Building graph neural networks for precision medicine in patient classification

Haoxuan Weng
hw499@georgetown.edu

Chia-chien Chiang
chia-chien.chiang@astrazeneca.com

December 8, 2023

Abstract

In the field of biomedicine, the inherent complexity of biological systems precludes a universal drug treatment approach for specific diseases. Instead, heterogeneous knowledge graphs integrate diverse sources of biomedical knowledge, commonly referred to as prior biomedical knowledge. Our research is centered on the integration of patient-level gene transcriptomic data, gathered from two distinct studies, into a publicly available knowledge graph known as PrimeKG. Utilizing the Clinical Embedding of Patients (CLEP) framework, we aim to develop a binary patient classifier capable of predicting responsive patterns to drug treatments in inflammatory bowel disease. Our findings reveal that the novel patient representations generated by the Relational Graph Convolutional Network (R-GCN) exhibit marginally improved predictive capabilities in comparison to solely relying on raw gene transcriptomic data.

Keywords: Precision Medicine, Inflammatory Bowel Disease, Graph Convolutional Network, Knowledge Graph, Drug Response Prediction

1 Introduction

1.1 Graph neural networks

The rapid progress in computational resources (e.g., GPUs), evolving deep learning architectures, and the abundance of training data have enabled the broad application of deep learning in various machine learning tasks.

Deep learning techniques are now pervasive in a multitude of downstream applications, including instance segmentation, object detection, speech recognition, text classification, and many more[1]. Notably, in the age of large language models, innovative concepts like zero-shot, one-shot, and few-shot learning have further boosted the capabilities of deep learning.

Deep learning has undoubtedly proven its prowess in capturing intricate patterns within Euclidean data (e.g., images, text, and videos). However, the landscape of

modern applications is witnessing a growing trend where data is represented not as traditional Euclidean data but rather in the form of graphs[1].

Graph neural networks (GNNs) hold immense potential and find applications in diverse domains. Consider, for instance, a bibliographic network, which serves as an academic network and is effectively represented as a heterogeneous graph (HG). Such a graph comprises various node types, including authors, papers, venues, and terms[2]. In recommendation systems, an entity like the Amazon Product Graph emerges, representing an incredibly complex task—associating every product on Amazon with both concrete and abstract concepts. This empowers shoppers to discover products that align with their specific needs.

1.2 Precision medicine

In the field of biomedicine, due to the complexity of the biological system, there is no one-size-fits-all drug treatment for curing a specific disease. Precision medicine requires both personalized diagnostic strategies and targeted treatments. In our research, we aim to predict the therapeutic efficacy of drugs used in treating a medical condition known as inflammatory bowel disease[3]. Our research objective is to focus on training a binary patient classifier, which is adept at distinguishing between patients who are responsive and those who exhibit non-responsive reactions to a drug treatment.

The advancement of heterogeneous graphs (HGs) [4] demonstrates the ability to bring together a wide array of diverse entities. Therefore, it is feasible to use the technology of graph neural networks to implement precision medicine in curing a specific disease. For example, biomedical companies such as Astrazeneca harness the potential of knowledge graphs to consolidate a wealth of data sources, encompassing genomics, disease information, and drug data. These comprehensive knowledge graphs play a pivotal role in biomedical research, including patient classification, drug discovery, and so on[5]. Notably, the diverse and multimodal nature of these networks serves to enhance the performance of predictive models compared to traditional machine learning models. Additionally, they generalize the model performance on unseen data that was not encountered during the training phase, ultimately enhancing the interpretability of the models[6, 7].

Diseases are typically categorized based on the symptoms presented by patients, which can often be attributed to molecular dysfunctions, including genetic mutations[5]. To identify important gene expression features that may explain the disease, we leverage the empirical cumulative distribution function (eCDF) based on the quantitative measurements for each feature in the dataset. By comparing the transcriptomic profiles of healthy individuals(i.e. control group) with those of patients(i.e.baseline group), we can uncover highly-expressive genes for identifying disease.

In our research, our main contributions are as follows. We utilize Precision Medicine Knowledge Graph (PrimeKG)[8] as our foundational knowledge base, integrated with patient-specific transcriptomic data. Employing graph representation learning methods, notably the Relational Graph Convolutional Network (R-GCN)[9]

and RotaE[10], we adhere to the framework provided by CLEP. This approach enables us to acquire a interpretable latent patient representation(i.e.node embeddings) and then use four machine learning techniques to classify patients based on the new patient representations.

2 Literature review

2.1 Related work

Knowledge graphs can be divided into homogenous and heterogenous graphs. Although there have been ample studies of embedding technology on homogeneous graphs[11] that consist of only one type of node and edge, they are rare in essence. The development of novel high-throughout experimental techniques has led to a broad availability of biological data from multiple entity types. In practice, integrating multiple data types can be advantageous, particularly in the context of complex diseases, where no single type of data can explain the cause of dysfunction. To enhance the utility of gene expression data, which can be subject to noise and variability, modern approaches involve transforming co-expression matrices into more topologically meaningful representations. Notably, recent advancements encompass the integration of co-expression matrices with established biomedical networks like GO annotations and PPI networks[12].

Precision Medicine Knowledge Graph (PrimeKG)[8] falls into the category of heterogenous graphs, providing a holistic and multimodal view of diseases. It integrates 20 high-quality resources, biorepositories, and ontologies to be curated into this knowledge graph. PrimeKG captures information on ten major biological scales, including disease-associated perturbations in the proteome, biological processes, molecular pathways, anatomical and phenotypic scales, environmental exposures, and the range of approved and experimental drugs together with their therapeutic action. PrimeKG demonstrates that it improves coverage of diseases, both rare and common, by one-to-two orders of magnitude compared to existing knowledge graphs. Moreover, textual descriptions of clinical guidelines for drug and disease nodes are supplemented to enable multimodal analyses in future work. This knowledge graph is recognized for its enhanced completeness and up-to-date information when compared to PPI network [12] mentioned previously.

Also, rather than relying on manual grouping supervised by experts, PrimeKG adopts a semi-automated and unsupervised approach to cluster similar disease concepts together. Specifically, it leverages ClinicalBERT[13], a BERT-based language model, which encodes medical semantics due to its pretraining on biomedical knowledge derived from PubMed. This approach empowers PrimeKG to intelligently group disease concepts based on their inherent similarities and connections, without the need for manual intervention.

Extensive research efforts have been dedicated to the development of techniques for creating embeddings in both homogeneous and heterogeneous graph structures. The field has witnessed the development of methodologies like Knowledge Graph Embedding Models (KGEMs). These models have been designed to encode entities

and their relations within a KG into a latent feature space, preserving the graph's structural information and node information. They open the door to a plethora of downstream applications in the fields of machine learning and artificial intelligence. In the medical field, KGEMs have predominantly been harnessed for link prediction tasks, spanning from the anticipation of side effects [14] to the forecast of disease-gene associations [15].

In this paper, we will follow the framework known as Clinical Embedding of Patients(CLEP). It is a proven integrative approach that is designed to integrate prior knowledge and patient-level transcriptomics data[16]. CLEP is a hybrid data and knowledge-driven framework conducted by Vinay and Mehdi. By integrating prior knowledge into original transcriptomics data, it is proven that new patient representations significantly improve performance on the binary patient classification tasks within the context of Alzheimer's disease (AD).

2.2 Limitations

Our study distinguishes itself from existing literature reviews in several crucial aspects. Previous reviews typically focus on deep learning applications, particularly on Euclidean data[17] or concentrate solely on homogeneous graphs[18]. Furthermore, some biomedically focused reviews have primarily explored graph neural networks in the context of molecular generation, drug discovery[19], or repurposing. A limited number of studies have explored the application of node classification in patient categorization, particularly when examining response patterns at the gene or protein level.

2.3 The current applications and our study

Previous framework like PPI-KG mainly focus on protein-protein interactions and draw data from just six sources[16]. Our study instead integrate information from a richer pool that has 20 high-quality resources. PrimeKG[8] is a vast and comprehensive knowledge graph that encompasses 17,080 diseases and incorporates a staggering 4,050,249 relationships[8]. Therefore, our study extends its analysis further, not only on protein-protein interaction but also on other biological entities in the network. Moreover, biological data often requires ongoing updates and timely maintenance. PrimeKG contains information up to July 10, 2023, which ensures the knowlegde graph accurately represent the latest information and development.

In addition to the breadth and timeliness of our data sources, our study focuses on a specific disease and its corresponding drug treatments. We concentrate our efforts on inflammatory bowel disease (IBD) and specifically investigate anti-p40 and anti-TNF[3] therapies among the available treatment options.

Our approach is geared towards integrating a wealth of resources and mapping them onto patient-level gene expression datasets. This integration allows us to gain deeper insights into how drugs interact with patients at the gene expression level. Specifically, our primary goal is to build a classification model using a graph neural network that performs well in binary patient classification tasks. Second, we aim to identify the roles of key biological components in the networks,

including genes, proteins, and biological pathways, in binary patient classification tasks with the help of explainable AI. Furthermore, we would explore potential treatment correlates that contribute to the effective management of IBD beyond the treatments we initially selected.

In summary, our survey endeavors to connect these diverse domains by integrating a heterogeneous public knowledge graph (i.e., PrimeKG) with patient-level transcriptomic data. In the process of generating latent representations of patient embeddings, we follow the steps outlined by CLEP as a foundational benchmark to produce a novel patient representation. We proceed to experiment with evaluating the predictive capabilities of patient representations that are generated by shallow and deep embedding networks in the binary patient classification task. Specifically, relational Graph Convolutional Networks (R-GCNs)[9] and RotatE[10] are used in this study. Finally, we aim to uncover how genes, proteins, and biological pathways play biological roles in patient classification tasks. Since deep learning is inherently a black box and hard to interpret in comparison to traditional machine learning, explainable AI techniques would be used to identify the essential biomarkers.

3 Materials and methods

3.1 Input data

To conduct our study, we need two key pieces of data: a patient-level transcriptomic dataset and a knowledge graph called PrimeKG. The patient-level dataset contains gene/protein expression values for each patient sample measured by Affymetrix microarray technology. To efficiently create the latent representation of the knowledge graph, it's crucial to integrate genes and proteins aligning with the public knowledge graph. Additionally, pruning the public knowledge graph to form a subgraph is vital step, ensuring it contains minimal noise and extraneous information.

3.1.1 Patients Transcriptomic data

In this work, we collect two transcriptomic datasets respectively experiemented by two different studies. The first gene expression profiling study[20] was conducted in which ileum biopsy samples were collected from patients (n=362) with moderate-to-severe Crohn's disease and from non-IBD subjects (n=26). It includes 86 baseline samples and 26 control samples. The second study[21] was conducted on patients with moderate-to-severe ulcerative colitis who treated with ustekinumab (n=364) or placebo (n=186) at week 0. Colonic biopsy samples were collected from 550 patients and 18 healthy subjects at baseline for RNA extraction and microarray profiling. It includes 364 baseline samples and 18 control samples.

Batch effects refer to variations in experimental data not related to the experimental design, often due to differences in sample handling, experiment locations, or processing methods. In our case, these effects stem from datasets analyzed by different researchers and originating from varied sample sources, such as the colon

and ileum. Therefore, our study was mainly executed on the second dataset due to its larger patient samples size in order to avoid the impact of batch effect.

For the binary classification task aimed at distinguishing between responders (R) and non-responders (NR) to ustekinumab treatment for inflammatory bowel disease, our data selection process focuses on the utilization of baseline and control patient group data.

Dataset	Treatment	Sample Type	N (subject) Groups
GSE112366	ustekinumab	ileum	112(86 baseline, 26 control)
GSE206285	ustekinumab	colonic biopsy	382(364 baseline ; 18 control)

Table 1: Lists of Assay Datasets

3.1.2 Public knowledge graph

In the context of our specific case, we employ a publicly accessible knowledge graph, referred to as PrimeKG. This knowledge graph seamlessly integrates data from 20 high-quality resources, collectively comprising a vast network of 4,050,249 relationships. It opens the doors for unraveling intricate associations among 10 distinct node types, including drugs, genes/proteins, biological processes, diseases, molecular functions, pathways, effects/phenotypes, cellular components, anatomy, and exposure. Such a comprehensive resource empowers systematic investigations into the intricate relationships between diseases and various biomedical entities.

Our studies focus on the predictive task associated with discerning patterns of patient response to drug treatments within the context of gene expression levels. Through a judicious combination of domain-specific knowledge and expert guidance, we refine PrimeKG into a subgraph characterized by interconnections among gene/protein nodes, disease, bioprocess, exposure and other pertinent biological entities. This process also entails the filter of unnecessary entities, which could introduce undesirable noise into the biological network, such as anatomical components and molecular functions. Notably, we introduce the link relationships of disease-disease within the network to gain insights into the associations between inflammatory bowel disease and other diseases of relevance.

3.2 Methodologies

3.2.1 Incorporating patients into the knowledge graph

Since patients transcriptomic data inherently contain a far greater number of features than samples, it is hard to conduct the binary classification task. To be specific, the effects of drug treatments are not limited to molecules to which they directly bind in the body. Instead, these effects spread throughout biological networks in which they act. The effect of a drug on a disease is inherently a network phenomenon. Consequently, the first step of our methodology involves the integration of each patient sample as an individual node within the PrimeKG framework.

Larger networks introduce the risk of diluting patient representations with irrelevant features. Our study aims to identify key features and integrate patients into the knowledge graph to refine the subgraph and improve patient representations. Our methodology leverages the empirical cumulative distribution function (eCDF)[16] based on the quantitative measurements for each feature in the dataset which can be mapped to the KG. We iterate over patients and calculate the eCDF for every gene in order to check if any of its features is significant. We select 5% quantiles of the eCDF as the threshold to build our biological network. It's important to note that our utilization of the empirical cumulative distribution function (eCDF) results in the creation of binary link relations(i.e. -1 or +1), between each patient sample and gene/protein feature known as down_reg(-1) and up_reg(+1). The sign of this link relation serves as an indicator, signifying whether a specific patient falls to a lower/higher extreme distribution of a particular feature.

Finally, the link relationships summary of the final pruning subgraph is listed below:

Relation type	**Count**	**Percentage (%)**
protein_protein	642,150	35.95
bioprocess_protein	289,298	16.20
up_reg	288,569	16.16
down_reg	259,954	14.55
disease_protein	160,822	9.00
pathway_protein	85,183	4.77
drug_protein	50,933	2.85
phenotype_protein	6,660	0.37
exposure_protein	2,424	0.14
disease_disease	26	0.00

Table 2: Link relation summary of the subgraph of PrimekG

The integration of patient gene expression data into the subgraph of PrimeKG serves as the subsequent generation of the feature representations for each patient (i.e. node embeddings) through the utilization of Knowledge Graph Embedding Models (KGEMs) in the next step.

3.3 Generating patient representations

We denote directed and labeled multi-graphs as $G = (V, \mathrm{E}, R)$, with nodes (entities) $v_i \in V$ and labeled edges (relations) $(v_i, r, v_j) \in E$, where $r \in R$ is a relation type. R-GCNs[9] are related to a recent class of neural networks operating on graphs, and are developed specifically to deal with the highly multi-relational data characteristic of realistic knowledge bases. It is motivated as an extension of GCNs that operate on local graph neighborhoods[22]. In the stage of message passing layer of RGCN, it can depicted as:

$$h_i^{(l+1)} = \sigma\left(\sum_{r\in R}\sum_{j\in N_i^r}\frac{1}{c_{i,r}}W_r^{(l)}h_j^{(l)} + W_0^{(l)}h_i^{(l)}\right), \tag{1}$$

where $h_i^{(l+1)}$ denotes the updated embeddings of node i at layer l+1. $\sum_{r \in R}$ represents the summation over all relation types in the knowledge graph, and $c_{i,r}$ is the normalization constant term. $W_r^{(l)}$ is a weight matrix specific to relation r at layer l. To ensure that the representation of a node at layer l + 1 can also be informed by the corresponding representation at layer l, single self-connection of a special relation type is added to each node in the data.

Intuitively, (1) accumulates transformed feature vectors of neighboring nodes through a normalized sum. Each update of node embeddings of node i at layer l+1 is learnt from local neighborhood in the computational graph and its corresponding representation at layer l.

3.3.1 Patient Classification

The latent representations of node embeddings of each patient sample generated from R-GCN were used to distinguish responders from non-responders to ustekinumab treatment for inflammatory bowel disease. The prediction power of are compared between the new patient representations and raw gene transcriptomic data.

Our study utilized a 5-fold nested cross-validation process, repeated five times, to evaluate four distinct machine learning classification models. Parameter tuning is performed via a grid search approach for each model. To assess the prediction performance on the hold-out test dataset, we used either the area under the receiver operating characteristic curve (AUC-ROC) or the area under the precision-recall curve (AUC-PR) as our evaluation metrics.

3.3.2 Uncovering important biological patterns

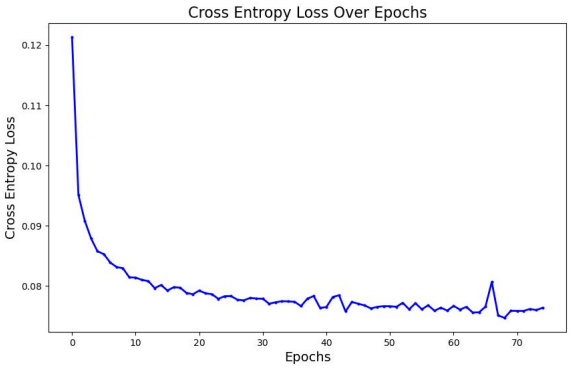
After identifying the model that exhibits the best performance in predicting drug treatment outcomes, our main objective is to unveil the pivotal genes/proteins, biological processes, and link relations within the knowledge graph that significantly influence the binary prediction task.

4 Experimental Results

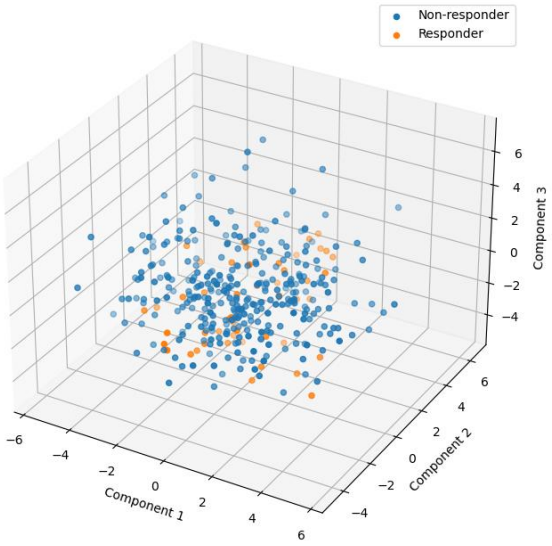
This research showcases the effectiveness of our method in improving the accuracy of machine learning models for binary classification in inflammatory bowel disease (IBD). By incorporating patient-specific transcriptomic data into PrimeKG, we facilitate the creation of patient node embeddings. These embeddings are particularly useful in node classification tasks. We employed a Relational Graph Convolutional Network (R-GCN) to generate these embeddings, with a training regimen of 75 epochs using cross-entropy as the loss function. The Adam optimizer was used with a learning rate of 0.001.

Figure 1(a) presents a graphical representation of the cross-entropy loss observed in the hold-out test dataset across 75 epochs. We decompose the 100 dimensions of the new patient representations into a 3D visual plot using principal component

analysis shown in figure1(b). It indicates our embeddings result has a marginally discernible predictive capability than the raw data.



(a) Cross entropy loss over 75 epochs



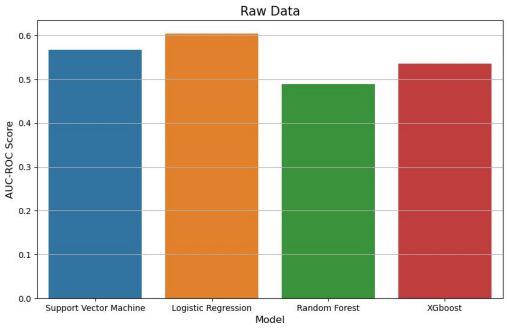
(b) 3D visualization of PCA components decomposed from 100 dimensions of latent representations

Figure 1: Visualizations of cross entropy loss and node embeddings representation

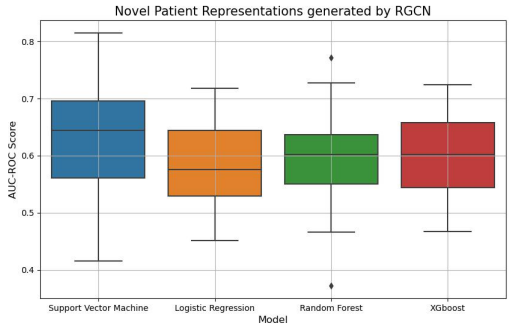
Furthermore, we engage in comparing the performance of several traditional ML methods in distinguishing between baseline patients treated with ustekinumab that are responders v.s. non-responders. Given the imbalanced nature of our dataset, we allocate 20% of the data as a hold-out test set and apply a 5 repeats of the 5-fold nested cross-validation procedure on the remaining dataset. We compare the performance of both the benchmark raw data and new patient representations in Figure 2,3 and 4.

Figure 2(a) reveals that using raw data alone for binary classification results in limited predictive accuracy on the test dataset, with logistic regression as the most

effective model, achieving just above a 0.6 AUC-ROC score. Other models like support vector machines, random forests, and XGBoost are less effective. Figure 2(b) shows that all models trained on new patient representations have improved, averaging around 0.6 AUC-ROC scores. This indicates that CLEP-generated patient representations enhance binary classification performance compared to original data, though the overall enhancement remains modest.



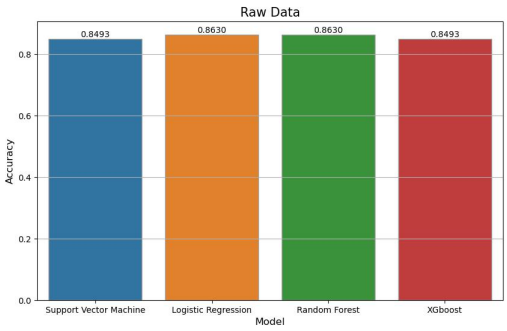
(a) Benchmarking experiments based on raw transcriptomic data



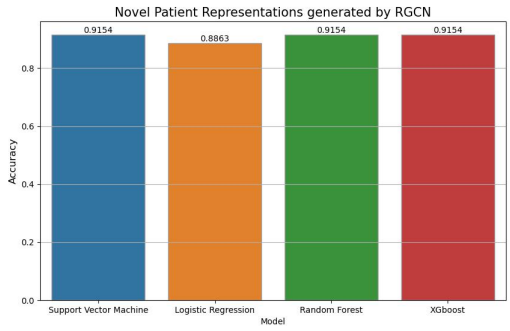
(b) R-GCN generated new patient representations

Figure 2: Comparative analysis evaluated by AUC-ROC

Figure 3 demonstrates that the model accuracy achieved using the new patient representations is marginally higher compared to that obtained by solely employing benchmark raw data.



(a) Benchmarking experiments based on raw transcriptomic data

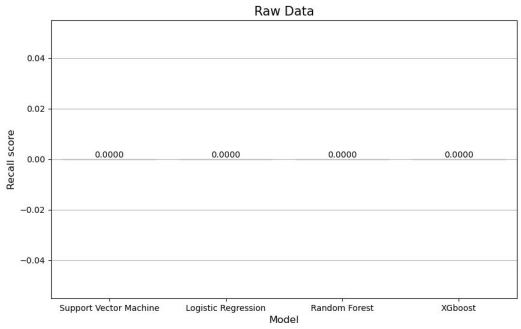


(b) R-GCN generated new patient representations

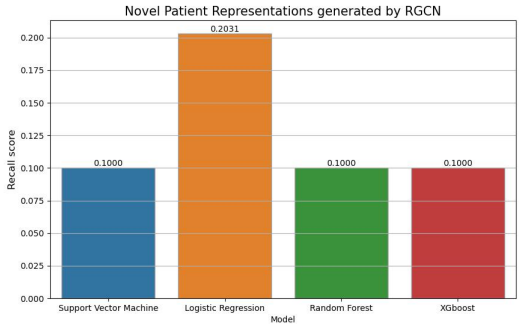
Figure 3: Comparative analysis evaluated by accuracy

In the context of the imbalanced dataset, the ability to accurately predict the positive samples is more valuable than achieving a high overall prediction accuracy across a larger portion of samples. Figure 4 finally illustrates the recall score performance on both new patient representations and raw data. Figure 4(a) indicates that all four models have the average 0% recall score on the test dataset, highlighting their inability to effectively predict positive samples, especially responders, using only original gene expression data. In contrast, figure 4(b) demonstrates that our model, when trained with improved data representations, does attain a

certain level of performance in detecting positive samples. Notably, logistic regression emerges as the most effective, successfully predicting approximately 20% of the positive samples. However, it is crucial to recognize that there is significant room to increase the recall rate further.



(a) Benchmarking experiments based on raw transcriptomic data



(b) R-GCN generated new patient representations

Figure 4: Comparative analysis evaluated by recall score

5 Discussion

In this research, we have integrated patient-level transcriptomic data into PrimeKG to investigate the capabilities of Knowledge Graph Embedding Models (KGEMs) in harnessing the informational context of graph neighborhood nodes. By following CLEP framework, we have illustrated how node representations generated by various KGEMS can be utilized in the downstream binary patient classification tasks. This method demonstrates potential in enhancing the predictive accuracy of patient response patterns to drug treatment for inflammatory bowel disease, as compared to the use of original gene transcriptomics data. The evaluation of all four models, based on accuracy, AUC-ROC, and recall scores, shows improved predictive capabilities with R-GCN generated embeddings.

However, there are several limitations worth to note that impact the learning outcomes. A primary concern pertains to the methodology employed for partitioning the graph into training, validation, and testing datasets. The current strategy of random edge splitting in our study may potentially introduce serious biases and affect the generalizability of the results.

Another aspect of consideration is the interpretability of deep learning models when compared to other traditional models. Nonetheless, recent advancements have focused on developing explanation techniques for deep learning models applied to various data types, including images, text, and graphs. Specifically, in the realm of Graph Neural Networks (GNNs), tools such as GNN Explainer[23] and SubgraphX[24] have emerged, concentrating on explicability at the node, edge, or graph levels.

Despite these advancements, our study has not delved into the biological interpretation of the learned patterns, primarily due to time constraints. Unraveling key

biomarkers that differentiate responders from non-responders remains a compelling and unexplored avenue in our research. Furthermore, the limited computational capacity of our GPU resources has restricted our ability to incorporate link weight and node feature enhancements into our knowledge graph. Addressing these limitations could potentially elevate the performance and applicability of our models in precision medicine.

6 Acknowledgements

We extend our profound gratitude to Professor Dr.Qiwei He for her invaluable discussions and technical guidance throughout this research. Her insights and suggestions have been instrumental in shaping the direction and execution of this study. We are especially grateful for the support provided by Dr.Chia-chien Chiang. Her contributions have been pivotal in multiple facets of our project, including data collection, subgraph knowledge graph generation, and model tuning. Her extensive domain knowledge and expertise have significantly enhanced our research methodologies and outcomes.

References

[1] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and S Yu Philip. A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems*, 32(1):4–24, 2020.

[2] Xiao Wang, Deyu Bo, Chuan Shi, Shaohua Fan, Yanfang Ye, and S Yu Philip. A survey on heterogeneous graph embedding: methods, techniques, applications and sources. *IEEE Transactions on Big Data*, 9(2):415–436, 2022.

[3] Sudarshan Paramsothy, Adam K Rosenstein, Saurabh Mehandru, and Jean-Frederic Colombel. The current state of the art for biological therapies and new small molecules in inflammatory bowel disease. *Mucosal immunology*, 11(6):1558–1570, 2018.

[4] Yizhou Sun and Jiawei Han. Mining heterogeneous information networks: a structural analysis approach. *Acm Sigkdd Explorations Newsletter*, 14(2):20–28, 2013.

[5] Michelle M Li, Kexin Huang, and Marinka Zitnik. Graph representation learning in biomedicine. *arXiv preprint arXiv:2104.04883*, 2021.

[6] Richard J Chen, Ming Y Lu, Jingwen Wang, Drew FK Williamson, Scott J Rodig, Neal I Lindeman, and Faisal Mahmood. Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. *IEEE Transactions on Medical Imaging*, 41(4):757–770, 2020.

[7] Charlotte A Nelson, Atul J Butte, and Sergio E Baranzini. Integrating biomedical research and electronic health records to create knowledge-based

biologically meaningful machine-readable embeddings. *Nature communications*, 10(1):3045, 2019.

[8] Payal Chandak, Kexin Huang, and Marinka Zitnik. Building a knowledge graph to enable precision medicine. *Scientific Data*, 10(1):67, 2023.

[9] Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling. Modeling relational data with graph convolutional networks. In *The Semantic Web: 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3–7, 2018, Proceedings 15*, pages 593–607. Springer, 2018.

[10] Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. Rotate: Knowledge graph embedding by relational rotation in complex space. *arXiv preprint arXiv:1902.10197*, 2019.

[11] Peng Cui, Xiao Wang, Jian Pei, and Wenwu Zhu. A survey on network embedding. *IEEE transactions on knowledge and data engineering*, 31(5):833–852, 2018.

[12] Hryhorii Chereda, Annalen Bleckmann, Frank Kramer, Andreas Leha, and Tim Beissbarth. Utilizing molecular network information via graph convolutional neural networks to predict metastatic event in breast cancer. In *GMDS*, pages 181–186, 2019.

[13] Emily Alsentzer, John R Murphy, Willie Boag, Wei-Hung Weng, Di Jin, Tristan Naumann, and Matthew McDermott. Publicly available clinical bert embeddings. *arXiv preprint arXiv:1904.03323*, 2019.

[14] Marinka Zitnik, Monica Agrawal, and Jure Leskovec. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics*, 34(13):i457–i466, 2018.

[15] Daniel S Himmelstein and Sergio E Baranzini. Heterogeneous network edge prediction: a data integration approach to prioritize disease-associated genes. *PLoS computational biology*, 11(7):e1004259, 2015.

[16] Vinay Srinivas Bharadhwaj, Mehdi Ali, Colin Birkenbihl, Sarah Mubeen, Jens Lehmann, Martin Hofmann-Apitius, Charles Tapley Hoyt, and Daniel Domingo-Fernández. Clep: a hybrid data-and knowledge-driven framework for generating patient representations. *Bioinformatics*, 37(19):3311–3318, 2021.

[17] Tiffany J Callahan, Ignacio J Tripodi, Harrison Pielke-Lombardo, and Lawrence E Hunter. Knowledge-based biomedical data science. *Annual review of biomedical data science*, 3:23–41, 2020.

[18] William L Hamilton, Rex Ying, and Jure Leskovec. Representation learning on graphs: Methods and applications. *arXiv preprint arXiv:1709.05584*, 2017.

[19] Thomas Gaudelet, Ben Day, Arian R Jamasb, Jyothish Soman, Cristian Regep, Gertrude Liu, Jeremy BR Hayter, Richard Vickers, Charles Roberts, Jian Tang, et al. Utilizing graph machine learning within drug discovery and development. *Briefings in bioinformatics*, 22(6):bbab159, 2021.

[20] Kelli L VanDussen, Aleksandar Stojmirović, Katherine Li, Ta-Chiang Liu, Patrick K Kimes, Brian D Muegge, Katherine F Simpson, Matthew A Ciorba, Jacqueline G Perrigoue, Joshua R Friedman, et al. Abnormal small intestinal epithelial microvilli in patients with crohn's disease. *Gastroenterology*, 155(3):815–828, 2018.

[21] Polychronis Pavlidis, Anastasia Tsakmaki, Eirini Pantazi, Katherine Li, Domenico Cozzetto, Jonathan Digby-Bell, Feifei Yang, Jonathan W Lo, Elena Alberts, Ana Caroline Costa Sa, et al. Interleukin-22 regulates neutrophil recruitment in ulcerative colitis and is associated with resistance to ustekinumab therapy. *Nature Communications*, 13(1):5820, 2022.

[22] David K Duvenaud, Dougal Maclaurin, Jorge Iparraguirre, Rafael Bombarell, Timothy Hirzel, Alán Aspuru-Guzik, and Ryan P Adams. Convolutional networks on graphs for learning molecular fingerprints. *Advances in neural information processing systems*, 28, 2015.

[23] Zhitao Ying, Dylan Bourgeois, Jiaxuan You, Marinka Zitnik, and Jure Leskovec. Gnnexplainer: Generating explanations for graph neural networks. *Advances in neural information processing systems*, 32, 2019.

[24] Hao Yuan, Haiyang Yu, Jie Wang, Kang Li, and Shuiwang Ji. On explainability of graph neural networks via subgraph explorations. In *International conference on machine learning*, pages 12241–12252. PMLR, 2021.