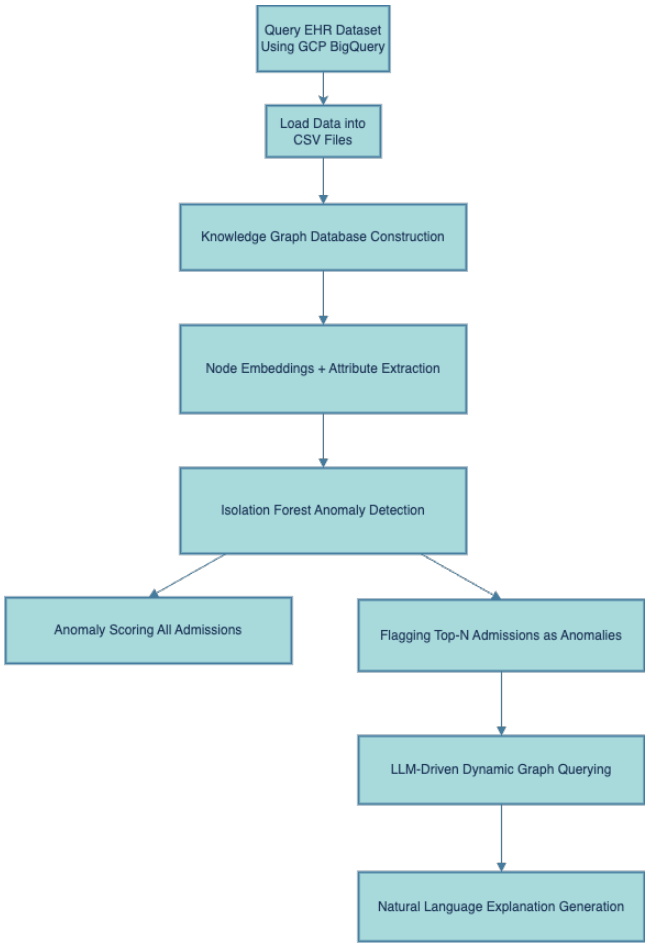

Figure 1: End-to-end workflow of EHR anomaly detection framework

2.1 Data Extraction and Preprocessing

The source data for this work was obtained from the MIMIC-III clinical database, a publicly available dataset containing de-identified health records from intensive care unit (ICU) admissions. Via Google BigQuery, relevant tables—patients, admissions, diagnoses, lab events, and medications—were extracted and preprocessed. The resulting datasets were exported as CSV files to create a static instance of the data.

2.2 Knowledge Graph Construction

The extracted CSV files were used to construct a structured clinical knowledge graph representing key entities and their relationships. Patients, admissions, diagnoses, laboratory results, and medications were modeled as distinct node types, each enriched with relevant attributes such as age, gender, diagnosis codes, and others. Relationships captured clinically meaningful associations between these nodes, including admissions linked to patients, diagnoses assigned during hospital stays, medications prescribed, and laboratory tests performed. See Figure 2. This property graph structure allowed both clinical entities and their features to coexist within the same framework, preserving the natural relational complexity of each hospital admission.

2.3 Graph Projection and Feature Engineering

To prepare the graph for complex statistical analysis, a vector projection was performed, focusing on admissions and their immediate clinical neighbors (diagnoses, labs, medications, and patient demographics). Node2Vec, an unsupervised graph embedding algorithm, was applied to generate low-dimensional vector representations of admissions, capturing the underlying graph topology and relational patterns. In addition to structural embeddings, node attributes such as gender, ethnicity, and hospital expiration were extracted to supplement the embeddings. Features like medications, labs, and diagnosis codes were all encoded using multi-label binarization. Categorical variables such as gender, ethnicity, and admission type were one-hot encoded, while medications and lab results were separately encoded based on both their names and associated dosage or values.

The final feature vector for each admission was a combination of structural embeddings and clinical attributes, enabling the anomaly detection model to leverage both relational and non-relational signals.

2.4 Anomaly Detection

Anomalies in hospital admissions were identified using the Isolation Forest algorithm, an unsupervised method designed to isolate rare or unusual data points in high-dimensional feature spaces. The combined feature matrix was used to train the Isolation Forest, with a contamination rate of 5% to approximate the proportion of anomalies expected in the dataset. Each admission received an anomaly score, with lower scores indicating greater divergence from typical patterns.

2.5 Anomaly Interpretation via LLMs

For flagged anomalous admissions, interpretability was achieved through dynamic graph querying and natural language explanation generation using LLMs. Upon detecting an anomaly, a natural language prompt was crafted instructing the LLM to query the underlying graph structure dynamically. Using the LangChain framework, the model generated graph queries in real time to retrieve contextual patient-admission information. The retrieved information was synthesized into concise natural language explanations, highlighting why the admission was identified as an anomaly.

2.6 End-to-End System Flow

The full pipeline consists of the following steps: (1) EHR data extraction from BigQuery, (2) construction of a clinically structured knowledge graph, (3) graph embedding and attribute extraction, (4) unsupervised anomaly detection using Isolation Forests, and (5) natural language explanation of flagged anomalies via dynamic LLM-driven graph querying. A framework that is easily scalable

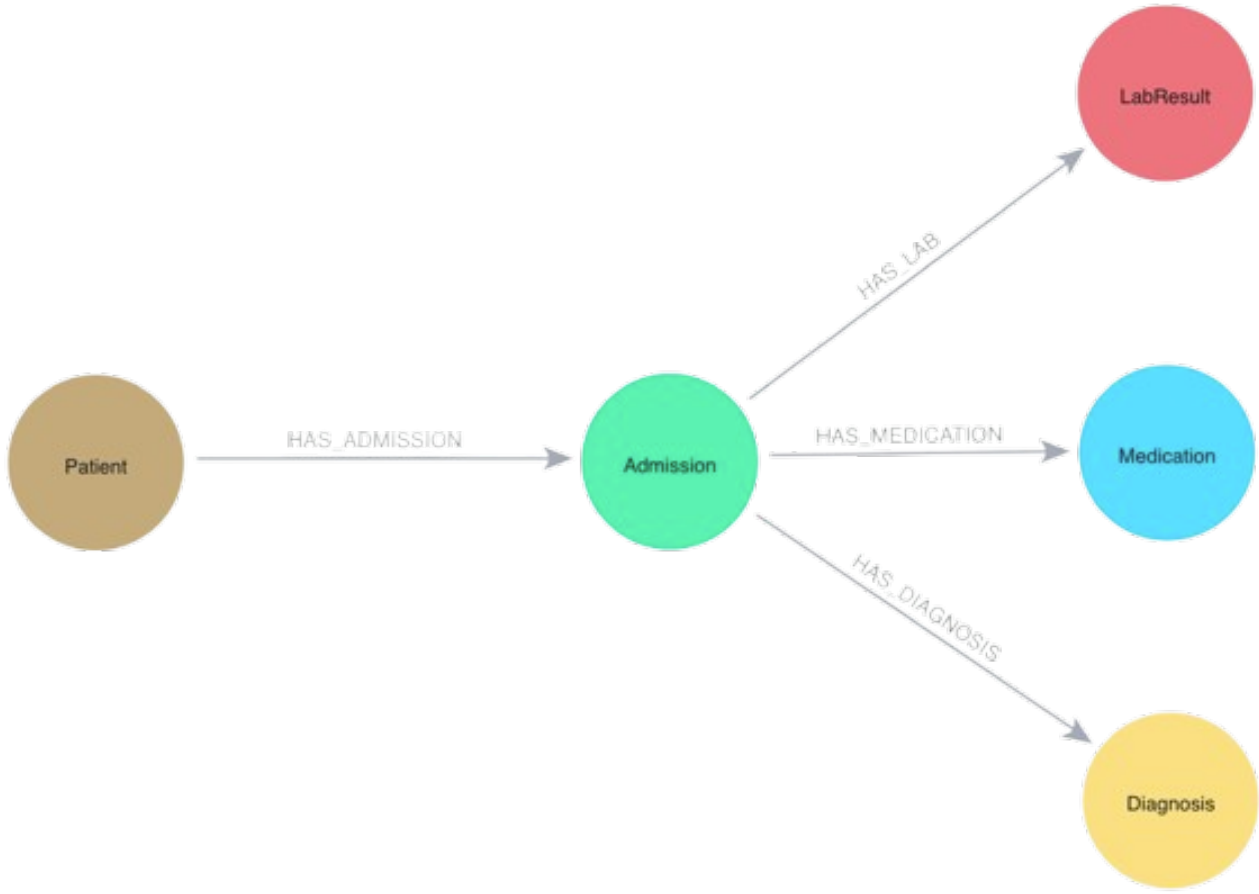

Figure 2: Knowledge graph schema showing node types and their relationships

for the identification of atypical admissions within larger and more complex EHR datasets.

3 RESULTS

The isolation forest algorithm was applied on the computed feature space of hospital admissions, producing a continuous anomaly score for each case. These scores reflect the degree of deviation from expected patterns across demographic, diagnostic, laboratory, medication, and graph-structural attributes. Admissions with the lowest 5% of scores were flagged as anomalous for further investigation.

3.1 Anomaly Score Distribution

The distribution of anomaly scores is visualized in Figure 3. Most admissions exhibit high anomaly scores clustered between 0 and 0.10, indicating strong alignment with typical clinical profiles. However, a long left-skewed tail emerges, capturing lower scores associated with potential outliers. The red dashed line represents the average score of flagged anomalies, revealing a clear divergence from the score distribution's peak. This demonstrates the model's ability to isolate admissions that differ meaningfully from the norm, rather than flagging noise or uniformly distributed points.

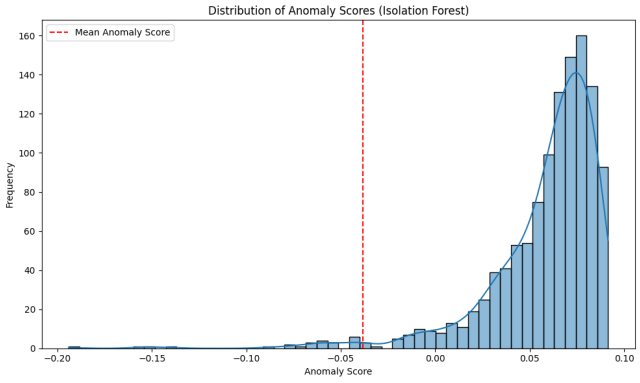

Figure 3: Distribution of anomaly scores assigned by an Isolation Forest. Red dashed line represents the average score of flagged anomalies

3.2 Lower-Dimensional Visualization of Admission Features

To investigate how anomalous admissions are distributed relative to the broader patient population, a t-distributed Stochastic

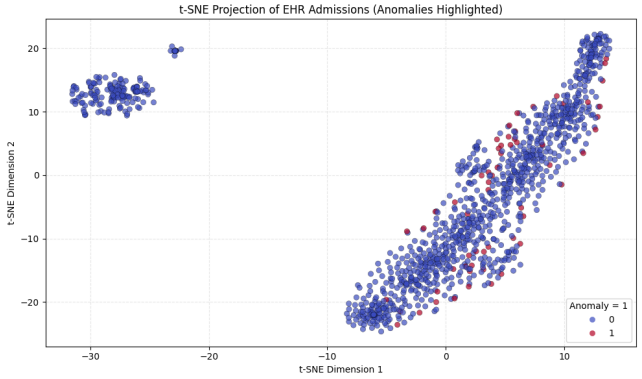

Figure 4: t-SNE projection of patient admissions onto two components. Anomalies are represented by red points

Neighbor Embedding (t-SNE) projection was applied to the high-dimensional feature matrix. This projection reduces the full representation—which includes graph embeddings, demographics, diagnoses, labs, and medication features—into a two-dimensional space suitable for visual interpretation. t-SNE excels at revealing clusters and small local anomalies.

As shown in Figure 4, most admissions form a dense cluster, suggesting shared clinical characteristics across the dataset. In contrast, admissions identified as anomalies by the Isolation Forest appear as scattered red points, often located at the periphery or outside of these dense clusters. Some anomalies are entirely isolated, while others lie within tight local groupings, indicating both global and context-specific deviations.

3.3 Case-Level Interpretation

Anomalous admissions were analyzed for clinicians by an LLM, which both generated dynamic graph queries and synthesized interpretations from the retrieved information. In one instance, the admission of a 38-year-old male was flagged because of an age-diagnosis mismatch. The patient was diagnosed with Acute Myelogenous Leukemia, something that according to the American Cancer Society *"is uncommon in people under the age of 45"*. The presence of psychiatric medications like Quetiapine, in combination with abnormal lab values such as elevated LDH and atypical lymphocytes, contributed to the anomaly flag.

3.4 Results Summary

Together, the anomaly score distribution, cluster projection, and case-level interpretations demonstrate the model's ability to recognize clinically meaningful deviations in EHR data. The framework enables not only statistical anomaly detection but also a semantic interpretation of the results, providing a foundation for clinical auditing.

4 CONCLUSION

This paper presented a KG-RAG framework for the detection and interpretation of clinically meaningful anomalies within electronic health record (EHR) data. By modeling hospital admissions as an interconnected knowledge graph, and applying node embeddings, we constructed a comprehensive feature representation of each patient encounter. Using the Isolation Forest algorithm, the framework effectively identified admissions that diverged from expectation.

Interpretability, often a major barrier in unsupervised anomaly detection, was addressed through the integration of large language models. Dynamic graph querying and retrieval-augmented generation(RAG) enabled the system to generate natural language explanations for each flagged admission, highlighting demographic mismatches, unusual medication-diagnosis pairings, and rare lab result patterns. These explanations provide essential context for clinicians seeking to validate or further investigate detected anomalies.

Overall, the integration of knowledge graphs, unsupervised detection, and LLM-based explanation offers a scalable, transparent, and clinically relevant method for exploring complex EHR datasets. Future work may extend this framework by incorporating temporal graph dynamics, evaluating detection outcomes against clinical expert labels, and adapting the approach to analyze other mediums for anomaly detection (e.g. images, sensor data, and genomic sequences). Future work could focus on calculating more accurate vectorization for complex features like lab and drug data.

5 CODE AVAILABILITY

The source code supporting this study is available at:
https://github.com/amankavil11/EHR_Anomaly_Detection/

6 REFERENCE LINKS

[1] Khan, W., Zaki, N., Ahmad, A. et al. Node embedding-based graph autoencoder outlier detection for adverse pregnancy outcomes. *Sci Rep 13*, 2023. https://doi.org/10.1038/s41598-023-46726-4

[2] Bopche, R., Gustad, L. T., Afset, J. E., Ehrnström, B.et al. Leveraging explainable artificial intelligence for early prediction of bloodstream infections using historical electronic health records. *PLOS Digital Health 3*, 2024. https://doi.org/10.1371/journal.pdig.0000506

[3] Soman, K. et al. Biomedical knowledge graph-optimized prompt generation for large language models. *Bioinformatics 40* 2024. https://doi.org/10.1093/bioinformatics/btae560

[4] Gao Y., Li R., Croxford E., Caskey J. et al. Leveraging Medical Knowledge Graphs Into Large Language Models for Diagnosis Prediction: Design and Application Study. *JMIR AI*, 2025. https://doi.org/10.2196/58670

[5] Niu, H., Omitaomu, O. A., Langston, M. A., Grady, S. K. et al. Anomaly Detection in Electronic Health Records Across Hospital Networks: Integrating Machine Learning With Graph Algorithms. *IEEE Journal of Biomedical and Health Informatics*, 2024. https://doi.org/10.1109/JBHI.2025.3527752