GraphEHR: Heterogeneous Graph Neural Network for Electronic Health Records

Minyoung Choe[1 *], **Juho Jung**[2 *], **Kushagra Agarwal**[3 *] and **Nivedhitha Dhanasekaran**[3 *]
[1]KAIST
[2]Sungkyunkwan University
[3]Carnegie Mellon University
minyoung.choe@kaist.ac.kr, jhjeon9@g.skku.edu, kagarwa2@andrew.cmu.edu, ndhanase@andrew.cmu.edu

Abstract

Electronic Health Records (EHR) are valuable for patient analysis and disease prediction using deep learning algorithms. However, previous approaches focused on single predictive tasks and feature selection, which may not be optimal in real clinical settings where all patient characteristics, including historical data, treatments, and past illnesses, are important. To address this limitation, we propose ***GraphEHR***, which leverages Heterogeneous Graph Neural Networks (HGNNs) for various predictive tasks using EHR data. ***GraphEHR*** aims to comprehensively predict multiple diseases by modeling all patient characteristics through a novel graph-based patient embedding. This approach effectively captures complex relationships within EHR and enables predictions across a wide range of diseases. In comparative experiments using the MIMIC-III EHR database, encompassing up to 13 predictive tasks, ***GraphEHR*** showcased its capability to grasp the complexity and multidimensional nature of EHR when compared against multiple baseline models across various tasks. By considering all aspects of a patient's medical history, this holistic modeling approach enhances clinical decision-making and facilitates patient management.

1 Introduction

Electronic Health Records (EHR) are digital patient charts that contain extensive patient information, including medical history, diagnoses, medications, treatment plans, and laboratory test results [Seymour *et al.*, 2012]. Given the potential of EHR to enhance patient care by streamlining processes and providing comprehensive medical histories, there has been a surge in deep learning-based approaches aimed at analyzing these digital clinical repositories [Shickel *et al.*, 2017; Zhu and Razavian, 2021; Choi *et al.*, 2020]. By incorporating historical diseases and treatment information along with multifactorial symptoms from EHR, various predictive modeling methods with EHR data have been introduced, aiming to advance personalized medicine and improve healthcare quality [Rajkomar *et al.*, 2018; Choi *et al.*, 2018].

These efforts to address healthcare predictive tasks using information from EHR necessitate modeling the relationships among various factors such as patient metadata (e.g., gender, age), past medical history, treatment records, and patient response to treatment. Since these factors affect and interact with each other, graphs have become a valuable tool for modeling multifactoral features in tabular EHR data, given their effectiveness in capturing relationships among numerous factors. However, previous studies have primarily concentrated on individual predictive tasks, overlooking the development of foundational models capable of addressing multiple predictive tasks using the abundant information available in EHR. By integrating multiple tasks into a single model, we can uncover new interactions among factors influencing diseases and symptoms, leading to enhanced effectiveness in addressing predictive models.

To address the intricacies and multi-dimensional structure of EHR, we propose a Heterogeneous Graph Neural Network (HGNN), *GraphEHR*, designed to tackle various predictive tasks effectively. By integrating patient-centric information from multiple hospital visits using a heterogeneous graph network, *GraphEHR* learns the complex relationships among medical concepts within EHR data. Additionally, through our proposed patient-embedding process, *GraphEHR* exhibits robustness in learning diverse graph structures, resulting in adaptive enhancements in performance across a spectrum of EHR predictive tasks. Extensive experiments conducted on widely-used EHR datasets such as MIMIC-III, along with comparisons against baselines, demonstrate *GraphEHR*'s capability to forecast the early stages of diseases like pancreatic cancer, which often manifest with similar initial symptoms. *GraphEHR* represents a significant advancement in fully leveraging EHR data across 13 predictive tasks, thereby demonstrating potential improvements in healthcare decision-making and patient outcome predictions.

2 Dataset: MIMIC-III

To extend the utility of *GraphEHR* and evaluate its extensibility for critical healthcare predictive tasks, we utilized the MIMIC-III [Johnson *et al.*, 2016] dataset. MIMIC-III is an open, accessible critical care database widely used for healthcare research, comprising records of 53,423 distinct

[*] These authors contributed equally to this work.

hospital admissions for adult patients spanning from 2001 to 2012. This dataset provides a wealth of information including demographics, vital signs, laboratory results, medication records, procedural details, diagnoses, and clinical notes. Within our study, we focused on 13 specific predictive tasks: Advanced Cancer, Advanced Heart Disease, Advanced Lung Disease, Alcohol Abuse, Chronic Neurological Dystrophies, Chronic Pain Fibromyalgia, Dementia, Depression, Developmental Delay, Non-Adherence, Obesity, Substance Abuse, and Schizophrenia and other Psychiatric Disorders.

2.1 Data Preprocessing

We extract various information from the complex relational files in the MIMIC-III dataset to create our features as follows:

- **Patient**: Details from the patient's initial hospitalization, such as Subject ID, HADM ID, ICU stay ID, gender, age, and obesity status.
- **Medication**: National Drug Code (NDC), Input CV item ID (liquids injected using the CareVue system), and Input MV item ID (liquids injected using the MetaVision system).
- **Procedure**: ICD9 billing codes assigned to surgeries or procedures undertaken.
- **Lab Result**: Item ID and corresponding value. The item ID indicates the type of measurement, such as red blood cells, while the value represents the actual measurement, such as the count of red blood cells for the patient.

Specifically, in cases where a patient (identified by Subject ID) made multiple visits to the emergency room (with unique ICU stay IDs) at the same hospital (with unique HADM IDs), multiple ICU stay IDs were assigned. Subject ID, HADM ID, and ICU stay ID serve as unique key values in this context. Our focus was on extracting specific details such as prescribed medications, procedures performed, and relevant measurements (e.g., blood pressure) from the dataset. The statistics of the dataset are summarized in Table 1.

3 Method: GraphEHR

To effectively manage the complexity of Electronic Health Records (EHR), which encompass diverse data types like medications, procedures, and lab results interconnected through relationships such as "causes" and "symptoms," we introduce *GraphEHR*. This approach involves constructing a heterogeneous graph from EHR data and employing a Heterogeneous Graph Neural Network (HGNN) to embed this multifaceted information, as depicted in Figure 1. Our methodology aims to enhance the model's ability to capture the intricate relationships within the data, thereby improving predictive accuracy concerning patient outcomes. Specifically, the heterogeneous graph helps in detecting patients with similar health profiles by recognizing patterns in their medical histories, such as commonly prescribed medications or frequently performed procedures. It also efficiently handles various types of nodes and edges, enabling a comprehensive representation of the EHR's complex data structure.

Table 1: Summary of Preprocessed MIMIC-III Dataset Statistics: Average counts of features per ICU stay and total counts across the dataset, with a distribution of diagnostic labels among patients.

(a) Feature Counts.

Feature	Avg. Count per ICU Stay / Total Count
Medication	28.9 / 2,042
Procedures	4.0 / 568
Lab Values	159.6 / 3,689
Patients	1,375

(b) Patient Counts with Positive Diagnostic Labels.

Diagnostic Label	Positive Counts
Mortality	828
Non-Adherence	120
Developmental Delay	29
Advanced Heart Disease	228
Advanced Lung Disease	137
Schizophrenia and Other Psychiatric Disorders	249
Alcohol Abuse	198
Other Substance Abuse	139
Chronic Pain Fibromyalgia	290
Chronic Neurological Dystrophies	324
Advanced Cancer	149
Depression	391
Dementia	104

3.1 Heterogeneous Graph Construction

We construct a heterogeneous graph comprising four node types: diagnoses, procedures, hospital visits, and lab results. Each hospital visit node is linked to nodes representing the diagnoses, procedures, and lab results associated with that visit. Additionally, hospital visit nodes are connected to each other if they correspond to the same patient. For node differentiation, diagnoses, procedures, and hospital visits are distinguished by their unique identifiers. Lab results, which are inherently continuous, are categorized into bins labeled high, medium, and low. This categorization discretizes the lab results, allowing us to treat these values as discrete nodes within the graph, facilitating more straightforward analysis and interpretation.

3.2 Learning Patient Representation

We utilize a Heterogeneous Graph Transformer (HGT) [Hu *et al.*, 2020] to extract node embeddings from the heterogeneous graph. The model incorporates unique embedding layers tailored for each node type and employs multi-head attention mechanisms to capture the relationships between neighboring nodes, taking into account their specific types and the types of edges connecting them. The final step in our approach involves making predictions using a linear layer that processes the embeddings of the most recent hospital visit nodes, thereby generating patient-specific outcomes.

4 Experiments

4.1 Experiment Setting

We assess our model across a variety of classification tasks, aiming to predict critical outcomes from patient data. Our

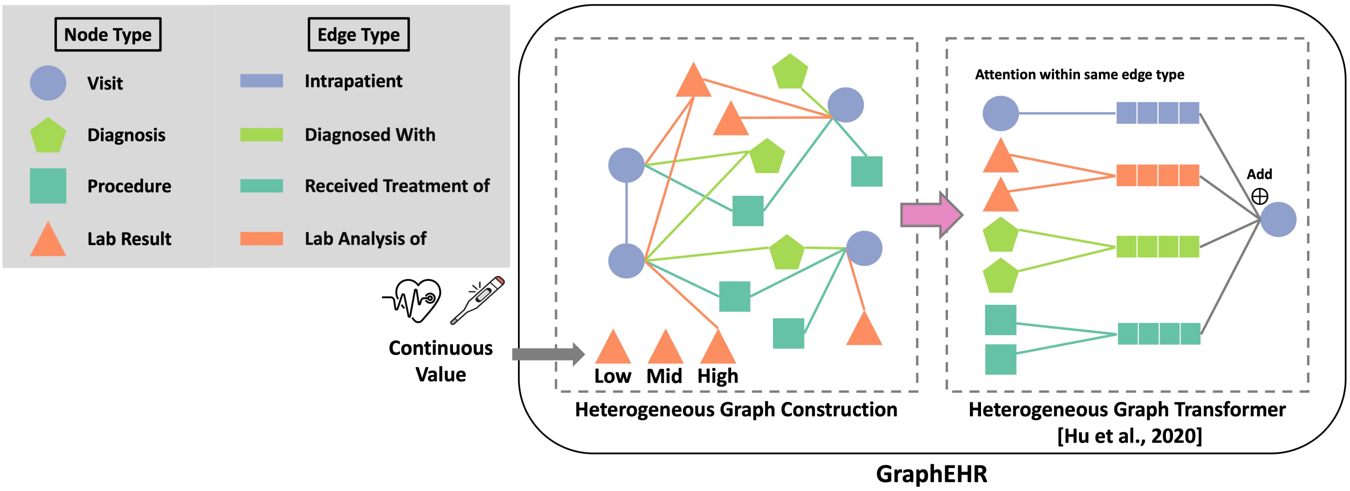

Figure 1: An overall architecture of the proposed method. The proposed heterogeneous graph identifies Electronic Health Record (EHR) features as distinct nodes, each connected by different edges to represent their relationships. The Graph Attention Transformer comprehensively understands the connections between these nodes and all patients, utilizing node embeddings to formulate representations. Finally, a task-specific representation is generated through the fusion of each node representation.

model is compared against seven baseline models, encompassing simple and competitive approaches, especially as we introduce novel prediction tasks beyond mortality. The evaluated models include: (1) *Logistic Regression* applies balanced class weights and predefined regularization parameters for binary classification tasks, (2) *Random Forest* utilizes decision trees with balanced class weights and ensemble learning, (3) *Gradient Boosting* employs a specific number of estimators to build weak learners sequentially, enhancing the model's predictive capabilities, (4) *SVM* (Support Vector Machine with Standard Scaler) uses balanced class weights and regularization, along with feature scaling, to identify optimal class separation, (5) *KNN* (K-Nearest Neighbors with Standard Scaler) classifies points based on the majority class among their nearest neighbors, (6) *MLP* is multilayer perceptrons with varying hidden layer sizes.

These models are refined with techniques like SMOTE for addressing class imbalance and include regularization to enhance their robustness. Additionally, the baseline *VGNN* [Zhu and Razavian, 2021] stands out for its graph-based approach, representing patients with nodes for diagnoses, procedures, and lab results. *VGNN* creates a complete graph with a self-attention mechanism for node updates, yet its patient-specific graphs limit broader analytical insights and face scalability issues with increasing features. We train *VGNN* for 50 epochs, using the initially proposed optimal hyperparameters except for the embedding size dimension. Due to memory constraints, we opt for a reduced embedding dimension of 128 instead of the original 256.

Our model introduces a heterogeneous graph embedding approach, incorporating node embeddings for each feature to capture inter-feature and patient-wide relationships. Unlike *VGNN*, which creates individual patient graphs, our model constructs a unified graph, enhancing comparative analysis across patients. Our model is trained for 200 epochs, selecting the iteration with the best validation performance.

4.2 Experimental Results

The results, detailed in Table 2, highlight the superior performance of our proposed model *GraphEHR* over the baseline models in almost all tasks, 9 out of 13. Notably, *VGNN* performs well in tasks like predicting advanced heart disease and psychiatric disorders, but its performance difference with *GraphEHR* is marginal. Impressively, *GraphEHR* outperforms *VGNN* significantly in predicting conditions like advanced lung disease or dementia, as depicted in the results. This superior performance is likely due to *GraphEHR*'s ability to effectively identify patients with common medications or procedures and its adept handling of various node and edge types. Moreover, the analysis with extensive predictive tasks indicates that simpler models like Logistic Regression and SVMs struggle with the complexity of EHR data, highlighting the necessity for more sophisticated, expressive models.

The Area Under the Precision-Recall Curve (AUPRC) generally appears lower in tasks other than mortality prediction, attributed to the class imbalance inherent in these tasks. This imbalance affects baseline models notably, with *VGNN* being especially sensitive to training data quality in imbalanced scenarios. For example, in dementia prediction, *VGNN*'s AUPRC of 0.08 highlights its struggle with imbalanced datasets.

Additionally, *GraphEHR* showcases a considerably shorter running time than *VGNN*, as illustrated in Table 3, demonstrating its efficiency. This is because *GraphEHR* utilizes the defined graph structure, whereas *VGNN* explores all potential connections between nodes to discover implicit connections, resulting in longer processing times for *VGNN*.

Table 2: Model evaluation of mortality prediction in our dataset using precision-recall curves (99% confidence interval)

Algorithm	LR	RF	GB	SVM	KNN	MLP	VGNN	GraphEHR
Mortality	0.68	0.65	0.68	0.72	0.61	0.73	0.76	**0.86**
Non-Adherence	0.08	0.07	0.09	0.06	0.06	0.06	0.23	**0.31**
Developmental Delay	0.33	0.01	0.10	0.01	0.01	0.33	0.23	**0.29**
Advanced Heart Disease	0.12	0.10	0.12	0.10	0.11	0.19	**0.56**	0.53
Advanced Lung Disease	0.10	0.12	0.10	0.09	0.10	0.14	0.16	**0.39**
Schizophrenia and Other Psychiatric Disorders	0.17	0.13	0.17	0.13	0.11	0.15	**0.46**	0.44
Alcohol Abuse	0.35	0.24	0.25	0.24	0.15	0.33	**0.52**	0.47
Other Substance Abuse	0.11	0.06	0.08	0.06	0.07	0.11	0.36	**0.42**
Chronic Pain Fibromyalgia	0.17	0.15	0.15	0.16	0.16	0.20	0.21	**0.44**
Chronic Neurological Dystrophies	0.23	0.24	0.25	0.24	0.25	0.28	0.29	**0.48**
Advanced Cancer	0.08	0.08	0.08	0.08	0.09	0.11	0.15	**0.30**
Depression	0.39	0.38	0.36	0.36	0.36	0.41	0.32	**0.54**
Dementia	0.10	0.11	0.26	0.09	0.09	0.29	0.08	**0.27**

Table 3: GraphEHR significantly outperforms VGNN regarding computational speed.

	VGNN	GraphEHR
Computation Time (Sec.)	13,308	921

5 Conclusion

In this paper, we introduced *GraphEHR*, aimed to revolutionize healthcare analytics by leveraging Graph Neural Networks (GNN) on Electronic Health Records (EHR) data. Through the proposed *GraphEHR*, we achieved a robust understanding of complex relationships, leading to enhanced diagnosis, treatment, and patient care. Moving forward, we plan to enhance interpretability by integrating techniques like attention mechanisms. This will shed light on influential nodes or edges in the heterogeneous graph, increasing trust in the model's decision-making process. Additionally, we aim to expand our dataset beyond MIMIC-III, considering alternatives like eICU to address limitations encountered during Exploratory Data Analysis. We envision our work benefiting medical professionals and patients, empowering informed decision-making, and improving healthcare outcomes.

Ethical Statement

There are no ethical issues.

Acknowledgments

This research was supported by the Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by Korea government (MSIT) (RS-2022-00143911, AI Excellence Global Innovative Leader Education Program).

References

[Choi *et al.*, 2018] Edward Choi, Cao Xiao, Walter Stewart, and Jimeng Sun. Mime: Multilevel medical embedding of electronic health records for predictive healthcare. *Advances in neural information processing systems*, 31, 2018.

[Choi *et al.*, 2020] Edward Choi, Zhen Xu, Yujia Li, Michael Dusenberry, Gerardo Flores, Emily Xue, and Andrew Dai. Learning the graphical structure of electronic health records with graph convolutional transformer. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 606–613, 2020.

[Hu *et al.*, 2020] Ziniu Hu, Yuxiao Dong, Kuansan Wang, and Yizhou Sun. Heterogeneous graph transformer. In *Proceedings of the web conference 2020*, pages 2704–2710, 2020.

[Johnson *et al.*, 2016] Alistair EW Johnson, Tom J Pollard, Lu Shen, Li-wei H Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. Mimic-iii, a freely accessible critical care database. *Scientific data*, 3(1):1–9, 2016.

[Rajkomar *et al.*, 2018] Alvin Rajkomar, Eyal Oren, Kai Chen, Andrew M Dai, Nissan Hajaj, Michaela Hardt, Peter J Liu, Xiaobing Liu, Jake Marcus, Mimi Sun, et al. Scalable and accurate deep learning with electronic health records. *NPJ digital medicine*, 1(1):1–10, 2018.

[Seymour *et al.*, 2012] Tom Seymour, Dean Frantsvog, Tod Graeber, et al. Electronic health records (ehr). *American Journal of Health Sciences (AJHS)*, 3(3):201–210, 2012.

[Shickel *et al.*, 2017] Benjamin Shickel, Patrick James Tighe, Azra Bihorac, and Parisa Rashidi. Deep ehr: a survey of recent advances in deep learning techniques for electronic health record (ehr) analysis. *IEEE journal of biomedical and health informatics*, 22(5):1589–1604, 2017.

[Zhu and Razavian, 2021] Weicheng Zhu and Narges Razavian. Variationally regularized graph-based representation learning for electronic health records. In *Proceedings of the Conference on Health, Inference, and Learning*, pages 1–13, 2021.