# HeteroKG: Knowledge graphs for multi-modal learning

**Chinmay Prabhakar[1], Anjany Sekuboyina[1,2], Suprosanna Shit[1,2], Bjoern Menze[2]**
[1]University of Zurich, [2]Technical University of Munich
chinmay.prabhakar@uzh.ch

## Abstract

Medical knowledge graphs (KG) are a source of highly granular, semantically rich, and curated medical ontologies. However, they have seen limited adoption in various multi-modal medical imaging tasks owing to their sheer size. Instead, semantic label embeddings from language models such as BERT and word2vec are currently employed. These embeddings are derived from word co-occurrences and encode rich semantic associations. However, they lack explicit relational information which KG intrinsically encode. On the other hand, the expressive power of KG is limited by its parsed size. In view of these observations, we propose a way to learn KG embeddings on the parsed heterogeneous graph and complement it with language embeddings. We test our hypothesis on generalized zero-shot learning of chest radiographs.

**Keywords:** Graph Neural Network, heterogeneous graphs, generalized zero shot learning

## 1. Introduction

There has been a surge in multi-modal text and image processing in the medical domain over recent years. Tasks such as few-shot learning (Mahajan et al., 2020), zero-shot learning (Hayat et al., 2021) and handling data-imbalance (Zhang et al., 2020) are using textual information to account for the semantic knowledge gap.

However, the traditional language-based models such as word2vec (Zhang et al., 2019) and BioBERT (Devlin et al., 2018) learn embeddings based on the distributional hypothesis and masked language modeling respectively. As such, although these embeddings are rich, they are local in their nature. They can not learn viable relationships between medical concepts that do not appear together in medical reports. On the other hand, we have knowledge graphs (KG) that are a curated knowledge bank of medical ontologies e.g. the Unified Medical Language System (*UMLS*) (Bodenreider, 2004). We can find explicit relationships between medical entities using such a KG. Intuitively, a KG gives us a more *global* view of these relationships. Thus, an approach combining these two sources of information can help us in various downstream tasks.

Despite the utility of KG, processing them is difficult owing to their size and the different kinds of relations they encode. In this work, we parse the KG conditioned on the downstream application and model the parsed graph as a heterogeneous graph. The nodes would be various medical ontologies while we have a separate edge for each relation type. A heterogeneous treatment of the KG allows us to learn node features that are enriched based on the different relationship types. This draws parallels from multi-head attention that has been used very successfully in the case of transformers (Vaswani et al., 2017) and graph attention networks (Veličković et al., 2017).

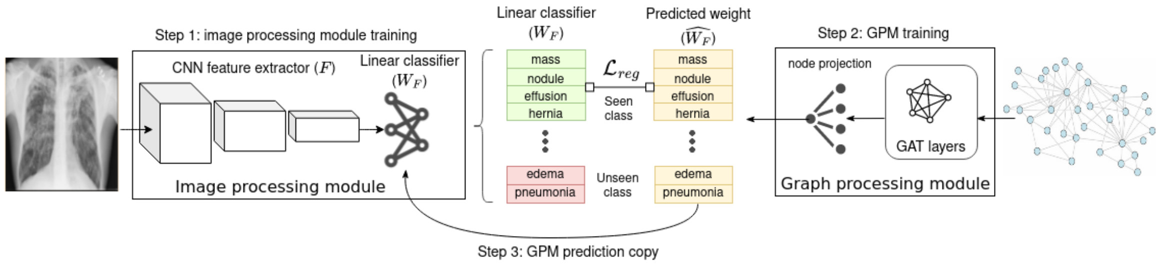


Figure 1: We first train the Vision backbone using seen classes to learn weights $W_F$ for the linear classifier. Second, the Graph processing module is trained to align predictions $\hat{W}_F$ with $W_F$ for seen classes. Finally, the enriched $\hat{W}_F$ are copied over and replace the $W_F$.

In this paper, we present an approach to using a knowledge graph for generalized zero-shot learning of chest x-rays. We show that modeling a KG as a heterogeneous graph allows us to learn information complementary to using only BERT embeddings. We show proof-of-concept results for generalized zero-shot learning on NIH chest x-ray dataset (Wang et al., 2017), where a combination of the two embeddings works better than the individual ones.

## 2. Method

We employ a multi-stage training process inspired from (Kampffmeyer et al., 2019). The method overview is presented in Figure 1.

**Graph construction** We use UMLS as the knowledge graph for our application. We limit ourselves to only five relationship types based on prior medical knowledge. These are: *is_associated_anatomic_site_of*, *finding_site_of*, *part_of*, *inverse_is_a* and *has_member*.

We look up the target labels in the UMLS. These act as our seed entities. Starting from each of these entities, we extract its 5-hop neighborhood. This operation is repeated for each of the different entity types. The different edge types are combined using a heterogeneous graph (Figure 2). We then compute the node2vec (Grover et al., 2016) embeddings for the parsed graph. Finally, we initialize each of the nodes with the concatenated node2vec and BioBERT embeddings.

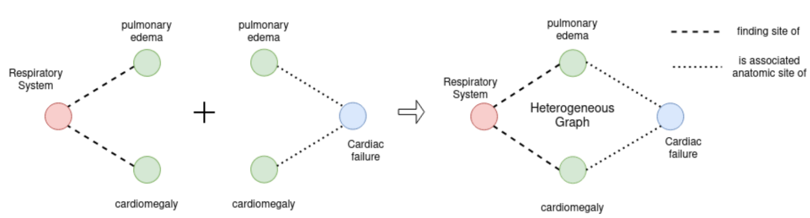


Figure 2: We parse graph based on each of the relations. We use two target nodes, *pulmonary edema* and *cardiomegaly* for reference. Target nodes are shown in green. The two subgraphs are obtained by parsing nodes based on two *different* relation types. The final graph is a heterogeneous graph containing both types of edge.

**Model:** The image processing module consists of a CNN feature extractor $F$ and a classification layer $W_F$. The Graph Processing module ($GPM$) consists of three graph attention layers and a node projection layer to predict the classifier weights $\hat{W}_F$.

**Training:** Firstly, we train the image processing module using only the seen classes. The classifier $W_F$ is semantically rich for the seen classes while it remains random for the unseen classes. We would use our parsed graph to bridge this information gap. Secondly, the parsed graph is processed using the *GPM*. The GPM predictions $\hat{W}_F$ are aligned with the classifier weights $W_F$ for seen classes. We hypothesize that during the alignment, weights of unseen classes are also enriched due to GPM weight sharing. Finally, we replace the $W_F$ with the enriched GPM predictions $\hat{W}_F$ (ref Fig 1).

To account for a surplus of positive or negative samples in a mini-batch, we use the multi-label classification loss(Chen et al., 2020) while training the image processing module. To train the GPM we use the following loss function:-

$$\mathcal{L}_{reg} = \sum_{j \in seen} ||W_F^j - \hat{W_F}^j||^2 \tag{1}$$

where $W_F^j$ is the weight of the $j^{th}$ disease. The loss is computed *only* for the seen classes.

Please note that the GPM module is not used during inference. Only the image processing module is required to classify the unseen test samples.

## 3. Experiment

**Dataset** We evaluate our method on the NIH Chest X-ray dataset (Wang et al., 2017). We have radiographs with multi-label annotations.The 112,120 frontal X-ray images are split into training (70%), validation (10%) and test sets (20%). Each image is associated with 14 class labels. We use *Atelectasis, Effusion, Infiltration, Mass, Nodule, Pneumothorax, Consolidation, Cardiomegaly, Pleural Thickening, and Hernia* as the *seen* classes while *Edema, Pneumonia, Emphysema, and Fibrosis* are the *unseen* classes (same as (Hayat et al., 2021)), resulting in 30,758 training, 4,474 validation and 10,510 test images.

**Evaluation metrics.** We report overall precision, recall, and f1 scores for the top $k$ predictions (where $k \in 2, 3$) and the average area under the receiving operating characteristic curve (AUROC) for *seen* and *unseen* classes and their harmonic mean.

Table 1: Performance Evaluation on the NIH Chest X-ray dataset. We report the results using Precision@k, Recall@k , F1@k for k ∈ {2, 3}. We also report AUROC for *seen (S) & unseen (U) classes and the Harmonic Mean (HM).*

| | k=2 | | | k=3 | | | AUROC | | |
|---|---|---|---|---|---|---|---|---|---|
| **Method** | **p@k** | **r@k** | **f1@k** | **p@k** | **r@k** | **f1@k** | **S** | **U** | **HM** |
| Node2Vec Grover et al. (2016) | 0.26 | 0.32 | 0.28 | 0.23 | 0.43 | 0.29 | 0.78 | 0.54 | 0.64 |
| BioBERT (Devlin et al., 2018) | 0.31 | **0.35** | 0.33 | **0.26** | 0.44 | **0.33** | **0.80** | 0.60 | 0.69 |
| **HeteroKG (Ours)** | **0.33** | **0.35** | **0.34** | **0.26** | **0.45** | **0.33** | **0.80** | **0.64** | **0.71** |

**Results** We summarize the results in Table 1. As we can see, while BERT embeddings are semantically rich, only a combination of node2vec and BioBERT gives the best performance. None of these embeddings are expressive enough on their own.

## 4. Conclusion

We propose a novel way to model medical knowledge graphs as a heterogeneous graph and use it for generalized zero-shot learning task. We show that it encodes information complementary to word embedding models and a combination of these two information sources might be the preferred way going forward.

## References

Olivier Bodenreider. The unified medical language system (umls): integrating biomedical terminology. *Nucleic acids research*, 32(suppl_1):D267–D270, 2004.

Bingzhi Chen, Jinxing Li, Guangming Lu, Hongbing Yu, and David Zhang. Label co-occurrence learning with graph convolutional networks for multi-label chest x-ray image classification. *IEEE journal of biomedical and health informatics*, 24(8):2292–2302, 2020.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

Aditya Grover et al. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 855–864, 2016.

Nasir Hayat, Hazem Lashen, and Farah E. Shamout. Multi-label generalized zero shot learning for the classification of disease in chest radiographs, 2021.

Michael Kampffmeyer, Yinbo Chen, Xiaodan Liang, Hao Wang, Yujia Zhang, and Eric P Xing. Rethinking knowledge graph propagation for zero-shot learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11487–11496, 2019.

Kushagra Mahajan, Monika Sharma, and Lovekesh Vig. Meta-dermdiagnosis: Few-shot skin disease identification using meta-learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 730–731, 2020.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.

Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald M Summers. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2097–2106, 2017.

Yijia Zhang, Qingyu Chen, Zhihao Yang, Hongfei Lin, and Zhiyong Lu. Biowordvec, improving biomedical word embeddings with subword information and mesh. *Scientific data*, 6(1):1–9, 2019.

Yuhao Zhang, Hang Jiang, Yasuhide Miura, Christopher D Manning, and Curtis P Langlotz. Contrastive learning of medical visual representations from paired images and text. *arXiv preprint arXiv:2010.00747*, 2020.